**Title**: Validation of the Pyruvate Kinase Deficiency Diary (PKDD): a patient-reported outcome measure for pyruvate kinase (PK) deficiency

**Authors**: Shayna Egan, Erin Zagadailov, Adrian Jewett, Junlong Li, Feng Tai, Audra Boscoe, David A. Andrae

## Introduction

Pyruvate kinase (PK) deficiency is a rare hereditary disorder characterized by chronic hemolytic anemia and long-term complications such as iron overload, osteoporosis, and pulmonary hypertension. There are no patient-reported outcome (PRO) instruments that are specific for this condition. The PK Deficiency Diary (PKDD), a disease-specific PRO, was designed as an evening diary to assess signs and symptoms for patients with PK deficiency. The seven-item instrument assesses tiredness, jaundice, bone pain, shortness of breath, and energy levels. Six of the seven items employ a numeric rating scale (NRS) ranging from 0-10 and one item uses a 5-point verbal descriptor scale. The instrument underwent psychometric validation using blinded data from the eighty patients participating in the ACTIVATE trial, a randomized, multicenter, double-blind, phase 3 trial of mitapivat, an allosteric activator of PK, in non-regularly transfused adults with PK deficiency (NCT03548220).

## Methods

Completion rates and baseline response distributions were characterized using descriptive statistics. Inter-item correlations were estimated and item response theory (IRT) modeling was applied. A scoring system was established with item weighting informed by IRT modeling. The resultant baseline scores were used to assess reliability (internal consistency and test-retest) and validity (convergent and known-groups). Change scores for group comparisons were evaluated via anchor-based methods using the Patient Global Impression of Severity (PGIS) changes from baseline to weeks 12 and 24. The PGIS had a 5-point verbal descriptor scale (Not at all, A little, Moderately, A lot, and Very much).

## Results

Baseline PKDD data were available for 77 patients (96% completion rate). Response distributions skewed right with sparse endorsements at the higher levels, especially for the NRS-rated items. To facilitate IRT modeling, responses were recoded to a 0-4 scale. The IRT model for the baseline item responses resulted in adequate fit [Root Mean Squared Error of Approximation (RMSEA) 0.09, 95% CI = (<0.01, 0.132)]. Daily sum scores were first calculated and then converted to T-scores (mean of 50 and standard deviation of 10) employing a scoring conversion (Table). The conversion of sum scores to T-scores is based on the expected a-priori scores (Z) from the IRT model. A higher score represents more severe symptomatology and higher disease burden.

Internal consistency was assessed with McDonald's coefficient, $\omega = 0.86$, indicating a high level of reliability. Test-retest reliability via a two-way mixed effects Intra-Class Correlation Coefficient (ICC) was also high (0.94). Convergent validity was established when correlating PKDD scores with the FACT-An ($|r| = 0.73$), the SF-12 PCS ($|r| = 0.54$), and the SF-12 MCS ($|r| = 0.30$). Known-groups validity was moderate with mean scores ordering approximately as expected across baseline PGIS ratings and a linear trend ($\eta^2 = 0.27$).

Changes in scores from baseline to weeks 12 and 24 were assessed by stratifying PKDD scores by changes in PGIS (no change as the reference group). By comparing patients who reported a 2-point worsening in severity to those who reported no change on the PGIS, estimates of a change in PKDD scores at weeks 12 and 24 ranged from 3.0 to 13.9 points. Per anchor-based methods, a change of approximately 5 to 8 points could be considered a meaningful change in PKDD score.

**Conclusions**

The PKDD is now the first validated disease-specific PRO to assess signs and symptoms in patients with PK deficiency. The PKDD showed high internal consistency, test-retest reliability, and convergent validity. Further research investigating meaningful change scores would be beneficial. Due to the rare nature of PK deficiency, the ACTIVATE study enrolled a small number of patients overall which was a limitation in the anchor-based determination of meaningful change. Thus, the recommended estimates of meaningful change, although interpretable for the current study, are not definitive and could benefit from further investigation.

**Table. Pyruvate Kinase Deficiency Diary (PKDD) scoring table**

| Sum Score | Z Score | T Score |
|---|---|---|
| 0 | -2.4659 | 25 |
| 1 | -2.2214 | 27 |
| 2 | -2.0098 | 29 |
| 3 | -1.7975 | 32 |
| 4 | -1.5792 | 34 |
| 5 | -1.3639 | 36 |
| 6 | -1.1626 | 38 |
| 7 | -0.9751 | 40 |
| 8 | -0.7956 | 42 |
| 9 | -0.6206 | 43 |
| 10 | -0.4484 | 45 |
| 11 | -0.2766 | 47 |
| 12 | -0.1047 | 48 |
| 13 | 0.0646 | 50 |
| 14 | 0.2305 | 52 |
| 15 | 0.3968 | 53 |
| 16 | 0.5698 | 55 |
| 17 | 0.7528 | 57 |
| 18 | 0.9435 | 59 |
| 19 | 1.1344 | 61 |
| 20 | 1.3180 | 63 |
| 21 | 1.4913 | 64 |
| 22 | 1.6548 | 66 |
| 23 | 1.8122 | 68 |
| 24 | 1.9751 | 69 |
| 25 | 2.1539 | 71 |
| 26 | 2.3439 | 73 |
| 27 | 2.5258 | 75 |
| 28 | 2.6618 | 76 |