PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Psychometric evaluation of the NTDT-PRO questionnaire for assessing symptoms in patients with non-transfusion-dependent beta-thalassaemia
AUTHORS	Taher, Ali T.; Musallam, Khaled M.; Viprakasit, Vip; Kattamis, Antonis; Lord-Bessen, Jennifer; Yucel, Aylin; Guo, Shien; Pelligra, Christopher; Shields, Alan L.; Shetty, Jeevan K.; Miteva, Dimana; Bueno, Luciana Moro; Cappellini, MD

VERSION 1 – REVIEW

REVIEWER	Lam, Joyce KK Women's and Children's Hospital
REVIEW RETURNED	31-Jul-2022

GENERAL COMMENTS	This is a paper describing the validation of a Patient Reported Outcome (PRO) questionnaire developed specifically for patients with non-transfusion dependent beta thalassaemia (NTDT). It is an important study as there are no other disease-specific QOL assessment tools for NTDT patients, and it can potentially be used in other clinical trials as a standardized method of assessing the QOL impact of interventions in this group of patients. I have a number of minor points for the authors to address: 1. In the abstract, the design of the study is described as "a retrospective study". The scores from the various assessment tools were captured prospectively during the course of the BEYOND trial, so why would this study be described as being retrospective in nature? 2. As the trial participants were from a number of geographic regions, were the QOL instruments available in languages other than English? 3. Was any qualitative data captured regarding the ease of use of the NTDT-PRO questionnaire as compared to the other QOL instruments, especially the longer instruments such as the SF-36v2 and FACIT-F? 4. Will the authors be providing a link or information as to how the actual NTDT-PRO instrument can be accessed?

REVIEWER	Elalfy, Mohsen Ain Shams University
REVIEW RETURNED	31-Jul-2022

GENERAL COMMENTS	Dear corresponding Author It's a good retrospective had evaluated the psychometric properties were using data from the BEYOND trial (NCT03342404). Scores were moderately correlated with changes in hemoglobin levels; Why moderate correlation? and please comment the cut-off Hb level with high degree of correlation (all were < 100gm/L please comment on variability of PRO in relation to different scores?) Authors mentioned the Strengths and limitations of this study • Strengths of this study include use of well-validated PRO instruments such as PGI-S, PGI-C, SF-36v2®, and FACIT-F. However validation was done for this particular study? • The data used in this analysis were from a phase 2 interventional study with participants from multiple geographic regions and spanning a range of NTDT symptom severities. (inter-symptoms variations could be added in a paragraph in the discussion ** patients with severe DRAEs were different in PRO scores or not; please add in results section and discussion • The use of data from an interventional study allowed for changes in symptom severity to be observed, validating NTDT-PRO's sensitivity to identify longitudinal changes in symptoms. • Given that NTDT is a rare disease, limitations of the present study include the reduced sample size (Please confirm that no patients in BEYOND were excluded from the survey) being a retrospective study for typical psychometric evaluations. • Cut-off values used to define different levels of improvement in the responsiveness analysis are not well established and were based on certain assumptions. Please clarify Please stress on PROs differences in 1st 12 weeks vs 13-24; specially more missing dates of reporting in this period and number of patients had dropped from 149 tat baseline to 110 at 24 weeks Any special scores are related to reported AEs according to their severity

REVIEWER	Premawardhena, Anuja University of Kelaniya, Department of Medicine
REVIEW RETURNED	12-Aug-2022

GENERAL COMMENTS	NTDT Pro is clearly a good tool in the armamentarium of the researchers in the field of Thalassaemia and the paper effectively describes its validity against the other more well established tools. The daily entry of details on one hand. makes it accurate but on the other less. practical in the absence of Apps. which patients. carry with them, or are sufficiently motivated to use. None of these comments are however strictly relevant to this paper!

REVIEWER	Vitoratou, S King's College London, Department of Biostatistics and Health Informaticss
REVIEW RETURNED	30-Sep-2022

GENERAL COMMENTS	This is a very well written work and very interesting indeed. With respect to the analysis it was conducted appropriately. However the authors have conducted only assessments of reliability (internal consistency and stability) assuming that the factor structure is unidimensional. The sample size is not enough for

	CFA (ideally we would want this to be above 200) so i will not suggest to the authors to conduct such analysis. It is important however in my opinion to replace the title wit investigations for the reliability and responsiveness as opposed to psychometric evaluation to reflect that the latter is only partially conducted. It should be mentioned clearly in the manuscript what is assumed for dimensionality and why and also state as a limitation that no cfa is presented in these data.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Dr Joyce Lam, KK Women's and Children's Hospital

Comments to the Author:

This is a paper describing the validation of a Patient Reported Outcome (PRO) questionnaire developed specifically for patients with non-transfusion dependent beta thalassaemia (NTDT). It is an important study as there are no other disease-specific QOL assessment tools for NTDT patients, and it can potentially be used in other clinical trials as a standardized method of assessing the QOL impact of interventions in this group of patients.

I have a number of minor points for the authors to address:

1. In the abstract, the design of the study is described as "a retrospective study". The scores from the various assessment tools were captured prospectively during the course of the BEYOND trial, so why would this study be described as being retrospective in nature?

We thank the reviewer for pointing this out. We agree and have updated the study design in the abstract (page 3).

2. As the trial participants were from a number of geographic regions, were the QOL instruments available in languages other than English?

The instruments were indeed translated and linguistically validated into multiple languages based on the geographic regions of the study sites and administered to participants in their preferred language. This information was added to the Methods section, under PRO assessments (page 8).

3. Was any qualitative data captured regarding the ease of use of the NTDT-PRO questionnaire as compared to the other QOL instruments, especially the longer instruments such as the SF-36v2 and FACIT-F?

These data were not collected in this study; however, qualitative work previously conducted established that the instrument was well understood by patients, covered all relevant symptoms, and included appropriate response options (Taher A, et al. Am J Hematol 2019;94:177–83). Additionally, compliance quotients (number of intent-to-treat [ITT] participants completing the questionnaire at a given week divided by the number of eligible ITT participants at a given week) remained relatively high in this study for all instruments through week 24 (>80%), indicating that most participants were able to complete the questionnaire. The non-transfusion-dependent β-thalassaemia patient-reported outcome (NTDT-PRO) questionnaire has only 6 items, thus the patient burden to complete the NTDT-PRO is expected to be much less than that for the 36-Item Short Form version 2 (SF-36v2) questionnaire.

4. Will the authors be providing a link or information as to how the actual NTDT-PRO instrument can be accessed?

The development of the NTDT-PRO instrument has been previously reported (please see Taher A, et al. Am J Hematol 2019;94:171–6). A copy of the instrument is included in the supplementary materials for that publication. Bristol Myers Squibb is in the process of establishing a method for investigators to request use of the instrument.

Reviewer: 2
Dr Mohsen Elalfy, Ain Shams University

Comments to the Author:

It's a good retrospective had evaluated the psychometric properties were using data from the BEYOND trial (NCT03342404). Scores were moderately correlated with changes in hemoglobin levels; Why moderate correlation? and please comment the cut-off Hb level with high degree of correlation (all were < 100gm/L) please comment on variability of PRO in relation to different scores?)

We thank the reviewer for the feedback provided. Moderate correlations (≥0.3 to <0.7) were observed between the change in Tiredness/Weakness (T/W) and Shortness of Breath (SoB) scores and changes in all potential anchors, with the exception of the Patient Global Impression of Severity (PGI-S). This is expected as each of the potential anchors measure a different construct than the T/W or SoB scores, and thus variability may be different between the T/W or SoB change scores and the potential anchors (PRO or haemoglobin level). Correlations >0.3 are generally considered to indicate responsiveness (Revicki D, et al. J Clin Epidemiol 2008;61:102–9); thus, the correlations with haemoglobin levels (ranging from –0.30 to –0.38) are acceptable. As the BEYOND inclusion criteria required participants to have a baseline haemoglobin level ≤10 g/dL, we cannot comment on how the correlation would have been for participants with baseline values >10 g/dL. However, no cut-off for haemoglobin values were used in this psychometric analysis; it was treated as a continuous variable. If this response does not address the comment, we ask that the reviewer kindly clarify the question.

Authors mentioned the Strengths and limitations of this study

• Strengths of this study include use of well-validated PRO instruments such as PGI-S, PGI-C, SF-36v2®, and FACIT-F. However validation was done for this particular study?

Validation of these instruments has been conducted as part of this analysis, but they have been well validated in other disease areas and were used as external anchors to validate the NTDT-PRO following standard approaches for psychometric evaluation.

• The data used in this analysis were from a phase 2 interventional study with participants from multiple geographic regions and spanning a range of NTDT symptom severities. Inter-symptoms variations could be added in a paragraph in the discussion

We kindly ask the reviewer to clarify the comment/concern about inter-symptom variations.

• Patients with severe DRAEs were different in PRO scores or not; please add in results section and discussion

The NTDT-PRO is designed to assess the severity of NTDT-related disease symptoms (i.e., T/W and SoB) rather than treatment-related symptoms. Thus, comparing NTDT-PRO scores between groups

specified as those who had a severe drug-related adverse event (DRAE) vs. those who did not have a severe DRAE is unlikely to yield interpretable differences. Additionally, no PRO questionnaires assessing DRAE or treatment-related symptoms were administered in the trial; therefore, no comparisons are possible.

• The use of data from an interventional study allowed for changes in symptom severity to be observed, validating NTDT-PRO's sensitivity to identify longitudinal changes in symptoms.

We thank the reviewer for pointing this out. The Discussion was edited per the reviewer comment (page 25).

• Given that NTDT is a rare disease, limitations of the present study include the reduced sample size (Please confirm that no patients in BEYOND were excluded from the survey) being a retrospective study for typical psychometric evaluations.

This analysis was done on the ITT population from the phase 2 BEYOND study, and all ITT participants completed the NTDT-PRO at baseline (i.e., no patients in BEYOND were excluded). Completion of the NTDT-PRO remained high (>87.3%) through week 24 among those ITT participants still being followed. The study design for this analysis was amended in the manuscript.

• Cut-off values used to define different levels of improvement in the responsiveness analysis are not well established and were based on certain assumptions. Please clarify.

Cut-off values are addressed in the Discussion section (paragraph describing limitations, page 25). Although assumptions on the values used to categorize improvement and worsening levels were required to conduct the ANCOVA analysis, the moderate to strong correlations between the NTDT-PRO domain change scores and the anchors as part of the responsiveness analysis provide additional evidence to demonstrate that the NTDT-PRO is sensitive to change.

Please stress on PROs differences in 1st 12 weeks vs 13-24; specially more missing dates of reporting in this period and number of patients had dropped from 149 tat baseline to 110 at 24 weeks Any special scores are related to reported AEs according to their severity.

Although the number of participants who completed the NTDT-PRO dropped from 145 at baseline to 110 at week 24, the majority of those missing NTDT-PRO scores at week 24 were due to discontinuation from the study. Completion rates among those in the study remained high through week 24 (>87.3%, Figure S1).

Reviewer: 3
Prof. Anuja Premawardhena, University of Kelaniya, Colombo North Teaching Hospital

Comments to the Author:

NTDT Pro is clearly a good tool in the armamentarium of the researchers in the field of Thalassaemia and the paper effectively describes its validity against the other more well established tools. The daily entry of details on one hand. makes it accurate but on the other less. practical in the absence of Apps. which patients. carry with them, or are sufficiently motivated to use. None of these comments are however strictly relevant to this paper!

We thank the reviewer for the feedback received. We clarified in the Discussion section that the instrument was developed for research purposes/analysis of trial endpoints and that its practical use in the clinical setting still warrants further evaluation (page 25).

Reviewer: 4
Dr S Vitoratou, King's College London

Comments to the Author:

This is a very well written work and very interesting indeed. With respect to the analysis it was conducted appropriately. However the authors have conducted only assessments of reliability (internal consistency and stability) assuming that the factor structure is unidimensional. The sample size is not enough for CFA (ideally we would want this to be above 200) so i will not suggest to the authors to conduct such analysis. It is important however in my opinion to replace the title with investigations for the reliability and responsiveness as opposed to psychometric evaluation to reflect that the latter is only partially conducted. It should be mentioned clearly in the manuscript what is assumed for dimensionality and why and also state as a limitation that no CFA is presented in these data.

We thank the reviewer for the feedback received. Although a confirmatory factor analysis (CFA) was not performed as part of this analysis, an exploratory factor analysis was performed in a previous evaluation of the NTDT-PRO, which supported the grouping of tiredness and weakness into one domain (Taher A, et al. Am J Hematol 2019;94:177–83), and we have added this into the manuscript, under Statistical analyses: Reliability (page 12). We feel that the current analysis comprises a psychometric evaluation as it included an assessment of construct validity (including known-groups validity and convergent-divergent validity as required by the FDA [FDA Draft COA Guidance 3; https://www.fda.gov/media/159500/download]), reliability (internal consistency and test-retest), and responsiveness (i.e., sensitivity to change, a special kind of assessment validity). The CFA may be conducted as part of a psychometric evaluation, but it is not a requirement and not typically conducted in studies using “psychometric evaluation” in the title. Thus, we have left the title unchanged.

VERSION 2 – REVIEW

REVIEWER	Lam, Joyce KK Women's and Children's Hospital
REVIEW RETURNED	27-Jan-2023

GENERAL COMMENTS	I thank the authors for addressing my comments and making the necessary changes to the manuscript.

VERSION 2 – AUTHOR RESPONSE