



EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

15 September 2022
EMADOC-1700519818-1033762
Executive Director

Letter of Support for TUMMY-UC – a Patient- and Observer-Reported Outcome for Paediatric Ulcerative Colitis (UC)

On 18/11/2021, the Applicant Shaare Zedek Medical Center requested a Qualification Opinion for their TUMMY-UC - Patient- and Observer- Reported Outcome for Paediatric ulcerative colitis (UC) pursuant to Article 57(1)(n) of Regulation (EC) 726/2004 of the European Parliament and of the Council.

TUMMY-UC - Patient- and Observer-Reported Outcome is intended for use in clinical trials investigating interventions to treat paediatric ulcerative colitis (UC).

A discussion meeting with the Applicant took place on 04/04/2022.

On 10/06/2022, the SAWP agreed on the advice to be given to the Applicant.

On 23/06/2022, the CHMP adopted the advice to be given to the Applicant.

The CHMP considered that uncertainties associated with the available evidence do not currently allow a positive conclusion with respect to a Qualification Opinion.

However, based on the previous and ongoing efforts and the submitted evidence, the SAWP agreed to provide a letter of support to facilitate and encourage further evidence generation which may support a future Qualification Opinion.

Potential sponsors of paediatric UC clinical trials are encouraged to incorporate the TUMMY-UC into their study protocols and development programmes to enable further evaluation and validation of this clinical outcome assessment with a view to supporting a potential future Qualification Opinion.

Official address Domenico Scarlattilaan 6 • 1083 HS Amsterdam • The Netherlands

Address for visits and deliveries Refer to www.ema.europa.eu/how-to-find-us

Send us a question Go to www.ema.europa.eu/contact **Telephone** +31 (0)88 781 6000

An agency of the European Union



The proposed Context-of-Use (COU)

The TUMMY-UC was developed as a tool to score disease-related signs and symptoms in paediatric UC, in turn reflecting disease-related physical state.

The proposed context of use (CoU) is to *"use TUMMY-UC as an outcome measure in paediatric UC (2-18yoa) [studies] together with an objective measure of endoscopic healing"*.

There are two versions of the TUMMY-UC: one for children 8-18 years of age (Patient Reported Outcome; PRO) and one for caregivers of children 2-7 years of age (Observer Reported Outcome; ObsRO) in which behaviours indicative of subjective concepts are scored, rather than the concept directly (e.g., caregivers are asked to score behaviours associated with pain rather than themselves scoring the degree of pain).

Summary of development and validation of the TUMMY-UC

The Applicant followed a stepwise development which is described for its content and results below.

Stage 1: Qualitative - concept elicitation interviews to explore items important to children with UC and determining their rank-order

During this initial stage of development of the tool, interviews were conducted in Israel, England, Scotland, Ireland, Canada and the USA intended to ensure cultural diversity. Data were collected directly from children and caregivers in the younger age group according to standardized methodology. A total of 79 interviews were performed, including 30 paired child-caregiver interviews. Item selection was based on importance and frequency of endorsement by the patients and caregivers.

Item reduction was based on importance and frequency ratings, as well as expert input, and removal of some of the items was based on low ranking and frequency of being mentioned. The following symptoms were finally selected: abdominal pain, rectal bleeding, stool frequency, stool consistency, general well-being/fatigue, urgency, and nocturnal stools. In addition, the item "rectal bleeding" was divided into "bleeding-amount" and "bleeding-frequency" to avoid asking two questions in one item.

The adequacy of a self-administered questionnaire in children was evaluated, and it was concluded that children younger than 8 years did not understand the questions sufficiently to reliably score the items. Consequently, for this age group an observer-reported TUMMY-UC was developed based on questionnaires with the caregivers.

The interviews also included questions aimed to determine the adequate recall period of the TUMMY-UC. Based on the results, it was decided to develop the TUMMY-UC for a 24-hour recall-period with an option to collate scores of more than one day in the final score.

Overlap in domains/items with the previously qualified PUCAI score (see EMA/CHMP/SAWP/801872/2015 with a relatively narrow CoU that is more endoscopy proxy-centred than the symptom/sign-centred CoU pursued for TUMMY) is large. Whereas some domains of TUMMY-UC offer more granularity (in terms of items queried and/or grading), are differently worded/conceptualised, and differ in weighting, such that sum scores are not immediately comparable, the content is almost identical between PUCAI and TUMMY, which is reassuring.

Stage 2: Cognitive debriefing interviews to finalize the TUMMY-UC design

Cognitive debriefing interviews with 107 children and 48 caregivers were conducted to establish how the included items are understood by patients and, when applicable, their caregivers, to determine the exact wording of the items. Recall period, potential weighting, and rank ordering of the items were also further explored in this stage. Another aim of this stage was to determine the conceptual equivalence of the included items between children 8-12 years of age and their caregivers to provide a scientific

basis for integration of the items completed by patients (children 8 years of age and older) and observers (caregivers of younger children). Understanding of all interview questions was reported as good by 99% of children while 85% correctly understood the time frame ("recall period" of 24 hours). Based on the rank order of the children's (and caregivers') scoring of importance and the univariate correlation matrix, the Applicant decided that three top items (amount of rectal bleeding, frequency of rectal bleeding, and stool frequency) should be weighted 50% higher than the other items (governed by the number of response options to allow for round numbers). Based on the advice given on this matter earlier, the adequacy of weighting of items was, however, to be further explored in the later development phases, along with the final algorithm for scoring.

To justify the similar content and structure of the PRO and ObsRO, 44 interviews were performed with 22 children 8-12 years of age and their 22 caregivers in parallel. The median scores of the TUMMY-UC PRO and the TUMMY-UC ObsRO were almost identical with good correlation. With respect to disease activity categories, 100% of the total ObsRO scores were within the same disease activity category as their corresponding PRO scores. However, scepticism remained about the conclusion to measuring similar concepts in young children below the age of 8 based on results of interviews conducted with children of the higher age group. A direct comparison of results of the psychometric evaluation for the PRO (in older children) and ObsRO (in younger children) was considered more adequate at that time.

For these two phases of development, it was concluded that TUMMY-UC domain/item coverage is comprehensive and meaningful, the format appears acceptable in terms of comprehensibility and burden to respondents, and content validity can thus overall be supported. It is noted that higher education among respondents was overrepresented during concept elicitation and cognitive debriefing. In this regard, the requested demonstration that educational level is not a relevant factor for the concept elicitation and cognitive debriefing for PRO/ObsRO development in general (and for the TUMMY-UC) remains to be provided. It is, however, acknowledged that overrepresentation of higher education reflects the epidemiology of paediatric UC.

Stage 3: Psychometric evaluation of validity, reliability and responsiveness

For this validation step, an observational, prospective multicentre study was conducted with the objectives to assess the psychometric and clinimetric properties of the TUMMY-UC including validity, reliability and responsiveness.

A total of 84 patients were included, of whom 52 underwent colonoscopy and 32 provided stool for faecal calprotectin. Analyses were performed both for one-day TUMMY-UC score and for a two-day average score.

Reliability was evaluated using Intraclass Correlation Coefficient (ICC), with a two-way mixed model reporting average measures with the corresponding 95% CIs. ICCs at or slightly above 0.9 were observed at the three time points (days 1-2, 11-12 and 25-26) (ICC range 0.87-0.95). Next, the reliability of the repeated visits from days 3-4 to days 11-12 was analysed in those judged to be stable, again with similar results for reliability (ICC range 0.89-0.94).

For construct validity, correlation coefficients (Pearson or Spearman, as appropriate) were calculated. The correlation between the TUMMY-UC and the UC endoscopic index of severity (UCEIS) was found to be around 0.7 (rho range 0.68-0.71). The correlation between the TUMMY-UC and calprotectin was 0.43 with similar correlations observed with CRP and albumin. The correlation with global assessment of disease activity as measured by patients and caregivers and the physicians ranged from 0.71 to 0.77). As already mentioned, the TUMMY-UC had a correlation with the PUCAI of almost 0.8 ($r=0.78-0.80$). The evaluation of the two-day against the one-day version of the tool did not show relevant differences.

For discriminative validity, the area under the ROC curve to differentiate clinical remission (defined as PUCAI<10) from active disease showed an AUROC of 0.97 (95%CI 0.93-0.99). The area under the ROC curve to differentiate endoscopic healing from active endoscopic colitis was determined at 0.85 (95%CI 0.74-0.96).

For the evaluation of responsiveness, data from 70 children who had evaluations at two time points were used. The TUMMY-UC showed responsiveness to change and differentiated very well children who improved, worsened or whose disease remained unchanged, whether defined by a change in at least 20 points in the PUCAI score or as a change of at least two categories in the Likert scale of the patient global assessment (AUROC range 0.88-0.94). The best cut-off of the weighted Δ TUMMY-UC to define response was a change of at least 10 points (AUROC 0.88-0.93). Similarly, the responsiveness statistics showed similar responsiveness of the TUMMY-UC with all three approaches (distributional, anchor-based and correlational).

To validate the conceptual equivalence between the PRO and ObsRO versions determined in stage 2 on 44 interviews, a further 36 interviews of children 8-12 years of age (PRO) and caregivers (obsRO) were evaluated. The agreement between the two assessments was determined with an ICC of 0.92 (95%CI 0.74-0.98). The mean and median values of the PRO and ObsRO TUMMY-UC versions were similar. In addition, similar psychometric properties were demonstrated for both versions for reliability, validity and responsiveness.

However, the submitted 'pivotal' validation data were considered rather limited, including the additional information made available for the Discussion Meeting. Overall, data presentation was difficult to base a structured assessment on, and a final study protocol or statistical analysis plan (SAP) compliant with usual regulatory standards, were not provided. A structured and comprehensive clinical study report (CSR) complying with usual regulatory standards to support this submission was also not provided and the study comprised an uncontrolled cohort only.

Generally, a clear separation of exploratory and confirmatory analyses was not apparent. This relates to weighting/scoring schemes, the choice and comparison of/to reference/anchor constructs and the strategy for missing data. Therefore, it was difficult to fully concur with the Applicants' conclusions.

Another limitation identified was the small study sample which appears mainly feasibility-driven considering the restriction of the development to a limited number of academic centres. This is further aggravated by the fact that two TUMMY versions (with two distinct target age cohorts) are concerned, and relevant and potentially relevant disease strata (at baseline and EOS) need to be considered from a perspective of representativeness. Only 84 patients (of 100 screened in total) contribute validation data with 13 subjects in the 2-7 years cohort, limiting the ability to assess ObsRO performance with high precision. In addition, in terms of clinical presentation, severe UC appears underrepresented (given that a total of ~ 20 subjects distribute across moderate & severe categories as per PUCAI, with actual distribution between the two remaining unknown).

Partly due to the limited time frame of the study (and with limited knowledge about study treatment and changes thereto), it appears that the vast majority of subjects remained stable (as per patient global assessment). Few patients improved or worsened during study which limits the precision achievable for respective analyses.

In support of the congruence of the PRO and ObsRO tools, the Applicant has presented a direct comparison of the PRO and ObsRO results for responsiveness and MID determination, which has demonstrated no relevant difference between the two using the PGA as anchor but show some divergence for the PUCAI, as well as for the precision of the estimates. An evaluation of responsiveness/minimal clinical important difference (MCID) determination could not be performed

with respect to (the anchor) Quality of Life, as no such scales were used in the patients aged 7 and below. Overall, the conclusions on similar properties of both tools appear to be only preliminary due to the small number of patients in the younger age group.

Conclusion

The TUMMY-UC has shown a fully acceptable content validity, and has shown promising results for relevant psychometric features, such as validity, reliability and responsiveness. Limited data also support conceptual equivalence of the PRO and the ObsRO versions of the TUMMY-UC.

Overall, the TUMMY-UC, with its fixed, weighted scoring rule, the severity categories and related numerical thresholds, the proposed MCID and item replacement rules, is recommended to be confirmed further for its validity in both its versions according to a-priori devised criteria in future, preferably randomized, controlled treatment trials. Preferably, this would be done with broad coverage in terms of demographics and disease characteristics, with sufficiently long follow-up and using established SoC and/or investigational treatment for which relevant changes of the disease state in established outcomes can reasonably be expected.

Potential sponsors of such studies are encouraged to incorporate the TUMMY-UC into their study protocols and development programmes to enable further evaluation and validation of this clinical outcome assessment with a view to supporting a potential future Qualification Opinion.

1. Marcovitch L, Nissan A, Mack D, Otley A, Hussey S, McLean B, et al. Item Generation and Reduction Toward Developing a Patient-reported Outcome for Pediatric Ulcerative Colitis (TUMMY-UC). *J Pediatr Gastroenterol Nutr.* 2017;64(3):373-7.
2. Marcovitch L, Focht G, Horesh A, Shosberger A, Hyams J, Bousvaros A, et al. Agreement on Symptoms Between Children With Ulcerative Colitis and Their Caregivers: Towards Developing the TUMMY-UC. *J Pediatr Gastroenterol Nutr.* 2021;73(2):e35-e8.