

# Assessing patient-reported adverse events in oncology – a review of the PRO-CTCAE instrument

## Erfassung von unerwünschten Ereignissen aus Patientensicht in der Onkologie – ein Review des PRO-CTCAE-Instruments

### Abstract

**Aim:** Patient reporting of adverse events (AE) is still rare in oncological drug development. It should however be a crucial part of clinical studies and medical decision making. The PRO-CTCAE instrument has been generated from observer based AE assessment to represent the patients' view in 124 items. This review examines the psychometric quality of the PRO CTCAE as tested in selected cancer patient groups and published in core papers.

**Method:** A literature review identified 3 league papers describing the original development of the PRO-CTCAE. Using the COSMIN quality criteria for patient-reported outcome measures, the psychometric performance of the PRO-CTCAE was systematically examined.

**Results:** Sufficient information on COSMIN quality standards were only present in 2 of the 10 criteria with EORTC QLQ-C30 used as a validation anchor, providing only limited evidence for reliability or validity. For 5 out of the 10 criteria no information at all is available.

**Conclusion:** PRO-CTCAE is a useful list of items which has been psychometrically tested on individual item level, but not on a scale level. The development of a measurement model and consecutive psychometric testing is recommended to provide information on test-theoretical performance on the aggregate level including scoring and comparison to reference populations. This would transform the PRO-CTCAE items from a list to a tool which can be used in patient-reported AE assessments within and across oncological trials. Future research should focus on explaining potential differences between physicians and patient ratings, on cross-cultural comparability of the tool, and on the impact of PRO-CTCAE as a unique instrument for patient reported outcomes in clinical research and medical decision making.

**Keywords:** PRO-CTCAE, validation, COSMIN criteria, psychometric quality

### Zusammenfassung

**Ziel:** Eine Erfassung der patientenberichteten unerwünschten Ereignisse (adverse events, AE) ist in der Arzneimittelentwicklung von Onkologika noch immer selten. Die Sicht des Patienten wird ein immer wichtigerer, bisher aber kaum beachteter Aspekt klinischer Information. Das Instrument PRO-CTCAE wurde aus den arztberichteten AE-Erhebungen entwickelt und stellt die Patientensicht in 124 Aspekten dar. Dieser Review untersucht die psychometrischen Eigenschaften des PRO-CTCAE, so wie er in ausgewählten Krebspatienten-Populationen untersucht wurde, die in die zentralen Publikationen zur Validierung Eingang fanden.

**Methoden:** Über eine Literaturrecherche wurden 3 zentrale Publikationen identifiziert, die die Entwicklung des PRO-CTCAE beschreiben. Mittels der COSMIN-Qualitätskriterien für patientenberichtete Fragebögen wurden die psychometrischen Eigenschaften und die Qualität des PRO-CTCAE (basierend auf dem verwendeten EORTC QLQ-C30 als Anker für die Validierung) systematisch untersucht.

Dietrich Knoerzer<sup>1</sup>

Martina Kron<sup>2</sup>

Armin Schöler<sup>3</sup>

Susanne Huschens<sup>4</sup>

Monika Bullinger<sup>5</sup>

1 Roche Pharma AG,  
Department of Biometrics &  
Epidemiology, Grenzach,  
Germany

2 Abbvie Deutschland GmbH &  
Co KG, Ludwigshafen,  
Germany

3 MorphoSys AG, Planegg,  
Germany

4 Janssen-Cilag GmbH, Neuss,  
Germany

5 Department for Medical  
Psychology, Medical Center  
Hamburg-Eppendorf,  
Hamburg, Germany

**Ergebnisse:** Ausreichende Information zu COSMIN-Qualitätsstandards waren nur für 2 der 10 Kriterien verfügbar, so dass nur eine beschränkte Evidenz im Hinblick auf die Belastbarkeit der Validität vorhanden ist. Für 5 der 10 Kriterien ist überhaupt keine Information verfügbar.

**Schlussfolgerung:** Der PRO-CTCAE ist eine nützliche Liste von Aspekten, die psychometrisch auf einer individuellen Aspekt-Ebene getestet wurden, aber nicht auf Skalenebene. Um die testtheoretische Leistung auf einer aggregierten Ebene zu beschreiben, wird die Entwicklung eines Messmodells empfohlen, sowie die nachfolgende psychometrische Testung desselben. Dies beinhaltet die Wertung und den Vergleich zu Referenzpopulationen. Damit würde der PRO-CTCAE von einer Liste zu einem Werkzeug weiterentwickelt, mit dem patientenberichtete AE-Bewertungen innerhalb und zwischen onkologischen Studien bewertet werden können. Zukünftige Forschung sollte sich mit der Erklärung möglicher Unterschiede zwischen Arzt und Patientenbeurteilungen, mit kulturübergreifenden Vergleichen und mit den Konsequenzen von PRO-CTCAE als alleinstehendes Instrument für die Erfassung von patientenberichteten unerwünschten Ereignissen in der klinischen Forschung und medizinischen Entscheidungsfindung befassen.

**Schlüsselwörter:** PRO-CTCAE, Validierung, COSMIN-Kriterien, psychometrische Qualität

## Introduction

In oncological clinical studies, adverse events (AE) are regularly recorded and documented by the investigator or treating physician using medical reporting systems such as Medical Dictionary for Regulatory Activities (MedDRA) [25]. Severity assessment of AEs is often based on Common Terminology Criteria for Adverse Events (CTCAE) of the National Cancer Institute (NCI) [11]. Both approaches are focusing on physician's judgement of adverse events.

There is, however, a common understanding that the patient perspective is of utmost importance in the evaluation of new treatments. Appropriate, meaningful, and valid instruments should be used to assess the benefits and risks of a new drug by patient reporting, and these should comply with internationally consented quality criteria for patient-reported outcomes assessment, such as the COSMIN guideline [13].

Until now, the patient's perspective has been assessed primarily in terms of morbidity and health-related quality of life. For safety aspects, e.g. adverse events, mainly the physician's perspective was documented so far, while the directly reported patient's perspective has not been established as a standard [7].

The patient-reported outcomes measurement system PRO-CTCAE was developed as an extension of the CTCAE assessment introduced by NCI as a tool to obtain patient reports on adverse events [8]. CTCAE and PRO-CTCAE do not exactly contain identical categories, it is advisable to compare and contrast the two scales to be informed about symptoms from different perspectives. It is crucial here that patients AND physicians fill in the forms carefully and completely [7].

In addition to the reported poor completion rate of the CTCAE items by physicians, it is debatable whether it is

an appropriate source for PRO-CTCAE item selection and if the EORTC QLQ-C30 is a well-chosen anchor. Finally, potential discrepancies between CTCAE and PRO-CTCAE should not be attributed to PRO-CTCAE alone. If data are missing or in case of a low association it is not possible to detect whether discrepancies are (i) due to different underlying concepts, (ii) due to incomplete answers of patients or (iii) physicians, (iv) due to both, physicians' and patients' incomplete answers or (v) due to actual differences in the perception of adverse events. There is also a different time perspective between the two instruments (7 days for PRO-CTCAE, no restriction for CTCAE [3]).

The current publication aims to provide a reflection of the information given by publications of the original validation of the PRO-CTCAE instrument, using the COSMIN quality criteria for a structured assessment of its measurement approach and psychometric characteristics including validity, reliability, and responsiveness as examined in a large, heterogeneous US sample of patients undergoing cancer treatment [9], [15], [20].

The paper thus examines whether – on the basis of the original papers introducing the instrument – important psychometric characteristics as identified with the COSMIN quality criteria are fulfilled by the PRO-CTCAE instrument.

## Characteristics of the PRO-CTCAE

The standard approach for documenting symptomatic AEs in cancer clinical trials involves investigator reporting using National Cancer Institute's Common Terminology Criteria for Adverse Events (CTCAE) for severity. Because this approach relies on observer data, it may fail to detect symptoms relevant to and noted by the patient, thus po-

tentially underreporting symptomatic AEs [33]. Acknowledging the value of patient reports, the PRO-CTCAE was developed from the criteria catalog of the CTCAE as a tool to include the patients' view by capturing the patient-reported adverse events. Newer developments have also established an IT-tool to ease input and review by physicians and patients [6].

In a first step, the existing 790 CTCAEs were analyzed to identify those events that are suitable for self-reporting by the patient, resulting in 78 items, i.e. one tenth of all CTCAE items, [8]. To be amenable to patient reporting, each of the 78 items is to be evaluated regarding presence or absence; and if present, up to four different attributes may need to be assessed, namely frequency, severity, interference with activities related to each AE and amount on a 5-point Likert scale [8], [15] (incl. supplements), [29]. Since items are assessed using different numbers of attributes, there are a total of 124 individual questions in the PRO-CTCAE item pool [8].

Not all AEs in this item pool are relevant to every disease or treatment context, and the large number of items in the PRO-CTCAE library may make it impractical to administer all items to all patients. Therefore, a reduction of the number of items is suggested: "A limitation to about 28 items is doable without overburdening patients" [8]. Other than MedDRA and CTCAE which both cover all potential AEs, PRO-CTCAE assessment does not require every item and attribute to be investigated in each clinical study.

The PRO-CTCAE is now available for patient reporting as a list of items to choose from for a given oncological study. The question is, however, if this tool can be considered not only as an item pool, but also a psychometrically sound instrument which fulfills test-theoretical quality criteria as a patient-reported outcomes measure. To evaluate the development and the validation of a PRO instrument, sets of criteria have been suggested, the most comprehensive of which are the *COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN)* criteria. The measurement properties described in the COSMIN were used to evaluate the original PRO-CTCAE validation approach [13], [27], [30], [32].

## Methods

### COSMIN criteria

The COSMIN steering committee developed and published a guideline for systematic reviews of patient-reported outcome measures including a list of 11 criteria which are essential for assessing the methodological quality of a PRO measure and its measurement properties [27].

Ten of these measurement properties were considered relevant for this evaluation of the PRO-CTCAE validation, the exception being face validity which can be subsumed under content validity based on the COSMIN list.

These ten properties, either defined as domain or measurement property, are listed below according to the numbering in the original COSMIN publication [27]. The publications describing the original validation approach of the PRO-CTCAE are used to examine whether these properties have been addressed:

- **Reliability**, containing the measurement properties internal consistency (1), reliability (2), and measurement error (3)
- **Validity**, containing the measurement properties content validity (4), construct validity (5), structural validity (6), cross-cultural validity (7) and criterion validity (8)
- **Responsiveness**, containing the measurement property responsiveness (9)
- **Interpretability** (10)

### Literature review

A systematic literature search was performed using the data bases EMBASE, MEDLINE and Cochrane Library. The searches include combinations and synonyms of the following terms: PRO-CTCAE, PRO, CTCAE, self-report, validity, reliability (see Appendix (Attachment 1) for the detailed search strategy for each library, access dates: Embase, Medline: May, 24<sup>th</sup> 2018, Cochrane: June 1<sup>st</sup>, 2018).

The systematic literature search resulted in 152 citations eligible for screening, which were inspected and consented by two independent reviewers. Screened publications were selected according to the criteria (a) methodological aspects covered, (b) validation aspects covered, (c) referring to the empirical patient data of the NCI data set used originally for validation [8], and (d) original data reflected and primary validation approach represented. This resulted in 3 publications, namely Hay et al. 2014 [20], Dueck et al. 2015 [15] (incl. resp. supplement), and Bennett et al. 2016 [9] (Figure 1).

In summary the assessment was as follows (Table 1):

a) Number of publications: After literature research 3 articles remained describing the original validation, which is too small a number for statistical analysis.

b) Analyzing measurement properties: The analysis was a three-step approach per property.

1. Is any information available on the measurement property?
2. Which type of analysis is the basis for information regarding the respective instrument property?
3. What is the conclusion drawn from the results?

In step 1 this included a binary decision (information available y/n), whereas in step 2 a methodological examination according to the COSMIN evidence levels was performed. Finally, step 3 is a contextualization of the numerical results.

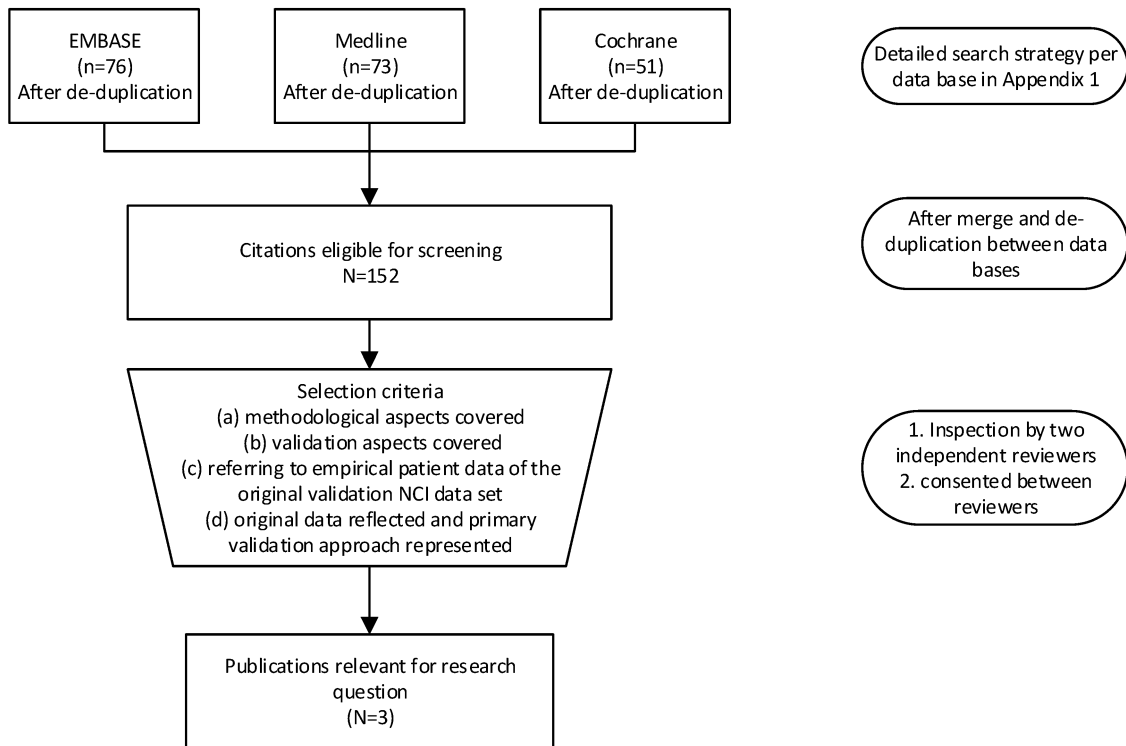


Figure 1: Flow chart of literature research for relevant papers (search criteria in appendix)

Table 1: Measurement properties and their evaluation based on original publications

Criterion number*	Properties addressed	Publication
1	internal consistency	–
2	reliability	Dueck et al. 2015 [15] Bennett et al. 2016 [9]
3	measurement error	–
4	content validity	Hay et al. 2014 [20]
5	(hypothesis testing for) construct validity	Dueck et al. 2015 [15]
6	structural validity	–
7	cross-cultural validity / measurement invariance	Dueck et al. 2015 [15]
8	criterion validity	Dueck et al. 2015 [15]
9	responsiveness	Dueck et al. 2015 [15]
10	interpretability	–

\* Criterion numbering according to Mokking et al. [27]

## Results

The 3 publications (and their supplements) were screened for information on the 10 COSMIN criteria, described in Table 1.

The degree of completeness of information on each of the criteria was rated along 4 categories: not available, insufficient, partly sufficient, and sufficient.

### Criterion 1: Internal consistency

Internal consistency is defined as consistency of responses to items of the same multi-item scale, where the items are intended to capture the same construct (or complementary interrelated aspects of it). Internal con-

sistency is measured through Cronbach's alpha coefficient [12].

No information on internal consistency was found in publications for the original US version of the instrument. Due to lack of information neither a review nor an evaluation is possible.

The overall rating of information on the measurement error of the PRO-CTCAE is 'not available'.

### Criterion 2: Reliability

Reliability is assessed through determination if a scale of measurement instrument yields reproducible and consistent results, e.g. test-retest reliability (reproducibility). During the assessment the patients are asked to

complete the same quality of life questionnaire on several occasions [17]. Test-retest reliability for PRO-CTCAE was assessed by Dueck and colleagues [15].

Items for validation were selected without providing a rationale: It is unclear how and why 49 out of 124 items [15] were chosen. The reported median Intraclass Correlation Coefficient (ICC) was 0.76, 13 out of 49 items had ICC values <0.7.

The ICC is the most commonly used method for assessing test-retest reliability with continuous data [17]. In general, an ICC of at least 0.90 is viewed as a high intraclass correlation while 0.70 is viewed as moderate. Since the magnitude of the ICC depends (among others) on the heterogeneity of the population, no interpretation of ICC values seems to be reasonably possible without further insight in the population, even if ICC values are above 0.7. However, such information was not provided.

The available information is not sufficient to conclude with reasonable certainty on the reliability of the PRO-CTCAE, both with regard to possible selection bias due to the non-transparent item selection and with regard to the heterogeneity of the population. Therefore, the level of information with regard to reliability is rated as 'insufficient'.

### Criterion 3: Measurement error

Measurement error is defined as the systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured. It can occur in (1) test-retest, (2) intra-rater and (3) inter-rater measurements [27].

No information was provided in any of the available identified publications.

The overall level of information on measurement error of the PRO-CTCAE is rated as is 'not available'.

### Criterion 4: Content validity

Three aspects of content validity are being distinguished and need to be addressed: (1) relevance (all items in a patient-reported outcome measure (PROM) should be relevant for the construct of interest within a specific population and context of use), (2) comprehensiveness (no key aspects of the construct should be missing) and (3) comprehensibility (the items should be understood by patients as intended).

Content validity was comprehensively assessed through cognitive interviews [20] which were conducted among patients undergoing chemotherapy or radiation therapy to evaluate comprehension, memory retrieval, judgment, and response mapping related to AE terms, attribute terms (regarding frequency, severity, or interference), response options, and recall period.

The above publication [20] provides sufficient information on content validity. Overall, the information for content validity of the PRO-CTCAE is rated 'sufficient'.

### Criterion 5: Construct validity

Construct validity is defined as the degree to which the scores of a PRO instrument are consistent with the hypothesis based on the assumption that the PRO validly measures the construct to be measured.

If comparison with a standard test is not available, construct validity is usually assessed via statistical testing of hypotheses regarding group differences in outcomes. This has not been reported by Dueck and colleagues [15]. The overall level of information on construct validity of the PRO-CTCAE is rated as is 'not available'.

### Criterion 6: Structural validity

Structural validity refers to the degree to which the scores of a PROM are an adequate reflection of the dimensionality of the construct to be measured and is usually assessed by factor analysis or Item Response Theory (IRT) / Rasch analysis [30].

Structural validity can be assessed via factor analytical or structural equation modelling approaches of the psychometric measurement model.

No information about structural validity of the PRO-CTCAE is available from any of the identified publications, thus neither an assessment nor an evaluation is possible.

The overall level of information on structural validity of the PRO-CTCAE is rated as is 'not available'.

### Criterion 7: Cross-cultural validity / measurement invariance

Questionnaires are not always translated appropriately before they are used in new temporal, cultural or linguistic settings. The results based on such instruments may therefore not accurately reflect what they are supposed to measure and need to be adapted cross-culturally including e. g. investigation of conceptual and item equivalence [18].

As described in Dueck and colleagues [15] the data base is built solely on US patients being able to comprehend English (with the risk of potential cultural bias), being able to make it to the waiting room (potential health status or financial bias), patients' willingness to participate and to comprehend electronic tools (potential educational bias). As for the generation of the original data base, no version other than English existed, there is a limitation to native speakers or those comprehending English. A selection bias regarding exclusion of Hispanic patients or other non-English speakers at the time of establishing the PRO-CTCAE cannot be ruled out [15].

Intercultural differences in AE reporting are well known, but rarely subject of research (e.g. [34]). For international studies using the PRO-CTCAE this would require an analysis on national levels and an inspection of differences across countries or languages (e.g. [10], [31], [28]). With respect to different reporting habits in different cultural or social segments of the US population, but also with

regard to populations outside the US, cross cultural aspects had at the time not been taken into account. Thus, the use of the PRO-CTCAE in other cultural populations can only be recommended after extensive cross cultural linguistic adaptation and validation. There are numerous translations/transfers of the PRO-CTCAE to other languages / cultural regions (currently 50 translations available, for current status refer to: <https://healthcaresdelivery.cancer.gov/pro-ctcae/>). These are not evaluated in this paper, as it refers to the original PRO-CTCAE validation.

The overall level of information on measurement error of the PRO-CTCAE is rated as 'partly sufficient'.

## Criterion 8: Criterion validity

Criterion validity involves, via correlation analysis, assessing an instrument against the true value, or against some other standard that is accepted as providing an indication of the true values for the measurements [17].

Dueck et al. [15] used patient-reported global health-related quality of life (EORTC QLQ-C30), clinician-reported ECOG performance status and other specific clinical variables (e.g. use of antiemetics or receipt chemotherapy containing taxane) as criteria in their study. The correlation coefficients between these anchors and PRO-CTCAE ranged between 0.0 (e.g. stretch marks (presence/absence)) and 0.74 (fatigue (interference with usual or daily activities)).

When looking at details of the statistical analysis, only a small subset of items is included. The statistical test for correlation was applied which is too sensitive to draw valid conclusions. A small p value does not provide evidence of a meaningful correlation. It needs to be taken into account when interpreting results that this just means that the correlation is different from 0. Generally well accepted thresholds to assess the strength of a correlation that could have been applied are e.g. 0 as no, <0.3 as negligible, <0.5 as low, <0.7 as moderate, <0.9 as high and 1 as perfect correlation [21].

Based on these thresholds, there was a weak correlation between EORTC QLQ-C30 symptom subscales and symptomatic AEs. This was observed also for items related to insomnia (0.48 (severity) and 0.52 (interference with usual or daily activities)), vomiting (0.40 (frequency) and 0.39 (severity)), and nausea (0.49 (frequency) and 0.51 (severity)) [15], which are very similar in both instruments, even though the scaling of EORTC QLQ-C30 and PRO-CTCAE is almost identical (the only difference being that the ECOG scale of performance status has one intermediate category more (EORTC QLQ-C30 has 4 whereas PRO-CTCAE and ECOG scale of performance status have 5 categories)).

Overall, the level of information on criterion validity of the PRO-CTCAE as an instrument (totality of items) is rated as 'insufficient' given the large number of items with 'no' to 'low' correlation.

## Criterion 9: Responsiveness

Item responsiveness examines whether changes over time are captured by the instrument [17].

As reported in Dueck et al. [15] the responsiveness of items was investigated by comparing change from first to second visit in 27 PRO-CTCAE items selected a priori. Since the analyses are limited to a subset of items with a possible selection bias, evidence for responsiveness still is weak.

Overall, the level of information on responsiveness of the PRO-CTCAE is rated as 'insufficient'.

## Criterion 10: Interpretability

Interpretability is defined as the degree to which one can assign qualitative meaning (that is, clinically or commonly understood connotations) to a PROM's quantitative scores or change in scores [30].

Interpretability is not considered a measurement property but an important characteristic of a measurement instrument.

There is no specific information available on interpretability in any of the publications.

Overall, the level of information on interpretability of the PRO-CTCAE is rated as 'not available'.

## Further aspects related to the validation

In addition to the ten criteria of the COSMIN statement, the following aspects of the data used for the validation by Dueck et al. [15] may represent further sources of bias in the validation of the PRO-CTCAE.

- Representativeness of sites: The number of sites is low, and the size of sites varies greatly (9 sites altogether, with patients ranging from n=9 to n=280 per site). The representativeness is unclear.
- Indication: Enrichment strategies are applied to increase patient numbers in (i) specific cancer types, (ii) specific symptomatic AEs and (iii) specific (higher) severity grades.
- Symptomatic AEs and items: By design the first 80% of patients had to answer 20 core symptomatic AEs (i.e. 26% of all) plus some specific ones. The last 20% of patients had to answer the remaining 58 symptomatic AEs.
- Study visit schedule: Three different kinds of schedules were followed to avoid additional visits for the validation effort. Visit frequency also differed between schedules but allowed for increasingly more distant visits. As the recall will decrease with longer time intervals, patient groups may differ with respect to results.
- Time course per patient: Post-baseline measurements are missing for a substantial amount of patients.

It is unclear how each enrichment strategy was implemented and whether or not the individual strategies interfered with the other enrichment strategies. These differing

strategies and their unknown interaction might have had a potential impact on data and results.

In general, it should be noted that the biases inherent in the results are not quantifiable as this would at least necessitate a check of the data base (i) in itself for internal validity, (ii) versus census data for external representativeness and (iii) for robustness of results with regard to missing (post-baseline) data.

The intention to generalize the results to (i) all cancer types, (ii) different classes of medication and (iii) a wide range of indications is a potential weakness of the validation. The overall results may mask different outcomes in different situations, i.e. different sets for AEs are needed in different cancer types treated with different medications based on different modes of action.

## Discussion

The NCI PRO-CTCAE tool offers an opportunity to capture the patients' perspective on symptomatic AEs during the development of new therapies. This is an important element for modern patient centered drug development, especially as discordances between physicians' and patients' ratings are common [2], [7], [35] and agreement often is poor [3]. Any discordances could be a) due to differences in patients' and clinicians' perception of AEs which would underline the importance to gather the patient's perspective, b) caused by different reporting schedules, or c) by the well-known difference between spontaneous reporting and being asked about AEs in general or even specific AEs.

It should be noted that the CTCAE, used to capture the physicians' view on AE, also has difficulties surrounding it. One can challenge if the CTCAE itself is a good source for selecting items for the PRO-CTCAE and whether it can serve as a reference with respect to reporting habits. As the league papers reporting the validation of the PRO-CTCAE used EORTC QLQ-C30 as an anchor, we are in our review bound to this anchor. Ideally and as a bonus feature, the categories of PRO-CTCAE should be comparable to the AE in content, wording and coding, so that patients' and physicians' views of defined symptoms can be taken into account. Newer developments of the PRO-CTCAE have also established an IT tool to ease input and review by physicians and patients [6].

The identified publications reflect intensive activities to validate the original PRO-CTCAE and reveal a solid basis for some of the quality criteria. **Content validity**, a key element for a PRO, in that patient reports on selected CTCAE items represent relevant aspects of patient experience, is present. However, assessing adverse events by a patient is limited to perceivable symptoms, therefore PRO-CTCAE cannot replace, but enrich, the classical safety profile provided by a physician. CTCAE and PRO-CTCAE do not exactly contain identical categories, it is advisable to compare and contrast the two scales to be informed about symptoms from different perspectives. It is crucial here that patients AND physicians fill in the forms carefully

and completely [7]. If data are missing, in case of a low association it is not possible to detect whether discrepancies are (i) due to different underlying concepts, (ii) due to incomplete answers of patients or (iii) physicians or (iv) due to both, physicians' and patients' incomplete answers.

Following the publication of the original PRO-CTCAE instrument, multiple linguistic validation studies are now available to account for the fact that original studies were solely based on US patients able to comprehend English [1], [4], [19], [22], [26]. The open question for **cross-cultural validity** to be further evaluated is the extent to which the data used for the validation might be subject to selection bias by excluding non-English speaking parts of the US population. Besides focusing on an English-speaking patient population, the validation of PRO-CTCAE is based on patients undergoing **chemotherapy or radiation therapy** only, including lung, head or neck, or breast cancer [15]. Considering recent developments in immunoncology, further evaluation on the validity of the PRO-CTCAE tool for checkpoint inhibitors might be valuable to assess whether the assessment of content validity remains unchanged when less toxic therapies are reviewed or other cancer types are investigated.

The threshold value ( $ICC \geq 0.7$ ) for **reliability** was not met for about a quarter of the selected 49 items. In addition, no information was provided on the selection of the 49 items evaluated for the ICC, so that a possible selection bias cannot be excluded. Therefore, reliability does not seem to be sufficiently demonstrated.

Two scales were used to assess the criterion validity of the PRO-CTCAE: besides the ECOG performance status, EORTC QLQ-C30 was used as an **anchor** for the validation of PRO-CTCAE. Despite similar or overlapping items and scoring method, only limited correlations were observed for symptoms like vomiting or nausea which are items within the EORTC QLQ-C30. This limits the **criterion validity**, especially as correlations of  $\leq 0.5$  should not be considered meaningful by the authors. The weak correlation between EORTC QLQ-C30 symptom subscales and symptomatic AEs surprises because of the similarities of questions and scaling of EORTC QLQ-C30 and PRO-CTCAE. Therefore, a higher correlation would have been expected. Even though EORTC QLQ-C30 and ECOG are well established instruments in oncology, they may be of limited use for validation purposes because they measure quality of life or performance status, whereas PRO-CTCAE is intended to measure adverse events. Thus, sufficient information on COSMIN quality standards were only present in 2 (criterion 4, 7) of the 10 criteria, providing only limited evidence for reliability or validity. For 5 (criterion 1, 3, 5, 6, 10) out of the 10 criteria no information at all is available.

In the context of PRO-CTCAE validation, there are additional aspects and concerns: Is the instrument EORTC QLQ-C30 suitable for such a validation? Are (solely) morbidity items of this instrument suitable? Are there sufficient items in the anchors to validate 78 items separately? Is the number of instruments sufficient?

This leads to the paradox that, on the one hand, developing a new instrument that would be very similar to an existing anchor would per se lead to a high construct validity but would not add value, while, on the other hand, developing a new instrument with a different focus would lead to only limited correlations with the anchor. The lack of a gold standard instrument is then, as in this case, a problem.

While **interpretability** seems to be given due to the chosen attributes of the selected items, no specific information is available in any publications on questions regarding distribution of values or potential ceiling effects, which could be of special interest if less or more toxic drugs are investigated. Limited information is available to assess the aspects of internal consistency, structural validity, measurement error, responsiveness, as well as study-related aspects such as site and indication selection. Further research in this direction is needed to be able to obtain a holistic view on PRO-CTCAE following the COSMIN approach.

The main problem identified in the review of the PRO-CTCAE is the nature of the instrument. In its current form, it is an item pool from which item lists can be created for use in specific clinical trials. However, it is unclear what the criteria for item selection should be, making the selected study-specific item list arbitrary.

There is no guidance on scoring, implying that items need to be scored individually. While the analysis of individual items and their change over time can be helpful in the clinical context, a scoring of scales according to a psychometrically-based scoring system is most economical and recommended especially for clinical research.

The lack of a scoring system is related to the *absence* of a theoretically founded measurement approach. Since there is no structured measurement model to identify individual domains and related elements, classical test theory cannot be applied. This leads to a lack of information regarding internal consistency and confirmatory factorial validity, which was identified in this paper. But probabilistic test theory was also not applied, which could have been used to characterize individual items, even though no domain or scale was postulated. The absence of a dimensional structure and associated scoring provides maximum flexibility at the expense of stable measurement indices, such as domain scores, that allow testing of differences between patient-reported adverse events of treatment strategies in clinical trials. Ultimately, the approach chosen, the group working on Patient Reported Outcomes Measurement Information System (PROMIS) could be useful in further developments of the PRO-CTCAE [16]. The basic approach in PROMIS is to identify domains of interest in patient-reported outcomes from qualitative content analyses (such as pain or emotional well-being) and to identify a theoretically indefinite item pool including items with the highest likelihood of providing valid and reliable information of a patients via computer assisted testing.

In reviewing the PRO-CTCAE, it should be noted that for the present work, only the original publications were se-

lected from the literature, which at the time of the review included over 150 potentially relevant articles, among them the validation of languages other than English, different patient populations, and longitudinal study designs including RCTs. A follow-up publication will address and critically appraise these more recent developments.

In addition, the use of a strict methodological system to assess psychometric quality in PROs may not be appropriate. The COSMIN criteria claim to be applicable to patient-reported outcomes and help to evaluate the psychometric quality of individual assessment tools. Yet they seem to be based on instruments in the true sense, i.e. those having been developed from a measurement model, containing a clear, domain-subdomain-item structure and being amenable to statistical testing. Since this is not the case for PRO-CTCAE, use of these criteria may not seem to be meaningful. So far, the PRO-CTCAE may not be considered as an instrument, but as a tool of heuristic value. Applying rigorous standards of PRO assessment to this new tool may help to improve PRO-CTCAE to best capture the patient's perspective which is of utmost importance in the evaluation of new treatments. So far, most of these standards have not been met in the original papers which implies reviewing subsequent publications and ensuring that upcoming validations of the tool address these methodological requirements.

In view of the topics mentioned above a discussion around all of these is warranted, to ensure the PRO-CTCAE and the patient's voice as such finds its place in the toolbox for clinical studies.

## Conclusion/outlook

PRO-CTCAE offers a flexible tool to assess patient-reported adverse events in cancer treatment which are not fully covered by standard PROs. While this flexibility offers new possibilities in day to day care by engaging patients in their care [5], [14], [24], the free choice of items may result in different items being selected across different studies, making a cross-study comparison difficult, if not impossible, from a clinical research perspective. There is also a risk of a selection bias if individual items (AEs) are selected with a focus on those that are expected to have a positive effect of the drug under investigation, while other key items are disregarded. Therefore, it would be helpful if disease- or substance class-specific standards were developed to overcome this challenge. The authors suggest using a combination of well-known standard PRO tools like EORTC QLQ-C30 and enrichments by selected AEs from the PRO-CTCAE toolbox. This would combine the advantages of the established instruments with the flexibility of PRO-CTCAE thus balancing flexibility and comparability across studies.

Categories of PRO-CTCAE should be comparable to the AE in content, wording and coding, so that patients' and physicians' views of defined symptoms can be taken into account. It would also limit patient burden from additional questions, thus avoiding lengthy questionnaires which



might lead to survey fatigue [23]. Even if there is no correspondence between CTCAE and PRO-CTCAE, it does not compromise the information gathered from PRO-CTCAE since it captures the patient's perspective, which does not necessarily have to align with the physician's perspective. In addition, discrepancies between the two perspectives may provide valuable information. Nevertheless, differences might also occur due to different/insufficient reporting on either or one of the tools.

PRO-CTCAE represents a scientifically interesting concept for capturing the patient perspective, which has been rare in drug safety assessment. It provides an item list for flexible combination of items for a given trial, but at the cost of being founded on a testable measurement model. The present publication evaluates the original psychometric testing of the instrument and finds that more research is needed to provide information on all COSMIN measurement properties. Fulfilling only 1 of 10 criteria is insufficient. A novel psychometric analysis of PRO-CTAE instrument properties, best based on new, also international, data sets, would be desirable. We can say that the validation reported in the 3 papers does not give the information needed to examine if psychometric standards are met. For PRO-CTCAE to be accepted as a psychometrically sound instrument more in-depth and state-of-the-art analyses are needed. An upcoming paper will address the consecutive publications to examine the additional evidence for psychometric performance across trials, diagnoses, languages, and cultures. This will also include an evaluation of further efforts undertaken, including translation and cross-cultural validation.

Based on the need to capture the patients' view, which is up to now rare in drug safety studies, the present publication evaluates the psychometric testing of an instrument designed to do so. The PRO-CTCAE represents a scientifically interesting concept which is based on single items which have been individually validated, but not in any of their combinations. Further research should address such more complex validation strategies and also provide evidence for the psychometric performance of the tool, generated from the item list, in multiple clinical and cross-cultural contexts.

## Notes

### Competing interests

- Dietrich Knoerzer is employee of Roche Pharma AG.
- Martina Kron is employee of Abbvie Deutschland GmbH & Co KG.
- Armin Schüler is employee of MorphoSys AG.
- Susanne Huschens is employee of Janssen-Cilag GmbH.
- Monika Bullinger declares that she has no competing interests.

## Acknowledgment

The authors are grateful for comments of two reviewers which helped to improve the manuscript.

## Attachments

Available from <https://doi.org/10.3205/mibe000273>

1. mibe000273\_Attachment1.pdf (82 KB)  
Appendix

## References

1. Arnold B, Mitchell SA, Lent L, Mendoza TR, Rogak LJ, Barragán NM, Willis G, Medina M, Lechner S, Penedo FJ, Harness JK, Basch EM; PRO-CTCAE Spanish Translation and Linguistic Validation Study Group. Linguistic validation of the Spanish version of the National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). *Support Care Cancer*. 2016 Jul;24(7):2843-51. DOI: 10.1007/s00520-015-3062-5
2. Atkinson TM, Rogak LJ, Heon N, Ryan SJ, Shaw M, Stark LP, Bennett AV, Basch E, Li Y. Exploring differences in adverse symptom event grading thresholds between clinicians and patients in the clinical trial setting. *J Cancer Res Clin Oncol*. 2017 Apr;143(4):735-43. DOI: 10.1007/s00432-016-2335-9
3. Atkinson TM, Ryan SJ, Bennett AV, Stover AM, Saracino RM, Rogak LJ, Jewell ST, Matsoukas K, Li Y, Basch E. The association between clinician-based common terminology criteria for adverse events (CTCAE) and patient-reported outcomes (PRO): a systematic review. *Support Care Cancer*. 2016 Aug;24(8):3669-76. DOI: 10.1007/s00520-016-3297-9
4. Bæksted C, Nissen A, Pappot H, Bidstrup PE, Mitchell SA, Basch E, Dalton SO, Johansen C. Danish Translation and Linguistic Validation of the U.S. National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). *J Pain Symptom Manage*. 2016 Aug;52(2):292-7. DOI: 10.1016/j.jpainsymman.2016.02.008
5. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, Schrag D. Overall Survival Results of a Trial Assessing Patient-Reported Outcomes for Symptom Monitoring During Routine Cancer Treatment. *JAMA*. 2017 Jul;318(2):197-8. DOI: 10.1001/jama.2017.7156
6. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, Rogak L, Bennett AV, Dueck AC, Atkinson TM, Chou JF, Dulko D, Sit L, Barz A, Novotny P, Fruscione M, Sloan JA, Schrag D. Symptom Monitoring With Patient-Reported Outcomes During Routine Cancer Treatment: A Randomized Controlled Trial. *J Clin Oncol*. 2016 Feb;34(6):557-65. DOI: 10.1200/JCO.2015.63.0830
7. Basch E, Iasonos A, McDonough T, Barz A, Culkin A, Kris MG, Scher HI, Schrag D. Patient versus clinician symptom reporting using the National Cancer Institute Common Terminology Criteria for Adverse Events: results of a questionnaire-based study. *Lancet Oncol*. 2006 Nov;7(11):903-9. DOI: 10.1016/S1470-2045(06)70910-X

8. Basch E, Reeve BB, Mitchell SA, Clauser SB, Minasian LM, Dueck AC, Mendoza TR, Hay J, Atkinson TM, Abernethy AP, Bruner DW, Cleeland CS, Sloan JA, Chilukuri R, Baumgartner P, Denicoff A, St Germain D, O'Mara AM, Chen A, Kelaghan J, Bennett AV, Sit L, Rogak L, Barz A, Paul DB, Schrag D. Development of the National Cancer Institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). *J Natl Cancer Inst.* 2014 Sep;106(9):dju244. DOI: 10.1093/jnci/dju244
9. Bennett AV, Dueck AC, Mitchell SA, Mendoza TR, Reeve BB, Atkinson TM, Castro KM, Denicoff A, Rogak LJ, Harness JK, Bearden JD, Bryant D, Siegel RD, Schrag D, Basch E; National Cancer Institute PRO-CTCAE Study Group. Mode equivalence and acceptability of tablet computer-, interactive voice response system-, and paper-based administration of the U.S. National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). *Health Qual Life Outcomes.* 2016 Feb;14:24. DOI: 10.1186/s12955-016-0426-6
10. Bullinger M, Anderson R, Cella D, Aaronson N. Developing and evaluating cross-cultural instruments from minimum requirements to optimal models. *Qual Life Res.* 1993 Dec;2(6):451-9. DOI: 10.1007/BF00422219
11. Cancer Therapy Evaluation Program (CTEP). Common Terminology Criteria for Adverse Events. [last accessed 2021 Mar 17]. Available from: [https://ctep.cancer.gov/protocolDevelopment/electronic\\_applications/ctc.htm](https://ctep.cancer.gov/protocolDevelopment/electronic_applications/ctc.htm)
12. Cappelleri J, Zou K, Bushmakin A, Alvir J, Alemayehu D, Symonds T. Patient-Reported Outcomes: Measurement, Implementation and Interpretation. 1st ed. New York: Chapman & Hall/CRC Press; 2013. (CRC Biostatistics Series).
13. COSMIN. COnsensus-based Standards for the selection of health Measurement INstruments. [last accessed 2021 Mar 17]. Available from: <https://www.cosmin.nl/>
14. Denis F, Basch E, Septans AL, Bennouna J, Urban T, Dueck AC, Letellier C. Two-Year Survival Comparing Web-Based Symptom Monitoring vs Routine Surveillance Following Treatment for Lung Cancer. *JAMA.* 2019 Jan;321(3):306-7. DOI: 10.1001/jama.2018.18085
15. Dueck AC, Mendoza TR, Mitchell SA, Reeve BB, Castro KM, Rogak LJ, Atkinson TM, Bennett AV, Denicoff AM, O'Mara AM, Li Y, Clauser SB, Bryant DM, Bearden JD 3rd, Gillis TA, Harness JK, Siegel RD, Paul DB, Cleeland CS, Schrag D, Sloan JA, Abernethy AP, Bruner DW, Minasian LM, Basch E; National Cancer Institute PRO-CTCAE Study Group. Validity and Reliability of the US National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). *JAMA Oncol.* 2015 Nov;1(8):1051-9. DOI: 10.1001/jamaoncol.2015.2639
16. Evans JP, Smith A, Gibbons C, Alonso J, Valderas JM. The National Institutes of Health Patient-Reported Outcomes Measurement Information System (PROMIS): a view from the UK. *Patient Relat Outcome Meas.* 2018;9:345-52. DOI: 10.2147/PROM.S141378
17. Fayers PMM, Machin D. Quality of Life: The assessment, analysis and reporting of patient reported outcomes. 3rd ed. Chichester: Wiley Blackwell; 2016.
18. Gjersing L, Caplehorn JR, Clausen T. Cross-cultural adaptation of research instruments: language, setting, time and statistical considerations. *BMC Med Res Methodol.* 2010 Feb;10:13. DOI: 10.1186/1471-2288-10-13
19. Hagelstein V, Ortland I, Wilmer A, Mitchell SA, Jaehde U. Validation of the German patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE™). *Ann Oncol.* 2016 Dec;27(12):2294-9. DOI: 10.1093/annonc/mdw422
20. Hay JL, Atkinson TM, Reeve BB, Mitchell SA, Mendoza TR, Willis G, Minasian LM, Clauser SB, Denicoff A, O'Mara A, Chen A, Bennett AV, Paul DB, Gagne J, Rogak L, Sit L, Viswanath V, Schrag D, Basch E; NCI PRO-CTCAE Study Group. Cognitive interviewing of the US National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). *Qual Life Res.* 2014 Feb;23(1):257-69. DOI: 10.1007/s11136-013-0470-1
21. Hinkle DE, Wiersma W, Jurs SG. Applied Statistics for the Behavioral Sciences. 5th ed. London: Houghton Mifflin; 2003.
22. Kirsch M, Mitchell SA, Dobbels F, Stussi G, Basch E, Halter JP, De Geest S. Linguistic and content validation of a German-language PRO-CTCAE-based patient-reported outcomes instrument to evaluate the late effect symptom experience after allogeneic hematopoietic stem cell transplantation. *Eur J Oncol Nurs.* 2015 Feb;19(1):66-74. DOI: 10.1016/j.ejon.2014.07.007
23. Kluetz PG, Chingos DT, Basch EM, Mitchell SA. Patient-Reported Outcomes in Cancer Clinical Trials: Measuring Symptomatic Adverse Events With the National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). *Am Soc Clin Oncol Educ Book.* 2016;35:67-73. DOI: 10.1200/EDBK\_159514
24. Lee JK, Assel M, Thong AE, Sjoberg DD, Mulhall JP, Sandhu J, Vickers AJ, Ehdaie B. Unexpected Long-term Improvements in Urinary and Erectile Function in a Large Cohort of Men with Self-reported Outcomes Following Radical Prostatectomy. *Eur Urol.* 2015 Nov;68(5):899-905. DOI: 10.1016/j.eururo.2015.07.074
25. MedDRA. Medical Dictionary for Regulatory Activities. [last accessed 2021 Mar 17]. Available from: <https://www.meddra.org/>
26. Miyaji T, Iioka Y, Kuroda Y, Yamamoto D, Iwase S, Goto Y, Tsuboi M, Odagiri H, Tsubota Y, Kawaguchi T, Sakata N, Basch E, Yamaguchi T. Japanese translation and linguistic validation of the US National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). *J Patient Rep Outcomes.* 2017;1(1):8. DOI: 10.1186/s41687-017-0012-7
27. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol.* 2010 Jul;63(7):737-45. DOI: 10.1016/j.jclinepi.2010.02.006
28. Mühlan H, Bullinger M, Power M, Schmidt S. Short forms of subjective quality of life assessments from cross-cultural studies for use in surveys with different populations. *Clin Psychol Psychother.* 2008;15(3):142-53. DOI: 10.1002/cpp.573
29. National Cancer Institute. Patient-Reported Outcomes version Of The Common Terminology Criteria For Adverse Events (PRO-CTCAE). Quick Guide to the item library. [version 2020 Nov 03, last accessed 2021 Mar 17]. Available from: <https://healthcaresdelivery.cancer.gov/pro-ctcae/item-library.pdf>
30. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res.* 2018 May;27(5):1147-57. DOI: 10.1007/s11136-018-1798-3
31. Schmidt S, Bullinger M. Current issues in cross-cultural quality of life instrument development. *Arch Phys Med Rehabil.* 2003 Apr;84(4 Suppl 2):S29-34. DOI: 10.1053/apmr.2003.50244
32. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res.* 2012 May;21(4):651-7. DOI: 10.1007/s11136-011-9960-1

33. Trotti A, Colevas AD, Setser A, Basch E. Patient-reported outcomes and the evolution of adverse event reporting in oncology. *J Clin Oncol.* 2007 Nov;25(32):5121-7. DOI: 10.1200/JCO.2007.12.4784
34. Wu H, Fung M, Hornbuckle K, Muniz E. Impact of Geographic and Cross-Cultural Differences on Spontaneous Adverse Events Reporting. *Drug Inf J.* 1999;33:921-31.
35. Xiao C, Polomano R, Bruner DW. Comparison between patient-reported and clinician-observed symptoms in oncology. *Cancer Nurs.* 2013;36(6):E1-E16. DOI: 10.1097/NCC.0b013e318269040f

**Please cite as**

Knoerzer D, Kron M, Schüler A, Huschens S, Bullinger M. Assessing patient-reported adverse events in oncology – a review of the PRO-CTCAE instrument. *GMS Med Inform Biom Epidemiol.* 2025;21:Doc01. DOI: 10.3205/mibe000273, URN: urn:nbn:de:0183-mibe0002735

**This article is freely available from**

<https://doi.org/10.3205/mibe000273>

**Published:** 2025-04-10

**Copyright**

©2025 Knoerzer et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. See license information at <http://creativecommons.org/licenses/by/4.0/>.

**Corresponding author:**

Dietrich Knoerzer  
Roche Pharma AG, Department of Biometrics &  
Epidemiology, Emil-Barell-Str. 1, 79639 Grenzach,  
Germany, Phone: +49 (0)172 2047202  
[dietrich.knoerzer@roche.com](mailto:dietrich.knoerzer@roche.com)