Reflection Paper on the use of patient reported outcome (PRO) measures in oncology studies

6   Draft

Draft Agreed by Oncology Working Party	17 December 2013
Adoption by CHMP for release for consultation	22 May 2014
Start of consultation	17 June 2014
End of consultation (deadline for comments)	30 November 2014

Comments should be provided using this template. The completed comments form should be sent to ONCWPsecretariat@ema.europa.eu

Keywords	Patient reported outcome (PRO), Health related quality of life (HRQL)

Executive summary

The importance of the patient's point of view on their health status is fully acknowledged and such information may in principle be used in drawing regulatory conclusions regarding treatment effects. This reflection paper on the use of patient reported outcome (PRO) measures in patients with malignancies focuses on the value of these data from a regulatory perspective. The possible add-on value from a licensure perspective of such data to conventional efficacy and safety data is therefore emphasised. In particular the use of PRO data in order to estimate patient perception of side effects of therapy is highlighted.

This document has been named "reflection paper" in order to underline its preliminary status and to spur an open discussion on the value of PRO data in the development of medicinal products for the treatment of malignancies and in acknowledgment that PRO methodology is developing and evolving.

Important definitions

PRO   A PRO includes any outcome evaluated directly by the patient himself and based on patient's perception of a disease and its treatment(s). Patient reported outcome is an umbrella term covering both single dimension and multi-dimension measures of symptoms, health-related quality of life (HRQL), health status, adherence to treatment, satisfaction with treatment, etc.

HRQL   Health-related quality of life is a specific type of PRO and is a broad concept which can be defined as the patient's subjective perception of the impact of his disease and its treatment(s) on his daily life, physical, psychological and social functioning and well-being. The notion of multidimensionality is a key component of the definition of HRQL.

1. Background

PRO measure is an umbrella term for the capturing of health status, symptoms, HRQL, adherence to treatment, satisfaction with treatment, etc with the emphasis placed upon the patient's judgement. It is recognised that such data are subjective, change over time and are influenced by the treatment, the disease and other co-morbidities. HRQL is a concept referring to the effect of an illness and its therapy upon a patient's physical, psychological and social wellbeing, as perceived by the patient themselves. In clinical research, such measures may provide an additional means of capturing the personal and social context of the disease and treatment experience, as objective clinical measures may not necessarily correlate to a patients own feeling of wellbeing.

Over the last decades, HRQL objectives have frequently been incorporated in confirmatory oncology studies. However longitudinal HRQL data have rarely been informative from a licensure perspective, a main reason being the absence of demonstrated difference between the study arms. Whether this is related to poor sensitivity of the instruments, high attrition rates and informative censoring, or simply reflects the resilience and dynamics of the individual's perception of HRQL during the course of disease, remains unknown. In addition, there is often a lack of consensus regarding what degree of difference is clinically relevant, which together with poorly defined objectives may further hamper the usefulness of PROs from a licensure perspective.

More recently, time to significant deterioration in tumour related symptoms, as measured by PRO instruments, has been introduced and here differences have been demonstrated, paralleling what has been shown in terms of progression-free survival (PFS). This is of value, but it could be discussed whether repeat demonstration of parallelism between PFS and time to symptom deterioration, e.g. in

56 the treatment of lung cancer, is of value. More importantly, these data do not provide estimates of
57 longitudinal HRQL, i.e. do not provide a weighed approach of benefits and adverse reactions of
58 therapy.

59 In most cases, for a particular tumour type and disease stage, there is no reason to assume that the
60 potential benefit of a delay in tumour progression, if of similar magnitude, is product specific. However,
61 the tolerability and toxicity profiles may differ considerably between medicinal products. The
62 differential impact on patient wellbeing is harder to estimate from conventional adverse event
63 reporting, even though withdrawal rates prior to tumour progression may provide some insights. In
64 relation to active compound comparative trials and from a licensure perspective, PRO data derived
65 from instruments capturing the consequences of adverse reaction on patient wellbeing, in an unbiased
66 way and in relation to the study drugs, are welcomed. However, at the time of this paper there is no
67 EMA/CHMP experience from the use of, e.g. the NCI's Patient-Reported Outcomes version of the
68 Common Terminology Criteria for Adverse Events (PRO-CTCAE).

69 In summary, PRO measures may provide important patient perspective on the disease and the
70 treatment received; an evaluation that provides clinically important information that is not captured by
71 conventional anti-tumour efficacy data and adverse event reporting. There are, however,
72 methodological obstacles that historically have reduced the impact of PRO data on regulatory
73 decisions. Key is careful planning and an in depth analysis of whether the inclusion of PRO measures is
74 likely to provide added value in the clinical trial setting; can the collection of PRO data make a potential
75 difference to the study conclusions.

76 ## 2. Scope

77 This reflection paper covers general aspects of the use of PRO endpoints in oncology studies such as
78 the designing and carrying out of clinical studies, the acceptability of instruments and the clinically
79 important differences and added value. This reflection paper does not cover the validation of
80 instruments nor does it make specific recommendations regarding the instrument to select.

81 ## 3. Legal basis

82 This document should be read in conjunction with Directive 2001/83/EC, as amended and Regulation
83 726/2004. In addition, relevant CHMP guidelines should be taken into account. These include but are
84 not limited to:

85 - Guideline on the evaluation of anticancer medicinal products in man -EMA/CHMP/205/95/Rev.4

86 - Statistical principles for clinical trials – CPMP/ICH/363/96 (ICH E9)

87 - Reflection paper on the regulatory guidance for the use of HRQL measures in the evaluation of
88   medicinal products - EMEA/CHMP/EWP/139391/2004

89 - Guideline on missing data in confirmatory clinical trials EMA/CPMP/EWP/1776/99 Rev. 1

90 - Points to consider on multiplicity issues in clinical trials CPMP/EWP/908/99

91 ## 4. Patient reported outcomes

92 A patient-reported outcome (PRO) is an umbrella term that can be defined as a measurement based on
93 a report that comes directly from the patient about the status of a patient's perception of the impact of

94   disease and treatment, without amendment or interpretation of the patient's response by a clinician or
95   anyone else. A PRO can be measured by self-report or by interview, provided that the interviewer
96   records only the patient's response. PRO measures must have acceptable responsiveness, reliability
97   and validity, and may include reference to symptoms, functional status, treatment adherence or
98   satisfaction with care. In clinical research, the use of a PRO measure is advised when measuring a
99   concept best known to the patient or best measured from the patient perspective. Clinical studies in
100 oncology may include PRO measures as secondary or exploratory outcomes and rarely as primary
101 outcomes, incorporated as part of the initial trial protocol. The general recommendations for the
102 incorporation of PRO measures in clinical studies include:

103 •   The extent to which the inclusion of PRO measures can provide added value in the clinical trial
104   setting; crucially can the collection of PRO data make a difference to the study conclusions.

105 •   PRO endpoints should be incorporated into the protocol development at the earliest stage and
106   should be explicitly stated as a specific clinical trial objective or hypothesis.

107 •   For specific therapeutic claims in Section 5.1 of the SmPC, a clear hypothesis lead strategy is
108   required and measures should be selected based on their 'fit' with the hypothesis.

109 •   Questionnaires & instruments should be administered to study subjects at time points when
110   there is a clear and hypothesis driven rationale for their use and when it is feasible to expect
111   high levels of completion. PRO instruments should match the abilities of the patient population.

112 •   PRO data should be treated like any other data in monitoring clinical site performance and
113   collection methods

4.1. Health Related Quality of Life (HRQL)

115 The impact of treatment and disease can be measured using self-reported questionnaires. HRQL is a
116 multidomain concept that represents the patient's general perception of the effect of illness and
117 treatment on physical, psychological and social aspects of life. HRQL instruments attempt to measure
118 complex aspects of life which are potentially modified by therapeutic interventions. HRQL is a personal
119 perspective and varies with gender, experience, age, education and cultural background. The inclusion
120 of HRQL assessment in clinical trials should have a strong scientific rationale and researchers should
121 utilise existing validated instruments where available. HRQL complements the range of traditional
122 indicators and the data can provide information regarding both positive and negative patient
123 experiences. Reasons to include HRQL assessment in the clinical development programme for oncology
124 medicinal products includes:

125 •   Provide a patient focused assessment of the burden and impact of disease
126 •   Understand how a novel treatment impacts on patient functioning
127 •   Add information on the positive and negative effects of a therapy by complementing efficacy
128   and safety data e.g. help assess the relationship between efficacy/ clinical endpoints (OS, PFS,
129   disease stabilisation) and HRQL
130 •   Identify treatment-related symptoms that need additional management and supportive care
131 •   Attempt to differentiate two treatments with similar efficacy
132 •   Facilitate more accurate patient-physician communication in terms of the quality of the time
133   remaining and the burden of treatment-related morbidities by detailing a more complete
134   evaluation of cancer treatment

5. Clinical trial design

General principles (see also Reflection paper on the regulatory guidance for the use of HRQL
measures in the evaluation of medicinal products, Doc. Ref.
EMEA/CHMP/EWP/139391/2004)

There is no standard approach to collecting, analysing or interpreting PRO data in clinical trials. As with
other aspects of clinical trial design, good science applies and objectives need to be justified alongside
realistic expectations. Careful thought must go into designing and implementing PRO measures in the
oncology clinical trial setting in order to investigate a well-formulated predefined hypothesis, whether
related to HRQL or a more targeted objective better captured by a more focused PRO instrument. In
the majority of circumstances, the patient is the best informant and the most appropriate way to
measure PRO is self-reporting direct from the patient. Importantly, measurements should not
constitute an undue burden to the patient.

There has been a general perception that only truly double-blind studies can provide trustworthy PRO
data. There is a paradox in this, as it implies that differences in side effects profiles should be
sufficiently small not to be detected by patient and treating physician. Whether such small differences
are sufficiently large to be detectable by PRO instruments are dubitable and such effects are perhaps
also of minor clinical relevance. It is obviously true that possible differences in positive effects on
tumour related symptoms might be detectable, but the added value to so called objective measures of
tumour response and delay in progression might be of relevance mainly in studies without an active
comparator. Thus whilst ideal from a "bias perspective", informative double blind studies may be
successfully conducted only in specific situations.

Whilst the concern in relation to open label studies remains, it might well be that data of clinical
interest a priori can be produced only under open label conditions. One example being an experimental
compound assumed to be more efficacious, but also more toxic or less well tolerated. Under these
circumstances extensive planning in advance is required to increase the credibility of study data. For
example, effects of neuropathy on functionality should be supported by conventional clinical measures
of neuropathy. As emphasised, it is of major importance to discuss in detail in the study protocol why
certain timings of assessments were selected and why the selected instrument is unbiased in relation
to the toxicity/tolerability profiles of study drugs.

Frequency and duration of assessments

Timing and frequency of assessment are key issues and frequency can greatly influence the scores
received. If assessments are too few, important changes may not be captured, if too frequent, the
subject may become sensitised to the instrument. The overall frequency of assessment depends on the
hypothesis being tested, the method of data analysis, the natural history of the disease and the nature
of the investigative treatment and anticipated side effects. It is generally recommended to determine
when expected changes in symptoms and or side effects are likely to occur over time and data
collection should cover the clinically most important periods. The duration of assessment depends on
the research questions being asked, but it is important to ensure that the duration of the clinical study
and follow up is of adequate length to robustly support any planned analysis, including reversibility of
adverse reactions.

In order to be able to accurately assess the PRO results on study therapy, continued assessment post-
progression and during next-line therapy may also be needed. Such next-line PRO data allows
contextualisation of the results observed on study treatment, which can be of particular importance in

178 the palliative or maintenance setting, and when therapeutic claims (in SmPC section 5.1.) are
179 intended. For example, when an active treatment is compared against placebo or other less toxic
180 therapy, worse scores for PROs may be seen during treatment in the active/experimental arm due to
181 toxicity. In this situation, if there is no gain in OS, or if OS cannot be assessed (e.g. due to lack of
182 power, immature data, or cross-over), next-line PRO data can help put the PFS gain into perspective
183 and could potentially affect the benefit/risk (B/R) balance. Apart from the need for contextualisation,
184 there is also a methodological rationale for collecting next-line or post-progression data when PROs are
185 studied. Patients in the comparator arm are normally expected (as a group) to experience progression
186 earlier than the patients in the experimental arm. Thus, if PRO assessments are stopped at
187 progression, patients in the comparator arm will automatically have a shorter observation period
188 compared with those in the experimental arm. This can be regarded as a form of informative
189 censoring, affecting the possibilities to draw conclusions from the PRO data.

190 *Data collection*

191 High compliance has been attributed to comprehensive educational programmes prior and during the
192 trial for both research staff and study participants. Assessments should be performed on schedule
193 irrespective of whether study treatment has been given. Collecting PRO data from patients with
194 advanced and progressive disease may be more difficult because of failing health and / or cognitive
195 challenges. PRO data can be collected by administering PRO instruments through different modes –
196 interviewing, telephone, mailing or self administration. Electronic data capture methods may offer
197 more convenience to some patients and may increase data quality, reduce missing data (allowing
198 automatic reminders to be sent) and potentially reduce data entry errors.

199 *Statistical methods and missing data*

200 Incorporating PRO instruments as clinical trial endpoint measures introduces challenges in the analysis
201 of clinical trial data, particularly because of their multi-dimensional nature and missing values. The
202 study protocol should describe the principal data analysis features in the statistical section with a
203 detailed elaboration of the analysis in the Statistical Analysis Plan, including how to control for
204 multiplicity. The clinical trial protocol should also describe how missing data will be handled in the
205 analysis (e.g. use of imputational techniques, sensitivity analysis). Missing data should be put into
206 context of underlying reason, but missing at random is hardly ever a justified assumption. It is
207 therefore essential to minimise data loss and to employ strategies to increase patient compliance, such
208 as, for example;

209 • Filling in baseline questionnaire as part of the eligibility criteria checklist

210 • Appoint a person responsible for PRO data collection in each study site

211 • Education and training to patients before completion of the questionnaire, including that there is no
212   incorrect answer and explaining the purpose of the assessment

213 • Explore the use of automated electronic data collection

214 • Checking for completeness of forms for omissions, clarifying reasons for non-completion

215 Of importance, whilst use of electronic data recording might be of benefit in some patient groups,
216 alternatives should be made available, e.g. for elderly patients so that differential loss of data is
217 minimised.

218 Depending on the chosen instruments, lack of linguistic and cultural validation of instruments may be
219 problematic in multinational and global studies. Investigation of PRO endpoints in a predefined
220 sufficiently large subgroup, including patients representative for the EU target population, may be
221 considered to avoid cross-cultural validity and translation issues. In these cases, implementation of all
222 measures for a high compliance rate is expected in order to provide a substantial amount of
223 interpretable longitudinal data.

5.1. Instruments

225 PRO instruments should be relevant, reliable, validated and responsive to change. An instrument can
226 be described as a means to capture data, such as a questionnaire, plus all the information and
227 documentation that supports its use. Generally, this includes clearly defined methods and instructions
228 for administration or responding, a standard format for data collection, and well-documented methods
229 for scoring, analysis and interpretation of results in the target patient population. Disease specific
230 measures may be more acceptable to patients, providing a more in-depth relevant analysis. However,
231 they may fail to capture unexpected changes. Generic measures are useful for comparisons across
232 treatments. However, they may be less sensitive to change and the relative importance of the different
233 PRO domains needs to be determined a priori.

Selection of an instrument

235 It is beyond the scope of this reflection paper to make specific recommendations regarding valid
236 instrument selection, but in general, the instrument should be shown to measure the concept it is
237 intended to measure, be appropriate for the research objective, the disease and patient population
238 characteristics and the practical considerations (respondent burden, feasibility). Instruments should be
239 culturally valid and translated versions should be as true to the original as possible (linguistic
240 validation).

Carer/ proxy input

242 There is generally discordance between 'patient' reported PRO and 'proxy' reported PRO. The
243 evaluation of PRO by carers or other proxy judges may be utilised where it is clear that the patient
244 themselves cannot contribute (e.g. very small children, patients with cognitive impairment, severe ill
245 health), but in general proxy reporting should be avoided.

5.2. Special patient populations

Paediatric

248 Specific issues to consider are development stage (maturation may also differ because of disease and
249 or experiences) and meaning of self. As with adult patients, the best informants are the patients
250 themselves and it is important to collect as much information directly from the patient wherever
251 possible, using creative and age related approaches. However it is acknowledged that some patients
252 will be too young or too sick to contribute to the data collection.

Elderly

254 Elderly patients present particular characteristics and instruments should be calibrated to the special
255 requirements of older patients wherever possible. In elderly patients, concomitant diseases are more
256 frequent, affecting psychological status and general performance. It is important to consider that HRQL

257 is affected by comorbidities, multiple medications (polypharmacy), functional status, ability to carry
258 out activities of daily living, mental status (depression, cognitive functioning) and social support.

259 ***Palliative setting (for definition see 7.4, main anticancer document)***

260 Successful patient palliation has been described as disappearance or improvement of symptoms,
261 improvement of a specific symptom from baseline, change in the severity of a specific target symptom,
262 for example pain or composite outcomes of pain and analgesic requirements, a symptom difference
263 perceived as beneficial by the patient, HRQL score changes or increased duration of survival. In
264 patients with advanced cancer where the aims are palliative, the focus of care is promoting and
265 maintaining remaining quality of life. This aspect should be carefully considered in the clinical study
266 design, in particular as complicated multidimensional changes can occur relatively quickly and patient
267 survival time is relatively short. If appropriate, longitudinal HRQL data should be collected alongside
268 other PRO measures such as symptom assessment (see section 6).

269 6. Symptom PRO measures

270 Patients provide a unique and personal perspective of treatment effectiveness and measuring
271 symptoms is important in understanding the burden of cancer. Symptom response rates and symptom
272 control are particularly significant in the palliative setting. Assessment of palliation can be assessed by
273 changes in symptom scores in general or change in symptom scores considering only certain
274 prespecified symptoms. Symptoms (related to the disease, toxicity or multi-factorial) that are
275 commonly found in the advanced setting include anorexia, anxiety, constipation, depression,
276 dyspnoea, fatigue, insomnia, pain and neuropathy. However, patient reported symptoms to be
277 investigated should be evidence based and derived from feedback from patients and carers, clinicians
278 and other experts, as well as the literature. If symptom PRO measures are used to evaluate the impact
279 on specific symptoms, these should be accompanied by multidimensional HRQL measures to ensure
280 that a benefit in respect to specific symptoms is not accompanied by a negative impact on global
281 HRQL. As important as selecting an instrument that properly captures disease related symptoms is to
282 use an instrument that captures side effects of therapy in an unbiased way.

283 7. Clinical importance and added-value

284 PRO instruments and assessments should be capable of detecting clinically meaningful effects. The
285 Minimal Clinically Important Difference (MCID) has been described as 'the smallest difference in score
286 in the domain of interest which patients perceive as beneficial and which would mandate, in the
287 absence of troublesome side effects and excessive cost, a change in the patient's management'.
288 Situations where PRO measures, including HRQL, could potentially be of added value in terms of
289 possibly affecting the benefit risk profile, include the late line palliative setting, maintenance therapy,
290 and in studies comparing agents with similar efficacy but different safety profiles. In some disease
291 settings, symptom response and especially time to relevant deterioration might in principle be used as
292 primary outcome measures, provided that data are supported by ORR and PFS. Criteria used to assess
293 the potential added value of PRO data include:

294 • The relevance, reliability and responsiveness of the instrument/ assessment

295 • The appropriateness of the frequency and duration of data collection, in light of the patient
296 population, disease setting and treatment regimen

297 • The adequacy of the study design including the hypothesis and methods for appropriate
298 handling of multiple outcomes in the statistical analysis

299 • The rationale for the anticipated magnitude of effect - statistical significance should correlate
300 with clinically relevance

301 • Considerations of alternative explanations that may account for the observed changes or lack
302 of changes

References

304 Brundage et al., 2008. NCIC Clinical Trials Group experience of employing patient-reported outcomes
305 in clinical trials: an illustrative study in a palliative setting. *Exert Rev. Pharmacoeconomics Outcomes*
306 *Res.* 8(3):243-253

307 Calvert et al., 2013. Reporting of Patient-Reported Outcomes in Randomized Trials. The CONSORT PRO
308 Extension. Jama, Vol 309, No 8: 814-822

309 Ganz and Cotay., 2007. Use of patient-reported outcomes in phase III cancer treatment trials: lessons
310 learned and future directions. J. Clin Oncol. 25(32): 5063-9

311 Jaeschke et al., 1989. Measurement of health status. Ascertaining the minimal clinically important
312 difference. Control Clin Trials. 1989 Dec;10(4):407-15.

313 Joly et al., 2007. Quality of life and/or symptom control in randomised clinical trails for patients with
314 advanced cancer. Annals of Oncology. 18: 1935-1942

315 Pasetto et al., 2007. Quality of life in elderly cancer patients. European Journal of Cancer. 43: 1508-
316 1513.

317 Revicki D., et al 2008. Recommended methods for determining responsiveness and minimally
318 important differences for patient-reported outcomes. Journal of Clinical Epidemiology 61;102-109

319 Sloan JA., et al 2006. The clinical significance of quality of life assessments in oncology: a summary for
320 clinicians. Support Care Cancer. 14:988-998.