EUROPEAN JOURNAL OF CANCER 45 (2009) 347–353

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejconline.com

Review

Patient-reported outcomes: Assessment and current perspectives of the guidelines of the Food and Drug Administration and the reflection paper of the European Medicines Agency

Andrew Bottomley[a,*], *Dave Jones*[b], *Lily Claassens*[a]

[a]*EORTC Quality of Life Department, EORTC Headquarters, Avenue E. Mounierlaan, 83/11, Brussels 1200, Belgium*
[b]*AstraZeneca, Alderly Park, Macclesfield, England, United Kingdom*

ARTICLE INFO

Article history:
Received 10 September 2008
Accepted 30 September 2008
Available online 14 November 2008

Keywords:
Patient-reported outcomes
Regulatory guidance
Health-related quality of life

ABSTRACT

Aims: Patient-reported outcomes (PROs) have recently gained greater credibility with regulatory bodies aiming to standardise their use and interpretation in RCTs, thereby supporting medicinal product submissions. For this reason, the United States (US) Food and Drug Administration (FDA) and the European Medicines Agency (EMEA) have released guidelines. This review paper provides an overview of the current perspectives and views on these guidelines.

Method: To evaluate the FDA and EMEA PRO guidelines, 47 expert responses to the FDA guidance were qualitatively reviewed. Two reviewers independently extracted data from these letters and checked these responses to warrant consistency and agreement in the evaluation process. A PubMed literature review was systematically examined to obtain supporting evidence or related articles for both the guidance documents.

Results: Generally, there is agreement between regulatory authorities and the research community on the contents of the FDA and EMEA PRO draft guidance. However, disagreements exist on significant philosophical topics (e.g. the FDA focuses more on conceptual models and symptoms than the EMEA) and design topics (e.g. the FDA is more restrictive on issues of recall bias, blinding of oncology trials and degrees of psychometric validation than researchers and the EMEA). This could influence the approval of PRO claims.

Conclusion: PRO guidance from the EMEA and FDA has been valuable, and has raised the profile and active debate of PROs in oncology. However, our review of the current opinion shows that there are controversial aspects of the guidance. Consequently, greater latitude should be given to how the guidance is interpreted and applied.

* *Corresponding author:* Tel.: +32 2774 1661; fax: +32 2779 4568.
E-mail address: andrew.bottomley@eortc.be (A. Bottomley).

0959-8049/$ - see front matter
doi:10.1016/j.ejca.2008.09.032

1. Introduction

In the recent years, the use of patient-reported outcomes (PROs) has increased significantly.[1] HRQOL measures involve subjective patient assessment or evaluation of important aspects of well-being[2] that are affected by current disease and/or treatment. Prominent examples of cancer-related HRQOL tools are the EORTC QLQ-C30 and the Functional Assessment of Cancer Therapy General (FACT G).[3,4]

Recently, the Food and Drug Administration (FDA) introduced the umbrella term PRO and attempted to standardise PRO use to provide a more systematic treatment-review process. For this reason, the FDA released draft guidance on PRO measures in February 2006: *Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*.[5] Also, the European Medicines Agency (EMEA) produced a reflective paper of regulatory guidance for the use of HRQOL measures in the evaluation of medicinal products in cancer in July, 2005.[6]

However, despite considerable effort to develop guidelines, it is not entirely clear what evidence the regulatory bodies would require in supporting claims when reporting PRO data. To date, established PRO measures are criticised by the regulatory agencies, leading to rejection of many PRO labelling claims.

The main objective of this paper is to review the current opinion relating to the guidance, and to make recommendations based on our review.

2. Materials and methods

The aim of the study is to provide an overview of the current opinion in relation to the two major guidance documents: the FDA Patient-Reported Outcomes draft (2006) and the EMEA HRQOL reflection paper (2005). A thematic qualitative approach was used to compare recommendations and requirements from each guidance. Written comments, invited by the FDA and submitted to the FDA web site (Table 1), have been reviewed. These reviewed responses related to the FDA guidance were also considered in relation to the EMEA document.

In order to seek evidence in support of the statements or recommendations within these regulatory guidance documents, a systematic literature search using PubMed was undertaken from January 1990 to December 2007. This searched for key words related to recurring issues which arose within the documents to identify scientific evidence to clarify debatable issues. All searches were restricted to English language articles only. In addition, literature references were checked to identify further evidence. Abstracts for major conferences, e.g. ASCO (2005–2008) and ESMO (2005–2008), were also reviewed.

3. Results

The FDA PRO guidance generated 47 written responses, totalling to 364 pages of comments, which are accessible on the FDA web site. These comments mainly came from professional groups in both academia and the pharmaceutical sector (Table 1). No documents were found on the EMEA web site about the EMEA HRQOL reflection paper, but the literature search identified several studies (discussed later) that looked at these guidelines. The search on PubMed and the search of grey literature generated additional documents referred to in the results below. The results are presented on the main thematic areas arising from the documents and from the responses to the documents. The comments' numbers refer to the authors who are listed in Table 1.

3.1. *Conceptual framework and end-point model*

According to the FDA guidance, PRO instruments must be based on an appropriate and clearly defined framework. This requires documentation of patient interviews, literature reviews and expert clinical opinion in order to support the concepts, domains and their associations (FDA, Section IV A, p. 7). The EMEA briefly noted the importance of incorporating the clinically relevant health-related domains of functioning that impact on HRQOL (p. 3).

Additionally, a submission to the FDA should be supported by an end-point model,[7] which displays an overview of the relevant end-points in an RCT and their relationships by mapping the treatment benefit and appropriate claims. Such a model should be hypothesis driven and incorporate a specific perspective,[8] i.e. the FDA prefers researchers to pre-specify expectations concerning the treatment impact, such as impact on disease symptoms, and the scales by which these outcomes will be measured. Some of the respondents who commented on this FDA opinion agree with a conceptual framework as a basis for PRO questionnaires (Comments 29, 33 and 36). Others suggested certain amendments, e.g. in its definition (Comments 6 and 38), or emphasised the distinction between HRQOL and symptoms (Comments 28 and 44). However, clearer guidance would be appreciated on the type of empirical data that need to support the conceptual framework (Comments 13 and 14), and two respondents recommended that the FDA should clarify that a conceptual framework precedes empirical analyses and may change during the validation process (Comments 15 and 20). Some respondents (Comments 10 and 34) questioned if a conceptual framework is needed in the manner the FDA stipulates (e.g. in a diagram).

3.2. *Patient involvement in instrument development*

The FDA aims to review instrument development to determine whether an adequate number of patients have supported the opinion that the specific items in the instrument are adequate and appropriate to measure the desired concept(s) (FDA, Section IV B, p. 10). No EMEA position is given on this issue. It is unclear what the appropriate number of patients involved in item generation ought to be according to the FDA. Eight respondents raised the question of how the FDA will evaluate if the sample size of patients, used in questionnaire generation, is adequate (Comments 8, 11, 15, 22, 24, 30, 39 and 43). One respondent (Comment 20) stated that item generation often involves small numbers of patients.

In addition, the characteristics of patients participating in early questionnaire development should match the charac-

teristics of the target population. The FDA plans to compare the patient population used in the PRO instrument development to the study populations enrolled in clinical trials (FDA, Section IV B, p. 9). The EMEA made no related statements on this matter. Ten respondents were of the opinion that population comparisons using the list of specifically age, sex, ethnic identity and cognitive ability are too specific (Comments 9, 19, 20, 21, 23, 29, 30, 34, 39 and 43). Many respondents argued that while the pertinent characteristics should be determined and compared, these might be different from those listed by the FDA (e.g. disease type and disease severity).

3.3. *Multi-domains versus single domains: making PRO claims*

The FDA and the EMEA stated that general PRO claims should be based on and supported by improvement in all domains (FDA, Section IV A, p. 8) or at least in the most important domains (EMEA, p. 3). Specific claims based on individual items may be proposed, though only if these single items are validated for this purpose and pre-specified in the Statistical Analysis Plan (SAP). Twelve respondents criticised or questioned this requirement in various ways. For example, patients may not be impaired in all domains (Comments 1, 43), and therefore, an improvement in the totality of domains is not deemed feasible. A flexible approach is preferred that requires improvement in the most important domains (Comment 15), and no trend towards worsening for the other domains (Comment 37). Furthermore, it was suggested that a single component could be the basis of a general claim (Comments 10, 29, 32 and 41), and one respondent stated it should be pre-specified in the Statistical Analysis Plan (Comment 34). Several respondents needed more details on this issue (Comments 13, 24, 34 and 39).

3.4. *Recall period*

An appropriate recall period is required when assessing the effects of treatment on oncology PROs. The FDA argues for the measurement of the current state (Section IV B, p. 11). In contrast, the EMEA made no statements on a recall period. There is a consensus among respondent views that averaging experience over a period of days does not necessarily invalidate measurement (Comments 12, 14, 15, 18, 19, 21, 23, 24, 32, 34, 39 and 40): e.g. it can be valuable to measure patient evaluation of change over time (Comments 1, 10 and 36). Besides, the current state assessments over a longer period of time may include bias as well: e.g. caused by response shift (Comment 11). Two respondents (Comments 20 and 32) agreed that a 'current state' approach gives rise to data being influenced to too high an extent by the extremes of patient state. Indeed the current state can also be influenced by the use of concomitant medications. Furthermore, symptoms are believed to be appropriate for the current state assessments while instrumental activities of daily living may not be appropriate for current state measurement (Comment 20). Several respondents considered that the guideline may be too prescriptive or too general (Comments 11, 12, 19, 21 and 26) or contradictory to the previous FDA advice (Comment 22). The choice of recall period should depend on the type of disease (Comments 12, 19, 20 and 43), the question asked (Comments 19, 24 and 43) and the situation related to patient health and disease (Comment 15). Moreover, the FDA opinion on a recall period was believed to be unsupported by consistent or sufficient evidence (Comments 12, 15, 23, 24, 41 and 43). However, three respondents did agree with the FDA proposal of the current state measurement (Comments 7, 17 and 28).

Previously published research addresses the accuracy of recall in pain patients[9–15] and shows variability in the results. Certain researchers claimed retrospective perception of change may not be accurate due to recall bias.[13,15] Other researchers considered recall of experience to be equivalent to assessment of the current state[14,9–11], concluding that measurement error and significant regression effect are the main concerns in momentary measurements.[11]

3.5. *Reliability and validity*

The FDA guidance emphasised test–retest reliability as the most important reliability type, while internal consistency may be used in the absence of the test–retest reliability (Section IV C, p. 18). EMEA did not address this topic. Eight respondents did not fully agree with the FDA: measuring test–retest reliability may be inappropriate or not feasible in some circumstances, and may be replaced by other reliability tests, e.g. internal consistency reliability (Comments 10, 11, 15, 23, 28, 29, 36 and 41). One respondent (Comment 6) noted that evidence is lacking for this FDA opinion.

Further, the FDA recommended that content-related validity should be addressed by providing evidence that items and response options are of a relevant and comprehensive nature with respect to the concepts that should be measured (Section IV C, p. 16). Evidence should contain a documentation of the issues abstracted from the literature; interview processes involving patients and healthcare providers (including interview transcripts) and information relating to the addition or deletion of items. Construct validity determines the extent to which items, domains and concepts relate to one another, supported by item-scale correlation analyses. Finally, predictive validity, the ability to predict the future outcomes through patient characteristics with prognostic value, is on the FDA list of psychometric properties (Section IV C, p. 16). However, it is questioned if this type of validity should be obligatory evidence in a submission to the agencies. Seven respondents agreed that flexibility is important with respect to the predictive validity (Comments 15, 18, 19, 21, 24, 39 and 43), since it may be unrealistic in some PROs, especially in new instruments. Our PubMed literature review showed the documentation of psychometric characteristics (e.g. construct validity) to be of great importance when trying to obtain FDA approval.[16] However, it would be appreciated if the FDA would emphasise that the demonstration of all measurement properties is an 'ideal' rather than a 'requirement' (Comment 34).

3.6. *Ability of PRO measures to detect change*

An instrument should detect changes in outcome measures if relevant clinical changes have occurred. For this reason,

evidence must be documented showing the degree to which the PRO instrument detected expected changes in values that are thought to have changed (part IV C, p. 18). This point was not debated significantly.

3.7. Interpretability of PRO measures

A Minimum Important Difference (MID) can be generated by applying a variety of methods, e.g. an anchor-based or a distribution-based approach or an empirical rule (FDA, Section IV C, p. 19). The EMEA stated that a MID 'should be based upon a combination of statistical reasoning and clinical judgment' (p. 5).

Respondents believe that MID should not be referred to as the only method for interpretability (Comments 34 and 35): some argued that it is not an exact science and has no clear evidence base (Comments 6 and 9). Three respondents explicitly agreed with the way the FDA proposes to establish a MID (Comments 14, 25 and 38). Some respondents advocated an anchor-based approach (Comments 8, 9, 14 and 41). Several respondents agreed on the additional value of input from patients or the clinical and research communities (Comments 25, 26, 29, 37 and 41), not using mathematical procedures alone (Comment 34). Although Sloan et al.[17] believe that combining these perspectives is favourable, they recognise that an even broader notion of HRQOL can be useful for the determination of clinically significant changes. Clarification is needed if MID refers to a 'between-group' change, a 'within-group' change or a 'within-patient' change (Comments 8, 12, 30, 35 and 36). The question arose as to when to use which approach: the MID or the responder analysis approach (Comment 15) and which one is the preferred methodology for the FDA (Comment 22). Joly et al.[18] note that the majority of advanced cancer RCTs (81%), published from 1994 to 2004, compute group differences. Nevertheless, they advocate defining a palliative response (i.e. observing the individual responder proportion), since HRQOL is the perception of an individual and not of a group. Finally, other respondents argued for the FDA to adopt a flexible approach towards the development of MID approaches (Comments 18, 21, 23 and 43), and some respondents noted that the MID can be influenced by many factors such as patient characteristics, the degree of the disease severity and finally how effective the therapy is (Comments 11, 14 and 39).

3.8. Blinding and randomisation related to PRO studies

According to the FDA (Section V A, p.23), open-label studies, in which patients and investigators are aware of the assigned therapy, are rarely credible. Open-label studies are also not recommended by the EMEA (p. 5). Many respondents challenged the FDA view: a majority believe that blinding in oncology clinical trials is hardly feasible in some circumstances (Comments 11, 12, 18, 22, 25, 32, 38, 39, 41 and 43) and required flexibility here (Comment 21). Current open-label studies are not by definition believed to provide invalid data (Comments 20, 23, 24, 26 and 43). It is even stated that non-blinded, naturalistic trials may give rise to more valid estimates than rigorously blinded or open trials.[19]

One of the five sources of bias as to why, up to date, HRQOL-based efficacy claims were disapproved by the FDA is believed to be the lack of randomisation.[20] In general, the procedure of randomisation seems to be commonly applied in the current oncology RCTs. Respondents and the EMEA released no significant statements.

3.9. Statistical concerns

A SAP should address the methodology of handling the missing data: a range of methods for doing this are listed by the FDA (Section VI D, p. 29). Also, the EMEA stressed the importance of discussing the missing data in the study design (p. 4). Many respondents replied in various ways, e.g. addressing inadequacy of worst-case scenario (Comments 18, 34, 37 and 43); complete case analysis (Comment 13); imputation methods (Comments 34, 36 and 41); the importance of advance methods such as mixture models or joint/shared parameter models (Comment 41) and the prevention of missing data by means of electronic collection by PRO methods (Comments 17, 28 and 44). Although a statistical correction can be applied, prevention through careful study design and execution is believed to be the best approach,[1] as well as ensuring, where possible, that the reasons for the missing data are captured. At least one respondent agreed on flexibility (Comment 18) in dealing with analysis. In addition, there was agreement about the pre-specification of methods in the SAP or protocol (Comments 23 and 38).

Further, the pre-specification of a sequence of testing, or a hierarchy of comparisons that need to be satisfied before others are considered for testing, is recommended in order to control for substantial increases in type I error (FDA, Section VI B, p. 27). According to the EMEA, multiplicity in PRO assessment may be overcome by the pre-specification of a subset of HRQOL domains, correction of *p*-values, hierarchical testing or global test procedures (p. 5). Three of five respondents advocated greater flexibility with respect to *post-hoc* or *ad-hoc* analyses. Such analyses are believed to offer additional value in identifying unanticipated patient benefit and in better clarifying results (Comments 15, 21 and 43). Other comments concerned the wish for clearer guidance (Comments 13 and 20).

The sample size used in a trial should depend on several factors, e.g. the number of end-points and the decision rule for success (FDA, Section VI B, p. 27). Generally, according to the EMEA (p. 5), the number of patients necessary to support the change in the primary end-point is sufficient to test for PRO change. However, Joly et al.[18] believe that most studies are not powered for PROs, since sample sizes are typically based on only one end-point. Three of five respondents noted that a sample size should be driven by primary PRO scales when PRO studies are conducted (Comments 32, 39 and 41).

3.10. Modification of instruments

The FDA announced that revised instruments should be viewed as different from the original, and therefore additional validation studies are recommended (Section IV D, p. 20). The EMEA guidance contains no statements on this topic. Many respondents argued that this FDA requirement is too restrictive (Comments 11, 14, 15, 23, 26, 35, 41 and 43) and may lead to the stagnation of instrument development (Comment 40).

Re-validation studies were found to be unnecessary in the case of relatively minor modifications, such as revisions in wording and differences in disease severity levels (Comments 1, 9, 12, 19, 24, 28, 32, 34, 39 and 44). Cognitive debriefing tests might be sufficient for these relatively small changes (Comments 11 and 34). It was advised that the level of modification should provide the best guide to the extent of re-validation that is required (Comments 13, 18, 20, 25 and 29).

3.11. *Translations*

The FDA guidance on PRO translations recommends instrument developers provide evidence that the methods and results of the translation process were adequate to warrant the validity of the responses. Accepted standards for translation and cultural adaptation must be applied to support their validity (FDA, Section IV D 5, p. 21). The EMEA released no translation statements. Respondents questioned what the generally accepted standards are (Comments 15 and 25), and expressed a wish for additional translation guidance (Comment 1). Five respondents doubted whether full validation should be required for each new translation (Comments 11, 22, 34, 36 and 40). Recently, Acquadro et al.[21] have stated that a multi-step approach gives rise to high-quality translation.

4. Discussion

In oncology phase III RCTs and registration trials, PROs are increasingly used for providing information about HRQOL in patients who undergo new treatments. Both the FDA and EMEA increasingly appear to be willing to accept PROs in support of medicinal labelling claims or in the evaluation of medical products such as cancer drugs.

The views of the FDA on PROs could be described as extensive, with detailed requirements and a restrictive nature. The EMEA has provided more global statements and broad advice which suggests that researchers should use the best available evidence and current standards in their trials. Another point of difference in PRO guidance from both agencies is the FDA's emphasis on the need for end-points that reflect direct consequences from disease and treatment and on the requirement for simple and easy to measure end-points such as symptoms, whereas the EMEA's focus goes beyond symptoms and includes HRQOL. Despite this, no major fundamental contradictions have appeared.

Although the attention on PROs has increased, PRO end-points have infrequently contributed to oncology product-approval to date. This was found in studies addressing regulatory approvals of oncology products with PRO (HRQOL) statements in their label claims.[22,23]

Also, former approvals with PRO claims do not reflect the current FDA and/or EMEA requirements or accepted current thinking. Consequently, no insight can be drawn from these earlier approvals, and therefore, they should not serve as examples for future submissions with PRO components.

Our review takes into account FDA and EMEA regulations as well as the perspectives of other key stakeholders, e.g. academia and pharmaceutical sectors. The guidance shows that it is important to provide sufficient documentation to support the PRO submission. Furthermore, systematic and ongoing correspondence with the regulatory authorities during the development process of a trial design and/or a PRO questionnaire is of major importance. Also, well-defined and hypothesis-driven end-points should be chosen that clearly reflect treatment benefit for the FDA. While the FDA appears to focus much on specific end-points such as symptoms, the EMEA appears more likely to accept domains such as overall health-related quality of life and functioning domains. Several issues remain unclear including the need for the FDA end-point model, describing the relationships among end-points, and a conceptual framework. Patient input in item generation should be demonstrated, but the extent to which interviews must be documented and the exact number of patients incorporated are unclear. The recall period chosen in the FDA guidance was heavily criticised by respondents. A substantial section of the research community argues that the FDA guidance is too prescriptive on this issue and is lacking evidence for much of the recommendations related to the recall period.

Regulatory agencies such as the FDA are likely to criticise the ability of the questionnaire to support certain label claims. Therefore, care should be taken with the psychometric evidence of each single scale by showing its validity and its ability to independently support a claim. Since improvement in all domains is frequently unrealistic, sponsors should propose specific claims, not merely broad claims, pre-specified in the protocol or the SAP. Several methods can be used in order to generate a MID, since the regulatory guidelines maintain flexibility in this matter. A preference exists to involve opinion from the clinical perspective, as supported by the EMEA and respondent statements. Therefore, an approach that integrates clinical and mathematical input should find support from the regulatory agencies. An acceptable statistical practice includes the pre-specification in the SAP of the way the missing data will be handled, the plan for coping with multiplicity and issues concerning sample size. Translation procedures required by the FDA PRO guidance are not explained in great detail, but following international standards should be adequate.

The FDA PRO regulatory guidelines are found to be stringent for well-established PRO or HRQOL measures with a long history of effective use and significant evidence of real world validity: it is questioned to what extent the existing questionnaires are supposed to meet the draft regulations, even with decades of supporting evidence (Comments 9, 15, 19, 23, 24, 26, 30, 34 and 37). An example of a widely used European questionnaire is the EORTC QLQ-LC13, a 13-item lung cancer-specific module – developed alongside the EORTC QLQ-C30[24] – capturing cancer-associated symptoms and therapy side-effects.[25] The generation of these measures included patient input, i.e. a certain number of patients for several development phases, and ensured population representation. Overall, they were developed to standards the FDA and EMEA note.[24,25] However, given the current guidelines, the FDA may incorrectly question the EORTC QLQ-C30 and the lung module and other major, well-established tools (e.g. FACT) given they were developed decades before the new guidelines were issued. Not all documentations the FDA may request will be maintained by all established instrument developers. In the

case of the EORTC QLQ-C30, there were no concerns relating to recall bias in thousands of trials involving 10,000s of patients carried out over several decades.[26] Hence, the FDA must take a much more liberal view of its guidance on such factors as present state assessment and recall period.

Our review has limitations. Specifically, qualitative evaluations of the data were made on the basis of submitted letters to the FDA web site. Some respondent comments may have included vague statements, and therefore are difficult to interpret or to include in this review. The use of two independent reviewers and a third arbitrator has limited the potential for bias, although an element of subjective interpretation was required. Furthermore, due to word limit restrictions, only the major themes were addressed.

In conclusion, broadly, oncology researchers and clinicians have welcomed the FDA PRO draft guidance and the EMEA HRQOL reflection paper. These documents are considered important steps towards the acceptation and appreciation of patient viewpoint, and the creation of significant evidence in the drug approval process. Nevertheless, continued dialogue and future FDA PRO guidance improvements on the key methodological issues raised in our review will help make PROs an important element in the fight to improve patients' HRQOL.

Conflict of interest statement

EORTC and Andrew Bottomley have received an unrestrictive academic research grant funding from AstraZeneca, UK. Dave Jones is an employee of AstraZeneca. Lily Claassens' Fellowship was in part funded by an academic grant from AstraZeneca.

Acknowledgements

We thank Dr. Bhash Parasuraman (AstraZeneca, USA) for her comments on our paper.

This research was funded by an educational academic grant from AstraZeneca, UK. In addition, part of the publication was supported by Grant Numbers 5U10 CA011488-37–5U10 CA011488-38 from the National Cancer Institute (Bethesda, Maryland, USA). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute.

Appendix A. Supplementary material

Supplementary data (Table 1) associated with this article can be found, in the online version, at doi:10.1016/j.ejca.2008.09.032, or at http://groups.eortc.be/qol/qolu_activities.htm.

REFERENCES

1. Bottomley A. Developing clinical trial protocols for quality of life assessment. *Appl Clin Trials* 2001;**10**:40–4.
2. Lipscomb J, Gotay CC, Snyder C. Introduction to outcomes assessment in cancer. In: Lipscomb J, Gotay CC, Snyder C, editors. *Outcomes assessment in cancer: measures, methods, and applications*. Cambridge, UK: Cambridge University Press; 2005. p. 1–14.
3. Bottomley A, Aaronson NK. European Organisation for Research and Treatment of Cancer International perspective on health-related quality-of-life research in cancer clinical trials: the European Organisation for Research and Treatment of Cancer experience. *J Clin Oncol* 2007;**25**(32):5082–6.
4. Lipscomb J, Gotay CC, Snyder CF. Patient-reported Outcomes in cancer: a review of recent research and policy initiatives. *CA Cancer J Clin* 2007;**57**:278–300.
5. Food and Drug Administration. Guidance for industry. Patient-reported outcome measures: use in medical product development to support labeling claims. US Food and Drug administration website, 2006 [cited 2007 April]. <http:www.fda.gov/cder/guidance/5460dft.pdf>.
6. European Medicines Agency. Committee for medicinal products for human use (CHMP). Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQL) measures in the evaluation of medicinal products. European Medicines Agency website, 2005 [cited 2007 April]. <http://www.emea.europa.eu/pdfs/human/ewp/13939104en.pdf>.
7. Burke L, Rock EP, Powers JH. Patient-reported outcome instruments: overview and comments on the FDA draft guidance. US Food and Drug administration website, 2006 [cited 2007 Oct.]. <http://www.fda.gov/cder/present/DJA2006/Bruke_Rock.ppt>.
8. Lenderking W, Stewart M, Merikle E. Coping with the validation requirements of the FDA PRO guidance: the intersection of science and practice. In: *Conference proceedings, 14th annual conference of the international society for quality of life research, Toronto, Canada*; 2007. p. 8.
9. Bolton JE. Accuracy of recall of usual pain intensity in back pain patients. *J Pain* 1999;**83**:533–9.
10. Jamison RN, Raymond SA, Slawsby EA, McHugo GJ, Baird JC. Pain assessment in patients with low back pain: comparison of weekly recall and momentary electronic data. *J Pain* 2006;**7**(3):192–9.
11. Middel B, Goudriaan H, de Greef M, Stewart R, van Sonderen E, Bouma J, et al. Recall bias did not affect perceived magnitude of change in health-related functional status. *J Clin Epidemiol* 2006;**59**(5):503–11.
12. Erskine A, Morley S, Pearce S. Memory of pain: a review. *Pain* 1990;**41**(3):255–65.
13. Stone AA, Broderick JE, Shiffman SS, Schwartz JE. Understanding recall of weekly pain from a momentary assessment perspective: absolute agreement, between- and within-person consistency, and judged change in weekly pain. *Pain* 2004;**107**:61–9.
14. Brauer C, Thomsen JF, Loft IP, Mikkelsen S. Can we rely on retrospective pain assessments? *Am J Epidemiol* 2003;**157**:552–7.
15. Gendrau M, Hufford MR, Stone AA. Measuring clinical pain in chronic widespread pain: selected methodological issues. *Best Pract Res Clin Rheumatol* 2003;**17**(4):575–92.
16. Revicki DA, Gnanasakthy A, Weinfurt K. Documenting the rationale and psychometric characteristics of patient reported outcomes for labeling and promotional claims: the PRO Evidence Dossier. *Qual Life Res* 2007;**16**:717–23.
17. Sloan JA, Frost MH, Berzon R, et al. The clinical significance of quality of life assessments in oncology: a summary for clinicians. *Support Care Cancer* 2006;**14**:988–98.
18. Joly F, Vardy J, Pintilie M, Tannock IF. Quality of life and/or symptom control in randomized clinical trials for patients with advanced cancer. *Ann Oncol* 2007;**18**(12):1935–42 [Epub 2007].

19. Marquis P, Arnould B, Acquadro C, Roberts WM. Patient-reported outcomes and health-related quality of life in effectiveness studies: pros and cons. *Drug Develop Res* 2006;**67**:193–201.
20. Rock EP, Scott JA, Kennedy DL, Sridhara R, Pazdur R, Burke LB. Challenges to use of health-related quality of life for Food and Drug Administration approval of anticancer products. *J Natl Cancer Inst Monogr* 2007(37):27–30.
21. Acquadro C, Conway K, Hareendran A, Aaronson N. European Regulatory Issues and Quality of Life Assessment (ERIQA) Group Literature review of methods to translate health-related quality of life questionnaires for use in multinational clinical trials. *Value Health* 2008;**11**(3):509–21. Epub 2007.
22. Rock EP, Kennedy DL, Furness MH, Pierce WF, Pazdur R, Burke LB. Patient-reported outcomes supporting anticancer product approvals. *J Clin Oncol* 2007;**25**(32):5094–9.
23. Coombs JH, McBurney CR, Gondek K, Copley-Merriman K. Evidence of patient reported outcomes in labelling for oncology products: evaluation of United States and European labels. *Qual Life Newslett* 2002:29.
24. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993;**85**(5):365–76.
25. Bergman B, Aaronson NK, Ahmedzai S, Kaasa S, Sullivan M. The EORTC QLQ-LC13: a modular supplement to the EORTC core quality of life questionnaire (QLQ-C30) for use in lung cancer clinical trials. EORTC study group on quality of life. *Eur J Cancer* 1994;**30A**(5):635–42.
26. Bottomley A. The journey of health-related quality of life assessment. *Lancet Oncol* 2008;**9**(9):906.