Patient-reported outcomes in daily clinical oncology practice: a tool for patient monitoring and quality of care assessment
Hilarius, D.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Level of Agreement between Patient Self-Report and Observer Ratings of Health-Related Quality of Life Communication in Oncology

Doranne L. Hilarius¹, Paul H.A.M. Kloeg¹, Symone B. Detmar², Martin J. Muller³, Neil K. Aaronson³

¹) Department of Pharmacy, Medical Center Alkmaar, Alkmaar, The Netherlands
²) TNO Prevention and Health, Leiden, The Netherlands.
³) Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands.

Patient Education and Counseling 2007; 65: 95–100
Abstract

Objective
To determine the level of agreement between patients and observers regarding the frequency with which health-related quality of life topics are discussed during outpatient clinical oncology visits.

Methods
The sample (n = 50) consisted of a consecutive series of cancer patients undergoing chemotherapy. Both the patients and observers used a checklist to report which HRQOL topics had been discussed during the consultation. Percentage of agreement, kappa and prevalence-adjusted kappa statistics were calculated.

Results
The percentage agreement between patients’ and observers’ ratings was generally high, ranging from 74% for fatigue to 96% for sleep problems and cognitive functioning. The average percentage of agreement over the 13 HRQOL topics rated was 86%. Cohen’s kappa varied between 0.41 (for pain) and 0.78 (for sleep problems). Prevalence-adjusted kappa’s were generally higher, ranging from 0.48 (for fatigue) to 0.92 (for sleep problems and social functioning). The average Cohen’s kappa and prevalence-adjusted kappa over the 13 HRQOL topics were 0.56 and 0.71, respectively. Level of agreement was not found to vary significantly as a function of patients’ background characteristics.

Conclusion
Oncology patients’ self-reports of the HRQOL-related topics discussed during outpatient chemotherapy visits are in reasonably close agreement with those provided by observers.

Practice implications
Our results indicate that the patient is a legitimate source of information about the HRQOL-related content of medical encounters, and thus can be used in communication studies where the collection of observational data (e.g., via audio- or videotaping) is either too costly or logistically impractical.
Introduction

The most optimal way to evaluate health care communication at the individual level is to make use of audiotapes or videotapes. However, audiotaping or videotaping conversations may not always be feasible for either technical or privacy reasons. For example, when treatment is given in open oncology wards where more than one person is treated at the same time, it may be difficult to isolate the relevant conversations to be recorded and there may be concerns with privacy and confidentiality. In addition, recording and evaluating tapes can be costly and time-consuming.

Alternative methods used for evaluating communication include medical chart reviews and patient self-report. Although medical record audits may be very useful in documenting observable behaviors linked to communication (e.g., medication prescription and referrals), they are less helpful in evaluating more general communication behavior, such as inquiries and advice about symptoms and functional limitations, as these are typically not noted systematically in the medical record.

Patient self-report data are frequently used to evaluate communication. Collecting self-report data is relatively efficient and low cost. However, such self-report data also have their limitations. Patients’ recall of conversations can be misinformed by lapses in memory and may be distorted when sensitive topics are at issue. In addition, patients’ worries and concerns may limit their ability to understand and to remember details of their medical visit. For these reasons, although they are widely used, the validity of patient self-report measures of health care communication can be questioned when used as the primary or only source of such information. Only a few studies have directly compared patients’ self-reports with audio- or videotapes of their medical visits (for a general review, see Boon and Stewart). Several studies report that patients tend to overestimate the occurrence of their physicians’ counseling regarding alcohol and tobacco habits.

Dimatteo et al. have reported that patients overestimate physicians’ counseling behavior, but underestimate the discussion of medications.

Over the past decade, communicative exchanges between health care providers and patients have become an area of increasing interest for research in oncology. An essential condition for optimal cancer treatment is that physicians, nurses and other ancillary health care providers are able to communicate effectively with their patients in order to obtain as complete a picture as possible of the patients’ physical and psychosocial health status.

Health-related quality of life (HRQL) considerations have come to play an increasingly important role in selecting among
available treatment options and in monitoring the effects of treatments over time\textsuperscript{13-15}. HRQL is typically defined in terms of physical, psychological and social functioning, and symptom experience\textsuperscript{13}. Yet, recent studies have reported that HRQL issues are often not addressed during outpatient medical visits in oncology. For example, Detmar and colleagues reported that, during outpatient palliative treatment consultations, physicians devoted 64\% of their conversation to medical/technical issues and only 24\% to HRQL issues. Importantly, in 20\textendash{}54\% of consultations in which patients were experiencing serious HRQL problems, no time at all was devoted to the discussion of these problems. In particular, patients’ fatigue and emotional functioning often went unaddressed (in 48\% and 54\% of the cases, respectively)\textsuperscript{16}.

A number of studies have investigated methods for improving HRQL-related communication. These efforts have focused primarily on training health care providers in communication skills, both as a part of the medical and nursing school curriculum, and in continuing education programs\textsuperscript{17\textendash{}20}. More recently, interventions have been developed that provide health care providers with individualized, standardized, patient self-reported HRQL information as a means of facilitating communication in daily clinical practice\textsuperscript{21,22}. In order to evaluate the effect of these efforts, the actual patterns and content of communication between patients and health care providers have to be evaluated in a systematic way.

To the best of our knowledge no studies have assessed the extent of agreement between patient self-reported and observational data on HRQL-related communications. The current study was undertaken to determine whether, in studies of patient-health care provider communication, it is possible (i.e., valid) to rely on patients’ self-reports of what actually was discussed during medical consultations. This is important because there are situations where it may not be possible to actually observe (or record) such conversations, and thus one may need to rely on the patients’ self-report. To investigate the validity of patient self-report data on HRQL-related communication, comparisons were made with data obtained via observer-raters. Given that patient characteristics may influence self-reports, we also examined whether the level of agreement between the self-reported and observed communication was associated with various sociodemographic and clinical characteristics, including age, education, sex, and treatment intent.
Methods

Study site and subjects
The study was conducted in the outpatient chemotherapy clinic of the Department of Internal Medicine of the Medical Center Alkmaar, a large community hospital located north of Amsterdam. In this clinic patients receive chemotherapy either in a private room or in the ward where a maximum of six patients are treated at the same time. Oncology nurses are responsible for the delivery of treatment, and the day-to-day management of patients.

The patient sample was composed of a consecutive series of patients with cancer who were receiving outpatient curative or palliative chemotherapy, and who were under the care of the participating nurses. Patients were excluded if they lacked basic proficiency in Dutch or exhibited overt psychopathology or serious cognitive dysfunction that would impede their being able to take part in the study.

Content of HRQL communication
A 13-item checklist was used to document the HRQL-related content of nurse–patient communication. The respondents (both the patients and the observer-raters) were instructed to indicate whether a specific HRQL topic has been discussed (dichotomous assessment; yes/no). To include the most salient aspects of HRQL for cancer patients, the checklist was designed to parallel the content areas of the EORTC QLQ-C30, a cancer specific HRQL questionnaire\(^2\). The checklist included the following topics: physical functioning, role functioning, cognitive functioning, emotional functioning, social functioning, fatigue, pain, nausea/vomiting, dyspnea, appetite loss, sleep disturbance, constipation and diarrhea.

Study procedure
For each patient, one consultation at the outpatient clinic was included. Immediately after the consultation, the patient was asked to fill in the HRQL communication checklist.

Two different procedures were followed to measure the actual (i.e., observed) content of the nurse–patient communication\(^1\). In the single patient rooms the consultations were audiotaped. The content of these consultations was analyzed by two research assistants using the same checklists as the patients. Although the interrater agreement was nearly 100%, regular meetings were held throughout the course of the study to identify any problems in carrying out the rating, and to achieve consensus where uncertainty in ratings existed\(^2\). On the ward, where multiple patients were being treated at the same time, it was not possible to record “isolated” conversations between the nurse and participating patients due to ambient conversation and noise, and because the nurses were treating several patients simultaneously. Therefore, in the ward setting, consultations were observed and directly scored by the research assistant, using the same checklist as the patients.
Data analysis

Descriptive statistics were used to describe the characteristics of the study sample and the frequency with which various HRQL topics were discussed during the consultations. The frequency with which HRQL topics were discussed as reported by the patients was compared with that based on the ratings of the research assistants, and the level of agreement between these two sources was assessed in two ways: (1) simple percentage of agreement; (2) Cohen’s kappa statistic, a chance-corrected concordance measure.

The level of agreement was established for all 13 HRQL topics included in the checklist.

In some cases, use of the kappa statistic as a chance-corrected measure of agreement can be misleading. Specifically, kappa can be affected by bias and/or the prevalence of the observations. Bias occurs when two observers differ in their assessment of the frequency of the observation, and the marginal distributions for the raters are unequal. When the observed prevalence of responses in one of the two available categories is low, there is a prevalence problem. It will result in high percentage agreement, but low kappa values. Byrt et al. have proposed a bias and prevalence-adjusted kappa. The current results will be reported based both on Cohen’s kappa and on the prevalence-adjusted kappa.

The conventional interpretation of kappa is as follows: $0.0-0.20 = \text{poor agreement}$, $0.21-0.40 = \text{fair agreement}$, $0.41-0.60 = \text{moderate agreement}$, $0.61-0.80 = \text{good agreement}$, and $0.81-1.00 = \text{excellent agreement}$. In general, kappa values of $0.60$ or higher are considered acceptable for self-report measures.

Separate analyses were conducted to determine if the level of agreement between the raters and the patients varied significantly as a function of the method used by the research assistants (i.e., ratings of audiotapes versus ratings based on direct observation of the nurse-patient encounters).

The chi-square statistic was used to examine whether level of agreement varied significantly as a function of patient background and clinical characteristics. For this analysis, patients were categorized in groups based on sex, age ($\geq 60$ and $< 60$ years), educational level (lower level high school, middle level high school or university), and treatment intent (palliative or curative).
Results

Sample characteristics
A total of 57 patients were invited to participate in the study, of whom 50 agreed (response rate = 88%). Of the seven non-participants, four declined due to very poor physical condition, one reported insufficient interest or lack of time, and two had other reasons not specified.

The demographic and clinical characteristics of the patient sample are shown in Table 1. Forty-six percent was female, with a mean age of 59 years (range = 27–79 years). Patients had a wide range of cancer diagnoses, with breast cancer being the most prevalent (36%). Half of the patients received treatment with a palliative intent. Forty-two percent of the patients was treated in a single patient room, and 58% on the ward. During nearly 80% of the visits someone, most often the spouse, accompanied the patient.

Level of agreement
As an initial step we investigated whether the procedures used for rating the content of the communication (i.e., based on audiotapes versus direct observations of consultations) was related significantly to the level of agreement observed between patients and observers. No significant differences were observed in levels of patient–observer agreement as a function of method. Therefore, all further analyses were carried out on the total study sample.

Table 2 shows the frequency with which the various HRQL topics were discussed according to the patients and to the observers, as well as the level of agreement between these two sources of information (percentage agreement, kappa, and prevalence-adjusted kappa). The five topics discussed most frequently, based on both patients’ and observers’ ratings, respectively, were nausea and vomiting (62% and 58% of the cases), appetite loss (40% and 32%), fatigue (38% and 40%), constipation (34% and 28%) and pain (26% and 24%). Those topics discussed least frequently were cognitive functioning (0% and 4%), dyspnea (4% and 6%), and sleep problems (10% and 10%).

The percentage agreement between patients’ and observers’ ratings was generally high, ranging from 74% for fatigue to 96% for sleep problems and cognitive functioning. The average percentage of agreement between patients and observers across the 13 HRQL topics that were rated was 86%.

Cohen’s kappa varied between 0.41 (for pain) and 0.78 (for sleep problems), indicating that the level of agreement was in the moderate to good range. As expected, the prevalence-adjusted kappa’s were generally higher (range = 0.48–0.92). These higher kappa values reflect primarily the adjustment for low prevalence (e.g., for dyspnea and cognitive functioning). The average Cohen’s kappa and prevalence-adjusted kappa across the 13 HRQL topics evaluated were 0.56 and 0.71, respectively.
Table 1.

<table>
<thead>
<tr>
<th>Demographic and clinical characteristics of the patient sample (n = 50)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Gender</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Education</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Cancer diagnosis</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Intent of treatment</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Treatment location</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Accompanying patient</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Number of treatment cycles</strong></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Discussion and conclusion

Discussion
This study was undertaken to investigate the level of agreement between patients and observers with regard to the frequency with which HRQoL-related topics are discussed in daily clinical oncology practice. An additional objective was to determine whether the level of agreement between patients and observers varied systematically as a function of the sociodemographic and clinical characteristics of the patients.

As has been suggested in the literature\(^{16}\), acute and more observable symptoms (e.g., nausea and vomiting, constipation) were reported as having been discussed most often. This is encouraging, in that emesis remains one of the most distressing side-effects of chemotherapy\(^{27}\). It is noteworthy, however, that fatigue, a chronic and less observable symptom, was discussed in more than one-third of the cases. This may reflect the increasing attention that has been paid to fatigue, in general, and in the context of cancer treatment, in particular\(^{28}\). Conversely, although the recent research literature has suggested that neurocognitive symptoms may be more common among patients receiving chemotherapy than had been suspected\(^{29}\), our data indicate that such problems are still rarely discussed explicitly during outpatient chemotherapy consultations.

The average percentage of agreement between patients and observers was high (86%; range 74–96%). The average value of kappa, a
Table 2.

<table>
<thead>
<tr>
<th>HRQL topic</th>
<th>Frequency discussed according to observer(^a) (%)</th>
<th>Frequency discussed according to patient(^a) (%)</th>
<th>Cohen's kappa</th>
<th>Prevalence-adjusted kappa</th>
<th>Agreement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain</td>
<td>26</td>
<td>24</td>
<td>0.41</td>
<td>0.56</td>
<td>78</td>
</tr>
<tr>
<td>Fatigue</td>
<td>38</td>
<td>40</td>
<td>0.45</td>
<td>0.48</td>
<td>74</td>
</tr>
<tr>
<td>Dyspnea</td>
<td>4</td>
<td>6</td>
<td>0.37</td>
<td>0.88</td>
<td>94</td>
</tr>
<tr>
<td>Nausea/vomiting</td>
<td>62</td>
<td>58</td>
<td>0.58</td>
<td>0.60</td>
<td>80</td>
</tr>
<tr>
<td>Sleep problems</td>
<td>10</td>
<td>10</td>
<td>0.78</td>
<td>0.92</td>
<td>96</td>
</tr>
<tr>
<td>Appetite loss</td>
<td>40</td>
<td>32</td>
<td>0.57</td>
<td>0.60</td>
<td>80</td>
</tr>
<tr>
<td>Constipation</td>
<td>34</td>
<td>28</td>
<td>0.58</td>
<td>0.64</td>
<td>82</td>
</tr>
<tr>
<td>Diarrhea</td>
<td>18</td>
<td>12</td>
<td>0.77</td>
<td>0.88</td>
<td>94</td>
</tr>
<tr>
<td>Physical functioning</td>
<td>24</td>
<td>18</td>
<td>0.70</td>
<td>0.80</td>
<td>90</td>
</tr>
<tr>
<td>Role functioning</td>
<td>24</td>
<td>18</td>
<td>0.58</td>
<td>0.72</td>
<td>86</td>
</tr>
<tr>
<td>Social functioning</td>
<td>26</td>
<td>12</td>
<td>0.52</td>
<td>0.68</td>
<td>84</td>
</tr>
<tr>
<td>Cognitive functioning</td>
<td>0</td>
<td>4</td>
<td>b</td>
<td>0.92</td>
<td>96</td>
</tr>
<tr>
<td>Emotional functioning</td>
<td>20</td>
<td>18</td>
<td>0.42</td>
<td>0.64</td>
<td>82</td>
</tr>
<tr>
<td>Mean</td>
<td>0.56</td>
<td>0.71</td>
<td></td>
<td></td>
<td>86</td>
</tr>
</tbody>
</table>

\(^a\) Frequencies were calculated by dividing the number of consultations during which a topic was discussed (according to the observers and the patients, respectively) by the total number of consultations.

\(^b\) Kappa could not be calculated due to empty cell.
chance-adjusted statistic, was 0.56, indicating a moderate level of agreement. When correcting for low prevalence rates for some HRQL topics, the average adjusted kappa was 0.71, and the adjusted kappa for 11 of the 13 HRQL topics met the 0.6 threshold generally recommended as criterion for an acceptable level of agreement. Only for fatigue and pain was this 0.6 threshold not reached.

No consistent, statistically significant differences in levels of patient-observer agreement were found as a function of patients’ sociodemographic characteristics or of the intent of the treatment (i.e., curative versus palliative). This suggests that patients’ self-reports of the HRQL-related content of medical encounters is equally valid for men and women, patients of different ages, with different education levels, and with different stages of disease. However, given the relatively small sample size of the current study, these findings need to be replicated.

**Conclusion**

In conclusion, the results of this study suggest that oncology patients’ self-reports of the HRQL-related topics discussed during outpatient chemotherapy visits are in fairly close agreement with those provided by observers. Collecting patient self-report data in communication research is typically less time-consuming and less costly than observer-based data.

**Practice implications**

Our results indicate that patients are a legitimate source of information on the HRQL-related content of medical encounters, and thus can be used in communication studies where the collection of observational data (e.g., via audio- or videotaping) is either too costly or logistically impractical.

**Acknowledgements**

The authors wish to thank Saskia de Bie, Edith Budding and Carien de Molennaar for assisting in the data collection. We thank the patients and the nursing staff of the oncology outpatient clinic of the Medical Center Alkmaar for their willingness to participate in this study.
References


26 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.

