Published in final edited form as:

Qual Life Res. 2017 August ; 26(8): 1925–1954. doi:10.1007/s11136-017-1540-6.

Systematic review of caregiver responses for patient health-related quality of life in adult cancer care

Jessica K. Roydhouse¹ and Ira B. Wilson¹

¹Department of Health Services, Policy, and Practice, School of Public Health, Brown University, 121 S. Main Street, Providence, RI 02912, USA

Abstract

Purpose—In surveys and in research, proxies such as family members may be used to assess patient health-related quality of life. The aim of this research is to help cancer researchers select a validated health-related quality of life tool if they anticipate using proxy-reported data.

Methods—Systematic review and methodological appraisal of studies examining the concordance of paired adult cancer patient and proxy responses for multidimensional, validated HRQOL tools. We searched PubMed, CINAHL, PsycINFO and perused bibliographies of reviewed papers. We reviewed concordance assessment methods, results, and associated factors for each validated tool.

Results—A total of 32 papers reporting on 29 study populations were included. Most papers were cross-sectional ($N = 20$) and used disease-specific tools ($N = 19$), primarily the FACT and EORTC. Patient and proxy mean scores were similar on average for tools and scales, with most mean differences < 10 points but large standard deviations. Average ICCs for the FACT and EORTC ranged from 0.35 to 0.62, depending on the scale. Few papers ($N = 15$) evaluated factors associated with concordance, and results and measurement approaches were inconsistent. The EORTC was the most commonly evaluated disease-specific tool ($N = 5$ papers). For generic tools, both concordance and associated factor information was most commonly available for the COOP/WONCA ($N = 3$ papers). The MQOL was the most frequently evaluated end-of-life tool ($N = 3$ papers).

Conclusions—Proxy and patient scores are similar on average, but there is large, clinically important residual variability. The evidence base is strongest for the EORTC (disease-specific tools), COOP/WONCA (generic tools), and MQOL (end-of-life-specific tools).

Keywords

Quality of life; Caregiver; Proxy; Observer; Adult

Correspondence to: Jessica K. Roydhouse.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Introduction

Collecting patient-reported outcomes (PROs), including health-related quality of life (HRQOL), is recommended for cancer comparative effectiveness research [1]. A significant challenge to the use of patient-reported data in cancer care is that patient illness or treatment side effects may affect the ability to complete measures, resulting in unavailable or missing data. Using proxy respondents such as family members is a commonly suggested strategy to address this problem [2–4]. Proxy respondents have been used in clinical trials [5] as well as national health surveys [6–12]. However, if proxies are unable to accurately report from the patient's perspective, this can produce a misleading picture.

Previous reviews [13–15] have considered this issue across a range of diseases, including cancer, as well as proxy types (health care professionals treating the patient and family/other non-health care proxies). Most cancer-specific reviews have focused on end of life care [16, 17]. Few, if any, reviews have looked at instrument-specific issues related to proxies. Results have been summarized across instruments and diseases, and tool-specific advice is lacking. The aim of this review is to examine the extent of evidence for family or other non-health professional caregiver proxy reporting for validated multi-dimensional HRQOL tools in adult oncology, in order to guide researchers who may need or want to employ proxy reports.

Materials and methods

Search strategy

Following search strategy review by a librarian, we searched PubMed, CINAHL and PsycINFO using a combination of terms for proxy, quality of life, and cancer (Appendix 1). We also perused the bibliographies of articles selected for review to identify additional papers. The search was conducted in March–April 2015, and repeated in February 2016 to ensure no new articles were missed.

Study selection

One reviewer (JKR) selected papers for full-text review based on paper title and abstract, if available. The focus at this stage was including all possible articles and thus articles were selected for full-text review even if there was uncertainty about an article's eligibility. Abstrackr [18], an open-source screening tool for systematic reviews [19], was used to manage abstract selection and review. Duplicate citations were identified using EndNote. Commentaries or editorials about previously identified articles were considered to be duplicates.

Articles were eligible for inclusion if: (1) the population studied was adult patients with diagnosed cancer, and the proxies were not exclusively health care professionals responsible for the medical/nursing care of the patient; (2) HRQOL was evaluated using a standardized, quantitative multi-dimensional instrument validated in a cancer patient population; (3) the proxy assessed the patient's HRQOL and patient-proxy concordance of patient HRQOL was evaluated; (4) the full manuscript was available in the English language; (5) the article was original research and not an existing systematic review or meta-analysis; (6) if diseases in

addition to cancer were included, cancer-specific results for all outcomes of interest were available. Studies that used both family/non-health care professional proxies and treating physicians/nurses were eligible if the two proxy types were differentiated and results were available for the family/non-health care proxy group. After all authors (JKR, IBW) reviewed the selection criteria, one reviewer (JKR) was responsible for review and data extraction. Data extraction was validated by an experienced researcher, who reviewed 8 (25%) of the extracted articles following training on two articles. Disagreements or different conclusions were discussed. For basic study information (e.g. countries, analytic cohort size, time from treatment to questionnaire administration, etc), no disagreements occurred. The validation identified one error in questionnaire administration timing, which was corrected. No disagreements were identified regarding the extraction of means and correlation coefficients. There were minor disagreements regarding how to consider the minimal clinically important difference in two papers, but these were resolved following discussion. Study authors were not contacted.

Data extraction

A customized Google spreadsheet (gsheet) developed by one reviewer (JKR) was used for full-text decisions and data extraction. We recorded the following information in data extraction: (1) HRQOL tool; (2) study design; (3) patient clinical characteristics (cancer type, cancer stage, treatment status, setting); (4) sample size; (5) if a minimal clinically important difference (MCID) was used to interpret concordance; (6) concordance assessment methods and results for three domains: global/overall QOL, emotional, and physical; (7) factors associated with concordance. For treatment status, treatment setting and MCID, unless studies explicitly noted these, they were considered unspecified. For factors associated with concordance, if proxy type (health professional vs not) was used as a factor for agreement, this was not considered or extracted. Results for other factors were still extracted, but the co-mingling of health care and non-health care proxies noted. Significance for factors, where evaluated, was only reported for the domains of interest (QOL, physical, and emotional). Where authors did not present all factors in tabular form, or used language such as “factors included,” we considered factors to be presented elsewhere in the tables/text as evaluated.

For the three domains, if a tool did not have an explicit physical/QOL/emotional scale, we used the scale and/or item which would be most appropriate for that domain. For example, the EQ-5D does not have a specific physical domain, but it does assess mobility; we therefore used this for the physical domain. Additionally, if authors did not provide information regarding summary scales, then the most appropriate subscale within the measure that mapped to the domains of interest was evaluated. The specific items/scales for each domain of interest in the included studies are in Table 1. As indicated by Table 1, the domains of these questionnaires covered very different topics.

When possible, the limits of agreement (LOA) were calculated to identify the extent to which patient and proxy responses diverged on average. If a mean difference between patient-proxy responses and associated standard deviation (SD) was available, the LOA were calculated as $\pm 1.96 \times \text{SD}$. Where only proxy mean and SD and patient mean and SD

were available, the mean difference was calculated by subtracting the two means. For ease of presentation, all mean differences were converted to absolute values. In cases where insufficient information was provided, a simplifying assumption of no covariance (e.g. mean variance equal to the difference of the two variances, calculated as the square of the two provided SDs) was made to facilitate LOA calculations. The LOA were then calculated as $\pm 1.96 \times \text{SD}$. As this calculation is an approximation and makes assumptions, all imputed LOAs and mean differences are noted when they appear. In several cases, the authors provided mean differences but not SDs; the calculation for the SD was then undertaken as described above. These cases are also considered to be imputed and noted as such. Descriptions of study scoring were taken from the studies reporting them, study scoring manuals or relevant published papers. All analyses were conducted at the score/domain level, as appropriate; analyses of items comprising domains of interest were not undertaken.

Results

Study selection and data extraction

The database searches yielded 6614 unique (non-duplicate) abstracts. From these, 202 papers were selected for full-text review (Fig. 1). Twenty-nine papers were eligible for data extraction; the primary reasons for non-eligibility were proxy not assessing patient HRQOL ($n = 80$ papers) and the use of study-specific tools or qualitative assessments ($n = 40$ papers). A further three articles were identified from bibliographic searching, and data were extracted from a total of 32 papers. Three pairs of papers assessed the same study population, either as separate analyses or sub-studies of a main study, yielding a total of 29 unique studies/populations.

Characteristics of included studies

Most papers ($N = 20$, 63%) were cross-sectional (Table 2). Most studies came from either Europe ($N = 14$, 44%), primarily the Netherlands, or North America ($N = 13$, 41%), primarily the USA. Half of the papers included a range of cancer types; among studies focused on a single disease, prostate and brain cancer were the most common. Care settings were infrequently specified ($N = 15$, 47%). Most papers ($N = 19$, 59%) used disease-specific tools: the European Organisation for Research and Treatment of Cancer (EORTC) ($N = 11$, 34%) [33, 37–46], Functional Assessment of Cancer Therapy (FACT)/Functional Assessment of Chronic Illness Therapy (FACIT) ($N = 7$, 22%) [47–53], Prostate Cancer Specific Quality of Life Instrument (PROSQOLI) ($N = 1$, 3%) [43], and Quality of Life Index (QLI) ($N = 1$, 3%) [54]. A variety of generic tools ($N = 8$, 25%) were used, including the Short Form-36 (SF-36) ($N = 2$, 6%) [55, 56], the EuroQol five dimensions questionnaire (EQ-5D) ($N = 1$, 3%) [33], the World Health Organisation Quality of Life Assessment-Bref (WHOQOL-BREF) ($N = 2$, 6%) [36, 57], and the Primary Care Cooperative Information Project/World Organization of National Colleges, Academies, and Academic Associations of General Practices/Family Physicians (COOP/WONCA) ($N = 3$, 9%) [32, 58, 59]. Of the papers using end-of-life specific tools ($N = 6$, 19%), most employed the McGill Quality of Life (MQOL) ($N = 3$, 9%) [29, 60, 61], followed by the Spitzer Quality of Life Index (SQLI) ($N = 2$, 6%) [31, 62] and the Hospice Quality of Life Index (HQLI) ($N = 1$, 3%) [27]. For those studies not restricted to spouses/partners as proxies, spouses usually

comprised half or more of the proxies. Baseline dyad sizes were variable, ranging from 23 dyads ($N = 46$ respondents total) to 614 dyads ($N = 307$ respondents total). Attrition was frequent and substantial for longitudinal papers, ranging from 12% to 61% missing at the first follow up. Reporting is restricted to baseline data in light of this selection bias and to facilitate comparison with the majority of the papers, which were cross-sectional.

When studies discussed the instructions provided for the proxies to answer questionnaires, the instructions were to view the questions from the patient's perspective. The exception were two studies [33, 41] which explicitly evaluated both the aforementioned "proxy-patient" perspective and "proxy-proxy" perspective [33], in which the proxy approached the questions from their own perspective. To facilitate comparability, all reported estimates are from the "proxy-patient" perspective.

Methodological evaluation of included studies

Methodological reporting was inconsistent across papers. Of the papers which were not pre-treatment ($N = 29$, 91%), $N = 12$ (41% of that group) specified the timing of questionnaire administration relative to treatment or hospice admission. Most papers ($N = 17$, 53%) specified timing relative to diagnosis. Questionnaires were specified as consistently administered at the same time/on the same day for patients and proxies in $N = 13$ papers (41%), while $N = 9$ (28%) did not specify and $N = 10$ (31%) noted that they were not consistently administered at the same time. While $N = 12$ papers (38%) considered a MCID, $N = 15$ papers (47%) evaluated the factors associated with proxy/patient concordance. Reporting of missing item data was infrequent, with many papers ($N = 19$, 59%) not discussing this explicitly. None of the longitudinal papers which discussed unit/form-level missingness ($N = 9$) used imputation, instead relying on listwise deletion methods such as complete or available case analysis.

Many papers ($N = 25$, 78%) compared patient and proxy means, typically through t tests or similar approaches (e.g. Wilcoxon signed-rank test). Comparison of means using effect size was undertaken in $N = 15$ (47%) papers, and $N = 27$ (84%) papers evaluated the correlation of patient and proxy scores. When correlation was evaluated, the intra-class correlation (ICC) was the most frequently employed approach ($N = 17$, 63% of papers evaluating correlation), followed by Pearson's r ($N = 9$, 33% of papers evaluating correlation) and the weighted kappa ($N = 4$, 15% papers evaluating correlation).

Patient/proxy concordance: disease-specific instruments

For the EORTC QLQ-C30 (or its predecessor, the QLQ-C36), the mean differences between patient and proxy estimates for the three domains of interest were generally small, and patient estimates were higher than proxy estimates (Table 3). However, standard deviations were large and spanned at least 20% of the scale for both patient and proxy estimates as well as the differences between them. Patient mean scores were higher/better for all scales in all studies. In other words, proxy estimation of patient physical and emotional function, QOL and the total score was lower on average than the patient's estimation. ICCs for each of the domains suggested moderate correlation, with the strongest correlations and narrowest range for physical function. For global QOL, ICCs ranged from 0.15 to 0.64 (mean 0.46); for

physical function, the range was 0.36–0.73 (mean 0.62), and for emotional function the range was 0.14–0.62 (mean 0.47).

For the FACIT/FACT suite, just five of the seven papers provided at least one mean score of some kind, and of these just three provided means for the sub-domains of interest as opposed to the total score. As studies used a variety of disease-specific tools (FACIT-Sp, FACT-G, FACT-P, FACT-Br), the maximum total scale score ranged from 112 to 200. Patient mean scores consistently exceeded proxy mean scores, and SDs again comprised a high percentage of the scale. Higher patient mean scores once again reflected patients reporting higher levels of function/QOL, compared to what proxies reported for them. The sole exception to this was one study in which patient and proxy mean scores for QOL were the same [49]. ICCs for each domain and the total score suggested moderate correlation, although correlations were weakest and the range widest for emotional well-being. The range for functional well-being was 0.45–0.73 (mean 0.60); for total score, it was 0.42–0.62 (mean 0.53); for physical well-being, 0.37–0.72 (mean 0.53), and for emotional well-being 0.07–0.58 (mean 0.35).

Higher proxy scores, relative to patient scores, were seen for the other two disease-specific instruments, the PROSQOLI and QLI. This was not consistent for the PROSQOLI, as proxies reported higher (better) scores for the physical domain but patients reported higher (better) scores for the emotional domain. The correlation for the PROSQOLI physical domain was moderate (0.4), but weak for the emotional domain (0.12). No ICC was available for the QLI.

Patient/proxy concordance: end-of-life-specific instruments

For the MQOL, study results for the physical and psychological domains were inconsistent. There was a difference in correlations between the two studies, with one study [29] having consistently moderate correlations (0.47 for the physical domain, 0.36 for the psychological, and 0.61 for the total score), while the other [60] had low to moderate correlations (0.14 for the physical, 0.37 for the psychological and 0.28 for the total score). For the MQOL total score and psychological score, in both studies patient mean scores were higher than proxy scores, reflecting higher patient-reported QOL and psychological well-being, relative to proxy-reported scores. However, in one study the patients' mean physical scale score was higher [60], indicating patients reporting higher/better physical well-being compared to what proxies reported for them, but in another the proxy score was higher [29].

None of the papers using the Spitzer QLI or Hospice QLI assessed an ICC. For the total score, where means for each group were provided [27, 62], the patient score was higher than the proxy score. One study using the Spitzer QLI [31] provided means for individual domains (which are items in the questionnaire), but no SDs; for both of these domains, the patient mean score was higher, indicating better QOL. However, as with other measures, the LOA were wide and encompassed a substantial proportion of the scale (Table 4).

Patient/proxy concordance: generic instruments

ICCs were infrequently reported for generic instruments. Only three papers included an ICC assessment [32, 33, 59], two of which used the COOP/WONCA [32, 59] and were drawn

from the same study population. For the COOP/WONCA, all ICCs were moderate: global quality of life (0.37 and 0.48), physical (0.56 and 0.57) and emotional (0.48 and 0.48). For the EQ-5D, ICCs were low (0.29) for emotional and moderate (0.46) for physical. As noted in Table 1, the physical domain for the EQ-5D refers to the assessment of mobility, while the emotional domain refers to the assessment of anxiety/depression.

Usable mean differences were reported for the COOP/WONCA, WHOQOL-BREF and EQ-5D only (Table 5). Once again, LOA were very wide even though mean differences were relatively small (0.5 points on a 1–5 point scale, for example). For the COOP/WONCA, ICCs for all domains were moderate: 0.37 and 0.48 for global QOL; 0.56 and 0.57 for the physical domain; and 0.48 for the emotional domain.

For the COOP/WONCA, higher scores meant worse QOL, physical fitness, or feelings. Patient mean scores were lower, indicating better QOL/fitness/feelings, for all studies save one [58], where the proxy and patient mean scores were identical. In the one study using the EQ-5D [33], patient mean scores were higher than proxy mean scores, reflecting better patient-perceived QOL compared to proxy-perceived QOL. Results for the WHOQOL-BREF were less consistent. For the QOL domain, in both studies proxies reported higher/better QOL compared to patient reports, although the differences were small. For the physical and psychological domains, however, patient mean scores were higher/better.

Assessment of factors associated with concordance

Of the 15 papers evaluating factors associated with proxy/patient concordance, there are 13 unique patient populations/studies as two pairs of papers are each analyzing the same population or a sub-set thereof in different ways. In one case, one [59] is a subset of the other [32], focusing on an inpatient group when the main study assessed both inpatients and outpatients. In the second set, one study [41] conducted an additional evaluation of the role of proxy viewpoint on concordance in a group of respondents from an earlier study [42].

Most of the papers conducting this analysis used the EORTC QLQ-C30 ($N=5$), followed by the MQOL ($N=2$), the SF-36 ($N=2$) and the COOP/WONCA ($N=2$). Other evaluated tools were the FACT-G ($N=1$), WHOQOL-BREF ($N=1$), Spitzer QLI ($N=1$), EQ-5D ($N=1$) and Hospice QLI ($N=1$). Due to the sub-studies mentioned above, the number of unique studies/patient populations is lower for two measures: the EORTC ($N=4$ studies, $N=5$ papers) and the COOP/WONCA ($N=1$ study, $N=2$ papers).

Proxy factors were evaluated in 11 papers. As Table 6 shows, a range of methods and factors were used. Statistical significance was evaluated, often through t tests or correlation. Multivariable analyses were conducted in $N=4$ papers. The patient–proxy relationship was frequently considered as a factor ($N=7$ studies), usually by comparing a spouse/partner proxy to other proxies such as children or friends. Results on the effect of the patient/proxy relationship on concordance could not be obtained for a further three papers [32, 59, 60], as for that analysis they did not separate health care and non-health care proxies. One paper [57] could not evaluate this as proxy selection criteria included a restriction to spouses/partners. Two studies found statistically significant proxy–patient mean score differences relating to spousal proxies. In one study [48], spouse–patient mean score differences on the

total QOL score for the FACT were smaller than that for other proxy types. In the other [55], spouse–patient differences were statistically significant for the mental health domain of the SF-36, but non-spouse–patient differences were not significant. Finally, one study found that spousal and sibling proxies had significantly better proxy–patient concordance than parental/child proxies [56]. In this study, concordance was defined as the proxy score being within the 90% confidence interval of the patient score.

A small number of studies considered other factors relating to the patient–proxy relationship, including the length of time the proxies and patients were living together, the frequency and intensity of proxy/patient contact, the quality of the relationship and of proxy–patient communication. The one study evaluating the proxy’s perceived knowledge of patient thoughts and feelings on disease and symptoms found it to be negatively associated with concordance (e.g. better knowledge was associated with increased absolute mean differences) in both univariable and multivariable analyses, although the amount of variance explained was low [61].

Few proxy socio-demographic characteristics also had significant effects on proxy–patient concordance. Five papers evaluated the role of proxy gender, and one paper [57] could not evaluate this as only male partner proxies were included by design. In pairwise analyses, male proxies were associated with significantly worse concordance (larger absolute proxy–patient differences) for the EORTC QLQ-C30 total score [42], but better concordance (less disagreement) for the emotional function scale [46] for the same questionnaire. In both of these studies, these associations were not significant in multivariable analyses. Proxy–patient response agreement levels were similar across gender groupings in another study [59]. Older proxies were associated with significantly larger absolute differences for the total score in pairwise analyses, but not in multivariable analyses [42]. Other proxy factors such as education and employment were not significant, although they were also infrequently examined ($N = 3$ and $N = 2$ papers, respectively).

Proxy health characteristics and caregiving burden had significant, albeit infrequently examined and inconsistent, associations with proxy–patient concordance. Worse proxy HRQOL [42] or the presence of a comorbidity [61] was associated with significantly larger absolute mean differences for the EORTC [42] and MQOL [61]. Higher caregiving burden as measured by the Caregiver Reaction Assessment Tool was associated with worse concordance (larger proxy–patient differences) in two studies in both pairwise and joint multivariable analyses [46, 61]. In one study, this association was found for the domain of impact of caregiving on proxy health [61], whereas for the other the domains of caregiver self-esteem and lack of family support were significant [46]. The effect for the MQOL was seen for the total score (the only MQOL score considered in that paper), whereas in the other paper effects were seen on the emotional function and QOL scales of the EORTC. Caregiver burden, assessed as the frequency of feeling burdened, was not significantly associated with absolute or directional proxy–patient score differences in another study. However, in that study higher reported caregiving intensity had significant negative associations with both measures (e.g. higher intensity associated with larger differences/worse concordance) in pairwise analyses, and a significant association in multivariable models for directional difference [42].

The two papers [33, 41] evaluating the impact of proxy viewpoint in instructions and questions had inconsistent results and neither the proxy–proxy nor the patient–proxy view was associated with better concordance on all evaluated scales. In one study, significant mean score differences between perspectives were found for the emotional and physical function scales of the EORTC, as well as for the EQ-5D VAS score [33]. For all these scales, proxy–patient differences were smaller under the proxy–patient perspective. However, in the second study there were no significant between-perspective differences, although both patient and proxy perspectives showed significant mean differences as indicated by *t* tests for emotional and physical function as well as QOL [41]. In terms of impact of proxy characteristics on how proxy perspectives affected concordance, this was only examined in one study [33]. Limited proxy health literacy was associated with more consistency across perspectives (e.g. smaller differences between the two proxy perspectives tested) for physical function, but proxies with depressive symptoms had significantly lower odds of exact agreement (100% concordance) between perspectives for the EQ-5D VAS than proxies without depressive symptoms.

Patient characteristics were more frequently assessed, although again results were inconsistent, and not significant in many of the assessing studies. As with the evaluation of proxy-specific factors, a range of methods and factors were considered. Although patient age was frequently examined (*N* = 8 papers), it was significant for both univariable and multivariable analyses in only one study [61], where older age was associated with smaller mean absolute differences in the MQOL total score. However, older patients had significantly larger absolute proxy–patient differences, relative to younger patients, for the EORTC total score, although the association in multivariable analyses was not significant [42] (Table 7).

Results for cognitive function were likewise inconsistent. Significantly more proxy–patient discrepancies were found for patients with mental confusion, compared to those with normal function [44]. In addition, significantly lower proxy scores were also found on the physical and emotional function and QOL scales for patients with mental confusion, whereas no significant proxy–patient differences were seen in the group of patients with normal function. Furthermore, worse (lower) proxy–patient agreement, as indicated by both exact agreement (same response category) and approximate agreement (within one response category) was seen for patients with mental confusion in the same study, compared to patients without. In this study, mental confusion was measured using a “newly developed 10-item instrument” administered as part of a neurologic exam. However, the opposite result was seen for patients with brain metastases (compared to those without) on the MQOL total score; the presence of brain metastases was associated with significantly smaller absolute mean differences between patients and proxies [61]. Finally, another study using the MQOL found that patients with poorer function, as measured by the Short Orientation-Memory-Concentration Test, had significantly smaller mean proxy–patient differences (e.g. better agreement) on both the psychological and total scores of the MQOL [60].

Patient performance status was likewise evaluated using a range of tools, and considered in five papers [32, 42, 44, 59, 60]. It was not significant in one paper [60]. One study used a *t* test to compare agreement among groups defined by the patient’s performance status, as

measured by the Eastern Cooperative Oncology Group (ECOG) score. Poorer/worse performance status was associated with significantly smaller differences for physical function and QOL, but significantly larger differences for emotional function [32]. While significance was not considered in another study [59], larger response agreement discrepancies were seen in patients with a status of ECOG 2 (poor), compared to ECOG 0 (good). Worse ECOG performance status was significantly associated with significantly larger absolute proxy–patient differences on the total EORTC QOL score in another study [42], although this was not the case for multivariable analyses.

Other aspects of patient health also provided inconsistent results; for example, more depressive symptoms as measured by the Beck Depression Inventory were associated with significantly smaller proxy–patient differences for the psychological scale of the WHOQOL-BREF [57], but poorer mental health as measured by the Mental Health Inventory-5 was associated with significantly larger absolute differences on the EORTC total score [42]. It was difficult to discern the effect of treatment as nearly half of the papers evaluating the correlates of proxy–patient concordance assessed either hospice or pre-treatment patients or patients receiving the same treatment ($N = 7$ of $N = 15$). Patient QOL was considered infrequently; in one proxy–patient correlations in groups defined by patient HQLI score relative to the median score were examined [27]. Significant proxy–patient correlations were found in the higher scoring group, and the correlation in the lower scoring group was both smaller and not significant. One other study looked at patient QOL using scatter plots, and found fewer proxy–patient differences at either extreme end of the patient QOL distribution; in other words, proxy–patient agreement was better if patients had either very good or very bad QOL, but worse if patients had moderate QOL.

Finally, relatively few papers reported on the quality of their models or approaches for evaluating the association between patient or proxy factors and concordance. Those that did noted the relatively low percentage of variance explained, which was $<20\%$ [42, 61].

Discussion

This paper has four main findings. First, group-level patient–proxy concordance as measured by mean differences and correlation is generally good for multidimensional HRQOL tools, although proxies consistently underestimate patient QOL and physical and emotional function compared to patient estimation of these outcomes. Second, despite the good concordance, there was substantial residual variability, suggesting the need to minimize this variability using appropriate adjustment factors. Third, more work is needed to identify additional adjustment factors using standardized measurement approaches and carefully designed protocols. Fourth, while additional work is needed, several tools currently have the strongest evidence base in terms of the extent of information available for concordance and adjustment factors. These are the EORTC QLQ-C30 (disease-specific tools), the MQOL (end-of-life-specific tools), and the COOP/WONCA (generic tools).

The finding of adequate group-level patient–proxy concordance and small mean differences between patient and proxy scores is consistent with previous reviews [15–17, 63]. The large levels of residual variability highlight the importance of identifying and using factors to

minimize it. For the EORTC results, for example, the mean differences for the QOL scales ranged from 1.8 to 11.27, with four of the ten papers having mean differences ≤ 4 , which is considered a trivial difference [64]. The largest mean difference of 11.27 is considered medium [64]. However, the extremely broad limits of agreement can encompass a large MCID for both scales (>15 points) [64]. While we have suggested the MQOL as an end-of-life-specific tool and the COOP/WONCA as a generic tool, the evidence base for these tools is relatively weaker than that for the EORTC, as a relatively small number of papers assessed the MQOL and the COOP/WONCA. However, we provided this recommendation due to our interest in discussing both disease-specific and non-disease-specific tools. Additionally, previous work in other populations provides some support for this recommendation. Kutner and colleagues looked at patient-proxy reporting using the MQOL in a hospice population (58% cancer patients) and demonstrated that patient-proxy correlation was moderate and mean differences were small, concluding that proxies could be used if patient responses were not available [65]. We were unable to find studies beyond those already identified in the review which evaluated proxy-patient concordance using the COOP/WONCA. We did identify studies which looked at patient and proxy reporting for the EQ-5D; as only one paper in our review used this measure, we could not provide a recommendation for it. None of the studies we found included cancer patients, and several were comparing health professional caregivers such as physiotherapists and nursing home staff, rather than family caregivers. Comparison of patient EQ-5D scores to proxy scores in frail elderly [66], community-based individuals with moderate dementia [67], and elderly patients who had visited the emergency department [68] identified large differences and none of the authors recommended the use of proxy ratings for this instrument.

A caveat to our finding about adequate concordance is that different methods were used across the studies for evaluating concordance. For example, while the weighted kappa and the ICC are equivalent under some conditions [69], the ICC is preferable to Pearson's r as the latter does not address systematic bias [70]. A related issue is that skewed data, which may be seen in quality of life research, may result in low correlations simply because variability is limited [15]. The range of scores should therefore be taken into consideration and reported when discussing proxy-patient correlation. Furthermore, use of unweighted kappa can be problematic due to the well-known bias and prevalence problems for that measure [69, 70]. While few studies employed graphical approaches such as Bland-Altman plots, these approaches can be helpful for understanding agreement [71].

The findings of this review suggest that caregiver burden, patient performance status and patient and proxy demographic characteristics should be considered as potential factors. Measurement approaches for these factors should be standardized, especially for factors such as caregiver burden and performance status. Additional exploratory work which examines the potential role of other factors is also needed, as most factors explained a relatively low proportion of variance in the outcome. At a minimum, researchers considering the use of proxy-reported data should collect information on caregiver burden and proxy socio-demographic characteristics such as age and gender. Proxy type should also be collected and reported, unless restricted to a specific type for study design reasons.

This work has several limitations. First, while limiting the included papers to published papers enabled a more thorough assessment of study methods and results, it may have resulted in missing other, potentially relevant findings. Second, exclusion of non-English language papers may have had a similar impact. Third, in many cases methodological issues such as item-level missingness, treatment setting or treatment status were classified as ‘unspecified’ in cases where papers did not explicitly describe the issue in question. This may result in evaluations of quality being more like evaluations of documentation quality; however, we felt it was more appropriate to classify these as ‘not specified’ rather than making possibly unsupported assumptions. Fourth, limiting the review to multi-dimensional, validated HRQOL instruments may limit the generalizability of the findings. Fifth, we did not consider item-level analyses, which may limit the usefulness of the work.

In conclusion, the moderate concordance between proxy and patient reports suggests that proxy reports can be used by researchers. However, it is important to adjust for the substantial residual variability which remains and researchers are encouraged to consider collecting, at a minimum, factors relating to caregiver burden, patient performance status and patient and proxy demographic characteristics. Standardized collection of such factors in future studies can provide important information about their predictive value and lead to better guidelines regarding proxy data collection and use. Although gaps in knowledge remain, we recommend that researchers who need to use proxy reports in surveys or other evaluations which will present group-level analyses employ either the EORTC QLQ-C30 (disease-specific), MQOL (end-of-life-specific), or COOP/WONCA (generic).

Acknowledgments

We would like to thank the reviewers for helpful feedback, Joanne Michaud for validating the data extraction, Tom Trikalinos for detailed advice regarding protocol development and interpretation of the findings, Gaelen Adam for her assistance with the search strategy and Shaun Forbes for helpful feedback on a conference abstract.

Funding Jessica Roydhouse completed this work while supported by an Agency for Healthcare Research and Quality (AHRQ) National Research Service Award Grant T32 HS000011-28 to Prof Vince Mor.

Appendix 1

See Table 8.

References

1. Basch E, Abernethy AP, Mullins CD, Reeve BB, Smith ML, Coons SJ, et al. Recommendations for incorporating patient-reported outcomes into clinical comparative effectiveness research in adult oncology. *Journal of Clinical Oncology*. 2012; 30(34):4249–4255. DOI: 10.1200/JCO.2012.42.5967 [PubMed: 23071244]
2. Cella, D., Hahn, EA., Jensen, SE., Butt, Z., Nowinski, CJ., Rothrock, N. Methodological issues in the selection, administration, and use of patient-reported outcomes in performance measurement in health care settings: Commissioned paper #1. Washington, DC: National Quality Forum; 2012.
3. Fairclough, DL. Practical considerations in outcomes assessment for clinical trials. In: Lipscomb, J.Gotay, CC., Snyder, C., editors. *Outcomes assessment in cancer*. Cambridge: Cambridge University Press; 2005.
4. Simes RJ, Gatrexore V, Gebiski VJ. Practical approaches to minimize problems with missing quality of life data. *Statistics in Medicine*. 1998; 17(5–7):725–737. [PubMed: 9549819]

5. Weinfurt KP, Trucco SM, Willke RJ, Schulman KA. Measuring agreement between patient and proxy responses to multidimensional health-related quality-of-life measures in clinical trials. An application of psychometric profile analysis. *Journal of Clinical Epidemiology*. 2002; 55(6):608–618. [PubMed: 12063103]
6. Stineman MG, Ross RN, Maislin G, Iezzoni L. Estimating health-related quality of life in populations through cross-sectional surveys. *Medical Care*. 2004; 42(6):569–578. [PubMed: 15167325]
7. Todorov A, Kirchner C. Bias in proxies' reports of disability: Data from the National Health Interview Survey on disability. *American Journal of Public Health*. 2000; 90(8):1248–1253. [PubMed: 10937005]
8. Edwards VJ, Anderson LA, Deokar AJ. Proxy reports about household members with increased confusion or memory loss, 2011 Behavioral Risk Factor Surveillance System. *Preventing Chronic Disease*. 2015; 12:E47.doi: 10.5888/pcd12.140427 [PubMed: 25855990]
9. Mosely RR 2nd, Wolinsky FD. The use of proxies in health surveys. Substantive and policy implications. *Medical Care*. 1986; 24(6):496–510. [PubMed: 3713288]
10. Zaslavsky AM, Zaboriski LB, Ding L, Shaul JA, Cioffi MJ, Cleary PD. Adjusting performance measures to ensure equitable plan comparisons. *Health Care Financing Review*. 2001; 22(3):109–126. [PubMed: 25372572]
11. Shields M. Proxy reporting of health information. *Health Reports*. 2004; 15(3):21–33.
12. Bjertnaes O. Patient-reported experiences with hospitals: Comparison of proxy and patient scores using propensity-score matching. *International Journal for Quality in Health Care*. 2014; 26(1):34–40. DOI: 10.1093/intqhc/mzt088 [PubMed: 24334233]
13. Sprangers MA, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease: A review. *Journal of Clinical Epidemiology*. 1992; 45(7):743–760. [PubMed: 1619454]
14. Lobchuk MM, Degner LF. Patients with cancer and next-of-kin response comparability on physical and psychological symptom well-being: Trends and measurement issues. *Cancer Nursing*. 2002; 25(5):358–374. [PubMed: 12394563]
15. Sneeuw KC, Sprangers MA, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease. *Journal of Clinical Epidemiology*. 2002; 55(11):1130–1143. [PubMed: 12507678]
16. Kirou-Mauro A, Harris K, Sinclair E, Selby D, Chow E. Are family proxies a valid source of information about cancer patients' quality of life at the end-of-life? A literature review. *Journal of Cancer Pain & Symptom Palliation*. 2006; 2(2):23–33.
17. Tang ST, McCorkle R. Use of family proxies in quality of life research for cancer patients at the end of life: A literature review. *Cancer investigation*. 2002; 20(7&8):1086–1104. [PubMed: 12449742]
18. Brown University. Abstrackr. 2012. <http://abstrackr.cebm.brown.edu>
19. Wallace, B., Small, K., Brodley, C., Lau, J., Trikalinos, T. Deploying an interactive machine learning system in an evidence-based practice center: Abstrackr. Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium; New York. 2012; New York: Association for Computing Machinery; 2012. p. 819-824.
20. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, 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. *Journal of the National Cancer Institute*. 1993; 85(5):365–376. [PubMed: 8433390]
21. Fayers, P., Aaronson, NK., Bjordal, K., Groenvold, M., Curran, D., Bottomley, A., et al. The EORTC QLQ-C30 scoring manual. 3rd. Brussels: European Organisation for Research and Treatment of Cancer; 2001.
22. Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, et al. The functional assessment of cancer therapy scale: Development and validation of the general measure. *Journal of Clinical Oncology*. 1993; 11(3):570–579. [PubMed: 8445433]
23. Cella D. Assessment methods for quality of life in cancer patients: The FACIT measurement system. *International Journal of Pharmaceutical Medicine*. 2000; 14:78–81.

24. Janda M, DiSipio T, Hurst C, Cella D, Newman B. The Queensland cancer risk study: General population norms for the Functional Assessment of Cancer Therapy-General (FACT-G). *Psycho-oncology*. 2009; 18(6):606–614. DOI: 10.1002/pon.1428 [PubMed: 18942661]
25. Stockler MR, Osoba D, Corey P, Goodwin PJ, Tan-nock IF. Convergent discriminative, and predictive validity of the Prostate Cancer Specific Quality of Life Instrument (PROSQOLI) assessment and comparison with analogous scales from the EORTC QLQ-C30 and a trial-specific module. European Organisation for Research and Treatment of Cancer. Core Quality of Life Questionnaire. *Journal of Clinical Epidemiology*. 1999; 52(7):653–666. [PubMed: 10391659]
26. Padilla GV, Presant C, Grant MM, Metter G, Lipsett J, Heide F. Quality of life index for patients with cancer. *Research in Nursing & Health*. 1983; 6(3):117–126. [PubMed: 6556698]
27. McMillan SC. The quality of life of patients with cancer receiving hospice care. *Oncology Nursing Forum*. 1996; 23(8):1221–1228. [PubMed: 8883070]
28. McMillan SC, Mahon M. Measuring quality of life in hospice patients using a newly developed Hospice Quality of Life Index. *Quality of Life Research*. 1994; 3(6):437–447. [PubMed: 7866362]
29. Tang ST. Concordance of quality-of-life assessments between terminally ill cancer patients and their primary family caregivers in Taiwan. *Cancer Nursing*. 2006; 29(1):49–57. [PubMed: 16557121]
30. Bruley DK. Beyond reliability and validity: Analysis of selected quality-of-life instruments for use in palliative care. *Journal of Palliative Medicine*. 1999; 2(3):299–309. DOI: 10.1089/jpm.1999.2.299 [PubMed: 15859762]
31. Moinpour CM, Lyons B, Schmidt SP, Chansky K, Patchell RA. Substituting proxy ratings for patient ratings in cancer clinical trials: An analysis based on a Southwest Oncology Group trial in patients with brain metastases. *Quality of Life Research*. 2000; 9(2):219–231. [PubMed: 10983485]
32. Sneeuw KC, Aaronson NK, Sprangers MA, Detmar SB, Wever LD, Schornagel JH. Value of caregiver ratings in evaluating the quality of life of patients with cancer. *Journal of Clinical Oncology*. 1997; 15(3):1206–1217. [PubMed: 9060565]
33. Pickard AS, Lin HW, Knight SJ, Sharifi R, Wu Z, Hung SY, et al. Proxy assessment of health-related quality of life in African American and white respondents with prostate cancer: Perspective matters. *Medical Care*. 2009; 47(2):176–183. DOI: 10.1097/MLR.0b013e31818475f4 [PubMed: 19169118]
34. Ware JE Jr. SF-36 health survey update. *Spine (Philadelphia, Pa 1976)*. 2000; 25(24):3130–3139.
35. Skevington SM, Lotfy M, O'Connell KA, Group, W. The World Health Organization's WHOQOL-BREF quality of life assessment: Psychometric properties and results of the international field trial. A report from the WHOQOL group. *Quality of Life Research*. 2004; 13(2):299–310. [PubMed: 15085902]
36. Awadalla AW, Ohaeri JU, Gholoum A, Khalid AO, Hamad HM, Jacob A. Factors associated with quality of life of outpatients with breast cancer and gynecologic cancers and their family caregivers: A controlled study. *BMC Cancer*. 2007; 7:102.doi: 10.1186/1471-2407-7-102 [PubMed: 17578579]
37. Blazeby JM, Williams MH, Alderson D, Farndon JR. Observer variation in assessment of quality of life in patients with oesophageal cancer. *The British Journal of Surgery*. 1995; 82(9):1200–1203. [PubMed: 7551995]
38. Sneeuw KC, Albertsen PC, Aaronson NK. Comparison of patient and spouse assessments of health related quality of life in men with metastatic prostate cancer. *The Journal of Urology*. 2001; 165(2):478–482. DOI: 10.1097/00005392-200102000-00029 [PubMed: 11176400]
39. Sigurdardottir V, Brandberg Y, Sullivan M. Criterion-based validation of the EORTC QLQ-C36 in advanced melanoma: The CIPS questionnaire and proxy raters. *Quality of Life Research*. 1996; 5(3):375–386. [PubMed: 8763806]
40. Giesinger JM, Golser M, Erharter A, Kemmler G, Schauer-Maurer G, Stockhammer G, et al. Do neurooncological patients and their significant others agree on quality of life ratings? Health and Quality of Life Outcomes. 2009; 7:87.doi: 10.1186/1477-7525-7-87 [PubMed: 19814834]

41. Gundy CM, Aaronson NK. The influence of proxy perspective on patient-proxy agreement in the evaluation of health-related quality of life: An empirical study. *Medical Care*. 2008; 46(2):209–216. DOI: 10.1097/MLR.0b013e318158af13 [PubMed: 18219250]
42. Sneeuw KC, Aaronson NK, Sprangers MA, Detmar SB, Wever LD, Schornagel JH. Comparison of patient and proxy EORTC QLQ-C30 ratings in assessing the quality of life of cancer patients. *Journal of Clinical Epidemiology*. 1998; 51(7):617–631. [PubMed: 9674669]
43. Wilson KA, Dowling AJ, Abdoell M, Tannock IF. Perception of quality of life by patients, partners and treating physicians. *Quality of Life Research*. 2000; 9(9):1041–1052. [PubMed: 11332225]
44. Sneeuw KC, Aaronson NK, Osoba D, Muller MJ, Hsu MA, Yung WK, et al. The use of significant others as proxy raters of the quality of life of patients with brain cancer. *Medical Care*. 1997; 35(5):490–506. [PubMed: 9140337]
45. Milne DJ, Mulder LL, Beelen HC, Schofield P, Kempen GI, Aranda S. Patients' self-report and family caregivers' perception of quality of life in patients with advanced cancer: How do they compare? *European Journal of Cancer Care*. 2006; 15(2):125–132. DOI: 10.1111/j.1365-2354.2005.00639.x [PubMed: 16643259]
46. Wennman-Larsen A, Tishelman C, Wengstrom Y, Gustavsson P. Factors influencing agreement in symptom ratings by lung cancer patients and their significant others. *Journal of Pain and Symptom Management*. 2007; 33(2):146–155. DOI: 10.1016/j.jpainsymman.2006.07.019 [PubMed: 17280920]
47. Hisamura K, Matsushima E, Nagai H, Mikami A. Comparison of patient and family assessments of quality of life of terminally ill cancer patients in Japan. *Psycho-oncology*. 2011; 20(9):953–960. DOI: 10.1002/pon.1802 [PubMed: 20623804]
48. Sandgren AK, Mullens AB, Erickson SC, Romanek KM, McCaul KD. Confidant and breast cancer patient reports of quality of life. *Quality of Life Research*. 2004; 13(1):155–160. [PubMed: 15058796]
49. Knight SJ, Chmiel JS, Sharp LK, Kuzel T, Nadler RB, Fine R, et al. Spouse ratings of quality of life in patients with metastatic prostate cancer of lower socioeconomic status: An assessment of feasibility, reliability, and validity. *Urology*. 2001; 57(2):275–280. [PubMed: 11182336]
50. Steel JL, Geller DA, Carr BI. Proxy ratings of health related quality of life in patients with hepatocellular carcinoma. *Quality of Life Research*. 2005; 14(4):1025–1033. [PubMed: 16041898]
51. Doyle M, Bradley NM, Li K, Sinclair E, Lam K, Chan G, et al. Quality of life in patients with brain metastases treated with a palliative course of whole-brain radiotherapy. *Journal of Palliative Medicine*. 2007; 10(2):367–374. DOI: 10.1089/jpm.2006.0202 [PubMed: 17472508]
52. Percy R, Waldron D, O'Boyle C, MacDonagh R. Proxy assessment of quality of life in patients with prostate cancer: How accurate are partners and urologists? *Journal of the Royal Society of Medicine*. 2008; 101(3):133–138. DOI: 10.1258/jrsm.2008.081002 [PubMed: 18344470]
53. Brown PD, Decker PA, Rummans TA, Clark MM, Frost MH, Ballman KV, et al. A prospective study of quality of life in adults with newly diagnosed high-grade gliomas: Comparison of patient and caregiver ratings of quality of life. *American Journal of Clinical Oncology*. 2008; 31(2):163–168. DOI: 10.1097/COC.0b013e318149f1d3 [PubMed: 18391601]
54. Curtis AE, Fernsler JI. Quality of life of oncology hospice patients: A comparison of patient and primary caregiver reports. *Oncology Nursing Forum*. 1989; 16(1):49–53. [PubMed: 2463614]
55. Forjaz MJ, Guarnaccia CA. Hematological cancer patients' quality of life: Self versus intimate or non-intimate confidant reports. *Psycho-oncology*. 1999; 8(6):546–552. [PubMed: 10607987]
56. Deschler DG, Walsh KA, Friedman S, Hayden RE. Quality of life assessment in patients undergoing head and neck surgery as evaluated by lay caregivers. *The Laryngoscope*. 1999; 109(1):42–46. [PubMed: 9917038]
57. Rabin EG, Heldt E, Hirakata VN, Bittelbrunn AC, Chachamovich E, Fleck MP. Depression and perceptions of quality of life of breast cancer survivors and their male partners. *Oncology Nursing Forum*. 2009; 36(3):E153–E158. DOI: 10.1188/09.ONF.E153-E158 [PubMed: 19403443]
58. Hoopman R, Terwee CB, Aaronson NK. Translated COOP/WONCA charts found appropriate for use among Turkish and Moroccan ethnic minority cancer patients. *Journal of Clinical*

- Epidemiology. 2008; 61(10):1036–1048. DOI: 10.1016/j.jclinepi.2007.11.018 [PubMed: 18539431]
59. Sneeuw KC, Aaronson NK, Sprangers MA, Detmar SB, Wever LD, Schornagel JH. Evaluating the quality of life of cancer patients: Assessments by patients, significant others, physicians and nurses. *British Journal of Cancer*. 1999; 81(1):87–94. DOI: 10.1038/sj.bjc.6690655 [PubMed: 10487617]
 60. Jones JM, McPherson CJ, Zimmermann C, Rodin G, Le LW, Cohen SR. Assessing agreement between terminally ill cancer patients' reports of their quality of life and family caregiver and palliative care physician proxy ratings. *Journal of Pain and Symptom Management*. 2011; 42(3): 354–365. DOI: 10.1016/j.jpainsymman.2010.11.018 [PubMed: 21454041]
 61. Tang ST. Predictors of the extent of agreement for quality of life assessments between terminally ill cancer patients and their primary family caregivers in Taiwan. *Quality of Life Research*. 2006; 15(3):391–404. DOI: 10.1007/s11136-005-2158-7 [PubMed: 16547776]
 62. Grassi L, Indelli M, Maltoni M, Falcini F, Fabbri L, Indelli R. Quality of life of homebound patients with advanced cancer: Assessments by patients, family members, and oncologists. *Journal of Psychosocial Oncology*. 1996; 14(3):31–45.
 63. von Essen L. Proxy ratings of patient quality of life—Factors related to patient-proxy agreement. *Acta oncologica*. 2004; 43(3):229–234. [PubMed: 15244245]
 64. Cocks K, King MT, Velikova G, de Castro G Jr, Martyn St-James M, Fayers PM, et al. Evidence-based guidelines for interpreting change scores for the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. *European Journal of Cancer*. 2012; 48(11):1713–1721. DOI: 10.1016/j.ejca.2012.02.059 [PubMed: 22418017]
 65. Kutner JS, Bryant LL, Beaty BL, Fairclough DL. Symptom distress and quality-of-life assessment at the end of life: The role of proxy response. *Journal of Pain and Symptom Management*. 2006; 32(4):300–310. DOI: 10.1016/j.jpainsymman.2006.05.009 [PubMed: 17000347]
 66. Sitoh YY, Lau TC, Zochling J, Cumming RG, Lord SR, Schwarz J, et al. Proxy assessment of health-related quality of life in the frail elderly. *Age and Ageing*. 2003; 32(4):459–461. [PubMed: 12851196]
 67. Orgeta V, Edwards RT, Hounsborne B, Orrell M, Woods B. The use of the EQ-5D as a measure of health-related quality of life in people with dementia and their carers. *Quality of Life Research*. 2015; 24(2):315–324. DOI: 10.1007/s11136-014-0770-0 [PubMed: 25129054]
 68. Tamim H, McCusker J, Dendukuri N. Proxy reporting of quality of life using the EQ-5D. *Medical Care*. 2002; 40(12):1186–1195. DOI: 10.1097/01.MLR.0000036431.38685.EE [PubMed: 12458301]
 69. Hallgren KA. Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*. 2012; 8(1):23–34. [PubMed: 22833776]
 70. Nelson LM, Longstreth WT Jr, Koepsell TD, van Belle G. Proxy respondents in epidemiologic research. *Epidemiologic Reviews*. 1990; 12:71–86. [PubMed: 2286227]
 71. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*. 1999; 8(2):135–160. [PubMed: 10501650]

Fig. 1.
Flow diagram: paper selection and extraction

Table 1

Topics covered by questionnaire items: domains of interest

Questionnaire	Domains/sub-scales/total score calculation	Physical domain	QOL domain	Emotional domain
Disease-specific				
EORTC QLQ-C30 [20, 21]	Role functioning Cognitive functioning Social functioning Fatigue Nausea/vomiting Pain Dyspnea Insomnia Appetite loss Constipation Diarrhea Financial difficulties Total score: calculated, not always used	“Physical functioning” Difficulty with physical activities Ability to get out of bed Ability to perform basic tasks	“Global health status/ QoL” Rating of health and quality of life	“Emotional functioning” Tenseness Worry Irritability Depression
FACIT Suite (FACIT-Sp, FACT-Br, FACT-G, FACT-P, FACT-Hep) [22–24]	FACT-G Physical well-being Social/family well-being Emotional well-being Functional well-being FACT-Br, etc: additional domain of disease-specific sub-scale Total score: calculated	“Physical well-being” Energy Nausea Meeting family needs Pain Treatment side effects Feeling ill	“Functional well-being” Ability to work Fulfillment from work Enjoyment of life and usual activities Rating of quality of life and sleep Acceptance of illness	“Emotional well-being” Sadness Coping with illness Hope Nervousness Concern
PROSQOLI [25]	Pain Physical activity Fatigue Appetite Constipation Family/romantic relationships Mood Passing urine Overall well-being Present pain intensity Total score: not calculated	“Physical activity” Ability to move	N/A	“Mood” Depressed feelings
Quality of Life Index (Padilla) [26]	Psychological well-being/general quality of life Physical well-being Symptom control Financial protection/concerns Total score: calculated	“General physical condition” Pain Nausea Vomiting Strength Appetite	“General quality of life” Rating of quality of life Social activities and enjoyment Life satisfaction Feeling useful Concern about cost of medical care	N/A
End-of-life specific				
Hospice Quality of Life Index [27, 28]	Social/spiritual Psychological/emotional Physical/functional Financial Total score: calculated	“Physical/functional” Constipation Engagement in enjoyable activities Ability to do usual work Tiredness Eating	N/A	“Psychological/emotional” Positive daily views Anger Loneliness Concern/worry for self Masculinity/femininity Pain relief Sadness

Questionnaire	Domains/sub-scales/total score calculation	Physical domain	QOL domain	Emotional domain
McGill Quality of Life (MQOL) [29, 30]	Physical Psychological Existential Relationships Total score: calculated	“Physical” Physical symptoms Physical health rating	N/A (single item which is not included in total scale score)	Concern/worry for friends and family Sleep quality “Psychological” Depression Nervousness Sadness Fear
Spitzer Quality of Life Index [31]	Activity Daily living Health Support Outlook Total score: calculated	N/A	N/A	“Outlook” Feelings about life
Generic COOP/WONCA [32]	Physical fitness Feelings Daily activities Social activities Overall health Pain Quality of life Total score: calculated	“Physical fitness” Ability to perform physical activities	“Quality of life” Quality of life rating	“Feelings” Emotional problems
EQ-5D [33]	Mobility Self-care Usual activities Pain/discomfort Anxiety/depression Total score: calculated	“Mobility” Ability to move	N/A	“Anxiety/depression” Problems with anxiety or depression
SF-36 [34]	Physical health Physical functioning Bodily pain General health Mental health Vitality Social functioning Role-emotional Mental health Total score: calculated. 2 scores, MCS and PCS	“Physical functioning” Ability to perform daily physical activities	N/A	“Mental health” Nervousness Sadness Calmness Peacefulness Happiness
WHOQOL-BREF [35, 36]	Overall Quality of Life and General Health (not in total score) Physical health Psychological Social relationships Environment Total score: not calculated	“Physical health” Pain Energy Sleep Daily activities Ability to work Need for medication Ability to get around	“Overall quality of life and general health” Rating of quality of life	“Psychological” Positive and negative feelings Ability to concentrate Body image Satisfaction with self Negative feelings

Table 2

Study characteristics

Study (year)	Country	Study design	Cancer type(s) and stage(s)	Treatment status	Treatment setting	Analytic cohort size (N dyads) ^d	%Spousal proxies ^e	HRQOL tool(s) used	Tool measure(s) evaluated in study for domains of interest	Concordance methods used ^f
Disease-specific: EORTC										
Blazeby et al. (1995) [37]	UK	Cross-sectional	Esophageal; some advanced, others unspecified	Mix: pre-treatment, post-treatment, palliative	Unspecified	78 (39 dyads)	64%	EORTC QLQ-C30	Global health status/QOL Physical functioning Emotional functioning	Weighted kappa Pairs with exact agreement
Sigurdardottir et al. (1996) [39]	Sweden	Cross-sectional	Melanoma: metastatic	Active	Inpatient and outpatient	60 (30 dyads)	79%	EORTC QLQ-C36	Global health status/QOL Physical functioning Emotional functioning	Wilcoxon matched-pairs signed-ranks test Pearson's r
Sneeuw et al. (1997) [32, 44]	UK, USA	Longitudinal cohort	Brain: stage unspecified	Mix of active treatment and not on treatment	Unspecified	206 (103 dyads)	75%	EORTC QLQ-C30	Global health status/QOL Physical functioning Emotional functioning	Cohen's d (effect size) <i>T</i> -test Repeated measures ANOVA ICC Pairs with exact agreement Pairs with agreement within 2 categories
Sneeuw et al. (1998) [42] ^c	Netherlands	Longitudinal cohort	Various cancers at unspecified stages; breast, GI most common	Active	Inpatient and outpatient	614 (307 dyads)	75%	EORTC QLQ-C30	Global health status/QOL Physical functioning Emotional functioning Total score	Cohen's d (effect size) <i>T</i> -test ICC Multitrait-multimethod analysis Cronbach's alpha Scatter plot
Wilson et al. (2000) [43]	Canada	Cross-sectional	Breast: metastatic	Mix of active treatment and supportive care only	Unspecified	142 (71 dyads)	70%	EORTC QLQ-C30	Global health status/QOL Physical functioning Emotional functioning	ANOVA ICC Pairs with exact agreement
Sneeuw et al. (2001) [38]	Netherlands	Cross-sectional	Prostate; advanced/metastatic	Active	Unspecified	144 (72 dyads)	100%	EORTC QLQ-C30	Global health status/QOL Physical functioning Emotional functioning	Cohen's d (effect size) <i>T</i> -test ICC Pairs with exact agreement Disagreement magnitude
Milne et al. (2006) [45]	Australia	Longitudinal cohort	Mix of advanced cancers; breast, bowel most common	Unspecified	Outpatient	102 (51 dyads)	74.5%	EORTC QLQ-C30	Global health status/QOL Physical functioning Emotional functioning	<i>T</i> -test Pearson's r
Wennman-Larsen et al. (2007) [46]	Sweden	Cross-sectional	Lung; various stages	Mix of treatment "phases"	Unspecified	106 (54 proxies, 52 patients)	81%	EORTC QLQ-C30	Global health status/QOL Physical functioning Emotional functioning	ICC <i>T</i> -test Effect size
Gundy and Aaronson (2008) [41] ^c	Netherlands	RCT	Various cancers; at unspecified stages; breast, GI most common	Active	Inpatient and outpatient	448 (224 dyads)	72%	EORTC QLQ-C30	Global health status/QOL Physical functioning Emotional functioning	<i>T</i> -test Effect size Pearson's r ICC % agreement within 10 points per scale

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Study (year)	Country	Study design	Cancer type(s) and stage(s)	Treatment status	Treatment setting	Analytic cohort size (N dyads) ^d	%Spousal proxies ^e	HRQOL tool(s) used	Tool measure(s) evaluated in study for domains of interest	Concordance methods used ^f
Giesinger et al. (2009) [40]	Austria	Cross-sectional	Brain; grades II–IV	Unspecified	Outpatient	84 (42 dyads)	73%	EORTC QLQ-C30	Global health status/QOL Physical functioning Emotional functioning	Multitrait-multi-method analysis Profile level Profile scatter Profile shape 7 tests Pearson's r Bland–Altman plots
Pickard et al. (2009) [33]	USA	Cross-sectional	Prostate; stage unspecified	Unspecified	Unspecified	174 (87 dyads)	63%	EORTC QLQ-C30	Global health status/QOL Physical functioning Emotional functioning	7 test/Wilcoxon signed rank test Cronbach's alpha Effect size ICC % exact agreement
Disease-specific: FACT/FACIT Knight et al. (2001) [49]	USA	Cross-sectional	Prostate; metastatic	Unspecified	Unspecified	72 (36 dyads)	100%	FACT-G	Physical well-being Emotional well-being Functional well-being	Cronbach's alpha 7 test ICC
Sandgren et al. (2004) [48]	USA	Longitudinal cohort	Breast; stages I–III	Unspecified	Unspecified	224 (112 dyads)	60%	FACT-G	Physical well-being Emotional well-being Functional well-being	7 tests ICC
Steel et al. (2005) [50]	USA	Longitudinal cohort	Hepatocellular carcinoma; stages I–IV	Unspecified	Unspecified	164 (82 dyads)	Unspecified	FACT-Hep	Physical well-being Emotional well-being Functional well-being	ICC Cronbach's alpha Comparison of mean scores (primarily graphically; no statistical comparison)
Doyle et al. (2007) [51]	Canada	Cross-sectional	Various metastatic cancers; lung; breast most common	Pre-treatment	Unspecified	120 (60 dyads)	55%	FACT-Br	Physical well-being Emotional well-being Functional well-being	7 test Lin's concordance
Brown et al. (2008) [53]	USA	RCT	Brain; advanced	Active	Unspecified	362 (181 dyads)	Unspecified	FACT-Br	Total score	Paired signed rank test Spearman's correlation Bland–Altman plots ICC % pairs with differences within 10 units
Pearcy et al. (2008) [52]	UK	Cross-sectional	Prostate; “all stages”	Pre-treatment	Unspecified	50 (25 dyads)	100%	FACT-P	Total score	7 test
Hisamura et al. (2011) [47]	Japan	Cross-sectional	Various advanced cancers; lung; GI most common	Palliative/hospice	Inpatient and outpatient	204 (102 dyads)	49.5%	FACT-Sp	Physical well-being Emotional well-being Functional well-being	ICC Weighted kappa Wilcoxon signed-rank test Cohen's d effect size
Disease-specific: PROSQOLI Wilson et al. (2000) [43]	Canada	Cross-sectional	Prostate; metastatic	Mix of active treatment and supportive care only	Unspecified	58 (29 dyads)	100%	PROSQOLI	Physical function Mood	ANOVA ICC Pairs with exact agreement
Disease-specific: QLI										

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Study (year)	Country	Study design	Cancer type(s) and stage(s)	Treatment status	Treatment setting	Analytic cohort size (N dyads) ^d	%Spousal proxies ^e	HRQOL tool(s) used	Tool measure(s) evaluated in study for domains of interest	Concordance methods used ^f
Curtis and Fernsler (1989) [54]	USA	Cross-sectional	Unspecified advanced cancers	Palliative/hospice	Home	46 (23 dyads)	57%	Quality of Life Index (Padilla)	Total score	<i>T</i> test
Generic: SF-36										
Deschler et al. (1999) [56]	USA	Cross-sectional	Head and neck; various stages including benign tumors	Pre-treatment	Unspecified	50 (25 dyads)	54%	SF-36	Physical functioning Mental health	% pairs where scores fell within each other's 90% confidence interval
Forjaz et al. (1999) [55]	Portugal	Cross-sectional	Various hematologic cancers at various stages; NHL, leukemia most common	Mix of active and no treatment	Outpatient	98 (49 dyads)	67%	SF-36	Physical functioning Mental health	<i>T</i> tests Eta effect size Pearson's <i>r</i>
Generic: EQ-5D										
Pickard et al. (2009) [33]	USA	Cross-sectional	Prostate; stage unspecified	Unspecified	Unspecified	174 (87 dyads)	63%	EQ-5D	Mobility Anxiety/depression	<i>T</i> test/Wilcoxon signed rank test Cronbach's alpha Effect size ICC % exact agreement
Generic: WHOQOL-BREF										
Awadalla et al. (2007) [36]	Sudan	Cross-sectional	Breast, cervical, ovarian, stage unspecified	Mix of active treatment and post-treatment	Outpatient	362 (181 dyads)	30.4%	WHOQOL-BREF	Physical health Psychological health General health/QOL	Cronbach's alpha <i>T</i> test ICC Kendall's tau
Rabin et al. (2009) [57]	Brazil	Cross-sectional	Breast, stages I–III	Unspecified	Unspecified	146 (73 dyads)	100%	WHOQOL-BREF	Physical Psychological Overall quality of life	<i>T</i> test ICC
Generic: COOP–WONCA										
Sneeuw et al. (1997) [32, 44] ^b	Netherlands	Longitudinal cohort	Various cancers at unspecified stages; breast, GI most common	Active	Inpatient and outpatient	590 (295 dyads)	74%	COOP–WONCA	Physical fitness Feelings Quality of life	% exact agreement % agreement within 1 response category ICC ICC for test–retest reliability <i>T</i> test Cohen's <i>d</i> effect size Relative validity estimates
Sneeuw et al. (1999) [59] ^b	Netherlands	Cross-sectional	Various cancers at unspecified stages; breast, lung most common	Active	Inpatient	180 (90 dyads)	76%	COOP–WONCA	Physical fitness Feelings Quality of life	<i>T</i> test Cohen's <i>d</i> effect size ICC % exact agreement % agreement within 1 response category % agreement within >1 response category
Hoopman et al. (2008) [58]	Netherlands	Cross-sectional	Various cancers at various stages; breast, head/neck most common	Mix of active treatment and "under control"	Outpatient	114 (57 dyads)	25%	COOP–WONCA	Physical fitness Feelings Quality of life	<i>T</i> test Effect size % exact agreement % agreement within 1 category % agreement within >1 category Limits of agreement for differences
End-of-life specific: MQOL										

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Study (year)	Country	Study design	Cancer type(s) and stage(s)	Treatment status	Treatment setting	Analytic cohort size (N dyads) ^d	%Spousal proxies ^e	HRQOL tool(s) used	Tool measure(s) evaluated in study for domains of interest	Concordance methods used ^f
Tang (2006) [29, 61] ^a , (concordance study)	Taiwan	Cross-sectional	Various advanced cancers; hematologic; lung most common	Palliative/hospice	Inpatient	228 (114 dyads)	41.6%	McGill Quality of Life	Physical well-being Psychological well-being Total score	Cronbach's alpha Weighted kappa <i>T</i> -test Cohen's d effect size Correlation
Tang (2006) [29, 61] ^a (predictors study)	Taiwan	Cross-sectional	Mix of advanced cancers; hematologic; lung most common	Palliative/hospice	Inpatient	228 (114 dyads)	41.6%	McGill Quality of Life	Physical well-being Psychological well-being Total score	N/A—predictors only, reported in a separate table
Jones et al. (2011) [60]	Canada	Longitudinal cohort	Various advanced cancers; lung, GI most common	Palliative/hospice	Inpatient	160 (80 dyads)	68%	McGill Quality of Life	Physical well-being Psychological well-being Total score	<i>T</i> -test Cohen's d effect size Linear mixed model for repeat measures ICC % within 1 point GEE for % within 1 point over time ICC for change scores Cohen's kappa for change score agreement
End-of-life specific: SQLI Grassi et al. (1996) [62]	Italy	Longitudinal cohort	Various advanced cancers; GI, GU most common	Palliative/hospice	Home	98 (49 dyads)	82%	Spitzer Quality of Life Index	Total/global score	<i>T</i> -test Pearson's r % exact agreement Kappa
Moinpour et al. (2000) [31]	USA	RCT	Mix of metastatic cancers; lung most common	Radiotherapy (treatment), observation (control)	Unspecified	80 (40 dyads)	43.9%	Spitzer Quality of Life Index	Total/global score	Lin's concordance Bland–Altman plots Weighted kappa Double repeated measures model
End-of-life specific: HQLI McMillan (1996) [27]	USA	Longitudinal cohort	Various cancers at advanced stages; lung, prostate most common	Palliative/hospice	Home and nursing home	236 (118 dyads)	74%	Hospice Quality of Life Index	Total score Psychological Physical/functional	Pearson's r <i>T</i> -test

^aThe papers use the same population; one evaluates concordance and the other looks at predictors of concordance

^bThe papers use the same population; the 1999 article examines a subset of the 1997 article's population (inpatients only rather than inpatients and outpatients)

^cThe papers use the same population, with the 2008 article looking at proxy perspectives in a sub-population of the 1998 article d Baseline reported for all longitudinal studies. This is the overall analytic cohort, numbers analyzed may vary per outcome e “Spouses” encompasses both spouses and partners f Analyses presented here are restricted to those relevant to proxy–patient concordance. For example, test–retest reliability within patients only would not be included. Analyses relating to factors affecting concordance are presented in Table 6 and not described here

Table 3

Concordance results across domains—patient vs proxy means, mean differences and limits of agreement (LOA): disease-specific tools

Authors (year)	Global QOL			Physical domain			Emotional domain			Total score		
	Patient mean (SD)	Proxy mean (SD)	Mean difference (LOA)	Patient mean (SD)	Proxy mean (SD)	Mean difference (LOA)	Patient mean (SD)	Proxy mean (SD)	Mean difference (LOA)	Patient mean (SD)	Proxy mean (SD)	Mean difference (LOA)
EORTC	<i>0–100</i> (100 = best)			<i>0–100</i> (100 = best)			<i>0–100</i> (100 = best)			<i>0–100</i> (100 = best)		
Sigurdardottir et al. (1996) [39]			3.45 (−43.7 to 50.56)			1.03 (−21.6 to 23.6)			1.01 (−29.9 to 31.9)			
Sneeuw et al. (1997) [32, 44]	64.6 (21.2)	61.4 (24.5)	3.2 (−34.4 to 40.8)	72.2 (30.3)	65.8 (30.7)	6.4 (−41.4 to 54.2)	74.8 (20.9)	71.2 (21.4)	3.6 (−32.5 to 39.7)			
Sneeuw et al. (1998) [42]	62.9 (22.1)	55.8 (23.8)	7.1 (−35.0 to 49.2)	63.6 (28.1)	58.4 (28.2)	5.2 (−34.4 to 44.8)	75.7 (20.6)	66 (23.1)	9.7 (−32.4 to 51.8)	76 (14.2)	72.3 (15.6)	3.7 (−17.9 to 25.3)
Wilson et al. (2000) [43]	59.9 (25.1)	53.5 (24.1)	6.4 (−61.8 to 74.6) ^b	62 (22.7)	58.5 (23.2)	3.5 (−60.1 to 67.1) ^b	64.3 (26.8)	56.1 (23.9)	8.2 (−62.2 to 78.6) ^b			
Sneeuw et al. (2001) [38]	66.9 (24.3)	65.3 (29)	1.6 (−45.1 to 48.3)	78.6 (24.7)	71.9 (28)	6.7 (−31.7 to 45.1)	78.9 (18.6)	75.1 (25)	3.8 (−40.1 to 47.7)			
Milne et al. (2006) [45] ^d	58.66 (24.3)	47.39 (26.6)	11.3 (−59.4 to 89.9) ^b	66.3 (28.4)	57.3 (28.0)	9.0 (−69.2 to 87.2) ^b	70.9 (24.7)	55.4 (25.7)	15.5 (−54.3 to 85.4) ^b			
Weerman-Larsen et al. (2007) [46]	55.7 (21)	47.5 (18.8)	8.2 (−47.0 to 63.4) ^b	62.9 (22.2)	54.8 (20.7)	8 (−51.5 to 67.5) ^b	69 (23.1)	58.6 (24.6)	10.4 (−55.7 to 76.5) ^b			
Gundy and Aaronson (2008) [41]			4.82 (1.3 to 8.3) ^b			4.78 (1.2 to 8.4) ^b			8.79 (5.3 to 12.3) ^b			
Giesinger et al. (2009) [40]	63.8 (23)	62 (21.6)	1.8 (−42.3 to 45.9) ^b	77.6 (27.3)	74.3 (28.8)	3.3 (−50.6 to 57.2) ^b	59.5 (30.4)	61.8 (23.8)	2.3 (−54.1 to 58.7) ^b			
Pickard et al. (2009) [33]	72.5 (24.1)	69.1 (22.8)	4 (−51.5 to 59.5)	72.8 (31.3)	65.8 (30)	7.4 (−50.8 to 65.6)	87.5 (19.9)	84.1 (19.6)	3.5 (−34.7 to 41.7)			
FACT-Sp	<i>0–28</i> (28 = best)			<i>0–28</i> (28 = best)			<i>0–24</i> (24 = best)			<i>0–152</i> (152 = best)		
Hisamura et al. (2011) [47]			1.2 (−9.0 to 11.4)			0.9 (−8.3 to 10.1)			3.4 (−8.8 to 15.5)			11.2 (−30.7 to 53.2)
FACT-G	<i>0–28</i> (28 = best)			<i>0–28</i> (28 = best)			<i>0–24</i> (24 = best)			<i>0–108</i> (108 = best)		
Knight et al. (2001) [49]	16.1 (6.1)	16.1 (7.4)	0 (−18.8 to 9.6) ^b	20.7 (6.1)	19 (6)	1.7 (−15.1 to 18.5) ^b	16.5 (3.6)	14.7 (5.3)	1.8 (−10.8 to 14.4) ^b	81.4 (17)	77.7 (19.7)	3.7 (−47.3 to 54.7)
Sundgren et al. (2004) [48] ^d	22.0 (4.9)	21.13 (4.6)	0.8 (−12.3 to 7.6) ^b	20.2 (6.1)	19.5 (6.1)	0.8 (−16.1 to 17.6) ^b	19.2 (4.3)	18.6 (4.3)	0.6 (−11.2 to 12.5) ^b	90.3 (14.4) ^c	87.3 (13.9)	3.0 (−36.2 to 42.2) ^b
FACT-P	<i>0–28</i> (28 = best)			<i>0–28</i> (28 = best)			<i>0–24</i> (24 = best)			<i>0–156</i> (156 = best)		
Pearcy et al. (2008) [52]										150 (10.5)	140 (6.1)	10 (−13.8 to 33.8) ^b
FACT-Br	0–28 (28 = best)			0–28 (28 = best)			0–24 (24 = best)			0–200 (200 = best)		
Brown et al. (2008) [53] ^d												
PROSQOLI				<i>0–100</i> (100 = best)			<i>0–100</i> (100 = best)			72.7 (12.9)	73.1 (14.4)	0.4 (−37.5 to 38.3) ^b
Wilson et al. (2000) [43]				69.5 (23.4)	73.7 (24.4)	4.2 (−62.1 to −57.9)	76 (20.3)	65.3 (25.3)	10.7 (−52.9 to 74.3)	N/A	N/A	N/A

Table is limited to those studies which provided at least one mean score

^aBaseline scores only

^bImputed by author

^cAuthors note their scale is 0–112 due to the addition of a study-specific item

Table 4

Concordance results across domains—patient vs proxy means, mean differences and limits of agreement (LOA): end of life-specific tools

Authors (year)	Physical domain			Emotional domain			Total score		
	Patient mean (SD)	Proxy mean (SD)	Mean difference (LOA)	Patient mean (SD)	Proxy mean (SD)	Mean difference (LOA)	Patient mean (SD)	Proxy mean (SD)	Mean difference (LOA)
Hospice QLI							25–250 (250 = best)		
McMillan (1996) [27] ^a							171.4 (31.5)	160.5 (36.3)	10.9 (−83.3 to 105.1) ^b
MQOL	0–10 (10 = best)			0–10 (10 = best)					
Tang (2006) [29, 61]	3.8 (2.9)	4.3 (2.6)	0.6 (−5.1 to 6.2)	22.1 (11.9) ^c	22.5 (11) ^c	0.3 (−25.0 to 25.5)	75.5 (27.9) ^c	72.8 (27.7) ^c	2.8 (−44.9 to 50.4)
Jones et al. (2011) [60] ^d	5.9 (2.7)	4.5 (2.3)	1.4 (−4.5 to 7.3)	7.2 (2.4)	5.6 (2.3)	1.6 (−3.7 to 6.9)	7.3 (1.4)	6.2 (1.4)	1.1 (−2.0 to 4.2)
Spitzer QLI							0–10 (10 = best) ^d		
Grassi et al. (1996) [62] ^a							4.8 (1.6)	4.6 (1.5)	0.2 (−4.2 to 4.6) ^b
Moinpour et al. (2000) [31] ^d									0.6 (−0.6 to 1.8) ^b

Restricted to studies which presented at least one mean score

^aBaseline scores only

^bImputed by author

^cAppears to use different scoring approach

^dThe scale typically has higher numbers reflecting worse function/more impairment, but the authors noted the use of a reversed scale where higher = better

Concordance results across domains—patient vs proxy means, mean differences and limits of agreement (LOA): generic tools

Table 5

Authors (year)	Global QOL			Physical domain			Emotional domain			Total score		
	Patient mean (SD)	Proxy mean (SD)	Mean difference (LOA)	Patient mean (SD)	Proxy mean (SD)	Mean difference (LOA)	Patient mean (SD)	Proxy mean (SD)	Mean difference (LOA)	Patient mean (SD)	Proxy mean (SD)	Mean difference (LOA)
COOP-WONCA	<i>L-5</i> (<i>5 = worst</i>)			<i>L-5</i> (<i>5 = worst</i>)			<i>L-5</i> (<i>5 = worst</i>)					
Sneeuw et al. (1997) [32, 44] ^a	3 (1.1)	3.3 (0.9)	0.3 (−1.7 to 2.3)	3.1 (1.2)	3.3 (1.2)	0.2 (−2.0 to 2.4)	2.2 (1)	2.5 (1.1)	0.3 (−1.7 to 2.3)			
Sneeuw et al. (1999) [59]	3.2 (1.1)	3.5 (0.9)	0.3 (−2.5 to 3.1) ^b	3.3 (1.3)	3.5 (1.2)	0.2 (−3.3 to 3.7)*	2.2 (1)	2.7 (1.1)	0.5 (−2.4 to 3.4) ^b			
Hoogman et al. (2008) [58]	3.4 (1)	3.4 (0.9)	0.0 (−2.5 to 2.6)	3.5 (1.2)	3.2 (1.4)	0.3 (−2.8 to 3.5)	2.8 (1.3)	2.8 (1.2)	0.1 (−2.9 to 3.0)			
EQ-5D										0–100 (100 = best)		
Pickard et al. (2009) [33]										73.4 (20.2)	69.4 (20.3)	3.8 (−38.9 to 46.5)
WHOQOL-BREF	<i>Raw: 2–10</i> (<i>10 = best</i>) <i>Transformed: 0–100</i> (<i>100 = best</i>)			<i>Raw: 2–35</i> (<i>35 = best</i>) <i>Transformed: 0–100</i> (<i>100 = best</i>)			<i>Raw: 6–30</i> (<i>30 = best</i>) <i>Transformed: 0–100</i> (<i>100 = best</i>)			N/A		
Awadalla et al. (2007) [36]	8.3 (1.9)	8.5 (1.4)	0.2 (−4.4 to 4.8) ^b	13.4 (1.7)	12.9 (1.7)	0.5 (−4.2 to 5.2) ^b	20.9 (2.9)	19.2 (1.6)	1.7 (−4.8 to 8.2) ^b			
Rabin et al. (2009) [57] ^c	65.2 (18.8)	67 (17.7)	1.7 (−48.8 to 52.4) ^b	62.5 (20.1)	59.3 (16.3)	3.2 (−47.5 to 53.9)	66 (17)	65.6 (12.7)	0.3 (−41.2 to 42.0) ^b			

SF-36 not included as 1 of the 2 studies using it did not assess mean differences, and the other only did it within subgroups rather than overall

^aBaseline scores only
^bImputed by author
^cAuthors report using 1–100 scale

Table 6

Proxy-specific factors evaluated for association with patient-proxy concordance

Study ID	Focus of evaluation	Measurement approach(es)	Factors evaluated	Significant results for QOL, PF, EF domains
Pickard et al. (2009) [33]	Comparing proxy perspectives in terms of impact on patient-proxy concordance	Effect size (standardized response mean) between perspectives, paired <i>t</i> tests ICC between perspectives Exact agreement (100% concordance) between perspectives Kendall's tau/Mann Whitney U for correlations between proxy factors and proxy-patient difference between perspectives Logistic regression to identify predictors of non-exact agreement between perspectives	Age Gender Race/ethnicity Education Employment status Living with patient Type of relationship with patient Health literacy (Rapid Estimate of Adult Literacy in Medicine score) Depressive symptoms (Center for Epidemiological Studies—Depression score) Proxy perspective (proxy-patient and proxy-proxy)	Significant mean score differences <i>between proxy perspectives</i> for EF, PF, EQ-5D VAS. Proxy-patient differences were smallest for the proxy-patient perspective Similar levels exact agreement between perspectives for EORTC and VAS (same for PF, EF; within 1–2% for QOL, VAS). Differences of 6–10% for mobility and anxiety for EQ-5D, favoring proxy-patient perspective Similar levels ICC across perspectives. Slightly better agreement for proxy-patient for mobility, EF, VAS; slightly better for proxy-proxy for anxiety, QOL; same for PF Significantly smaller differences between perspectives for PF for <i>proxies with limited literacy</i> Significantly lower odds of exact agreement between perspectives for VAS for <i>proxies with depressive symptoms</i>
Sandgren et al. (2004) [48]	Compare proxy-patient differences by proxy-patient relationship	<i>T</i> test for absolute value of difference between proxy and patient scores	Proxy-patient relationship: spouses vs other	Significantly smaller difference on total QOL scale for <i>spouses</i> relative to other proxies
Forjaz et al. (1999) [55]	Compare proxy-patient differences by proxy-patient relationship	Comparison of matched <i>t</i> test differences Comparison of effect size Comparison of significant correlations Mean proxy-patient correlation between groups	Proxy-patient relationship: spouses vs other	Significant mean difference (<i>t</i> test) between patient and proxy for mental health <i>for spouses</i> but not for non-spouse Significant proxy-patient correlations for physical and mental health <i>for spouses</i> ; for <i>non-spouses</i> , significant mental health correlation only Effect sizes not as large as significant differences No significant difference for mean correlation between groups
Rabin et al. (2009) [57]	Compare differences in scores by various characteristics	Hierarchical multiple linear regression	NB: study restricted to male partners Length of time proxy and patient have lived together	No significant difference found
Sneeuw et al. (1999) [59]	Compare response agreement by proxy characteristics	Percent large discrepancies (proxy, patient responses are >1 response category from each other) between groups	Age Gender Education level	NB: statistical significant not assessed Percent differences between groups ranged from 1%–5% Smallest difference for gender (1%), highest for education (5%: intermediate vs low)
Gundy and Aaronson (2008) [41]	Comparing proxy perspectives (proxy-patient and proxy-proxy) in terms of	Cronbach's alpha for scale reliabilities under each perspective <i>T</i> test for mean patient-proxy differences under each perspective	Proxy perspective (proxy-patient and proxy-proxy)	Cronbach's alpha similar for EF, better for proxy-proxy by 0.06–0.09 for PF, QOL

Study ID	Focus of evaluation	Measurement approach(es)	Factors evaluated	Significant results for QOL, PF, EF domains
	impact on patient-proxy concordance	<i>T</i> tests and standardized mean differences to compare bias across perspectives Pearson's <i>r</i> and ICC for patient and proxy ratings across perspectives Percent patient-proxy ratings within 10 points of each other Multitrait-multimethod analysis of patient-proxy correlations (convergence, discrimination evaluation across perspectives) Profile level, scatter and shape across perspectives	Mental health (Mental Health Inventory-5) Global health/QOL (EORTC QLQ-C30) Proxy-patient relationship Proxy living with patient Frequency of proxy-patient contact	Significant mean differences (<i>t</i> test) between patient and proxy for both perspectives for PF, EF, QOL, however no significant differences across perspectives Higher correlation for PF for proxy-proxy perspective, but higher for proxy-patient for EF, QOL. Differences not significant Similar convergence, discrimination across perspectives No significant differences for profile across perspectives No significant effect of proxy factors on differences across perspectives
Tang (2006) [29, 61] *	Identifying predictors of patient-proxy agreement	Multiple regression <i>T</i> test for mean differences Pearson's correlation with mean of absolute difference in scores Pearson's correlation with mean of differences	Age Gender Employment status Comorbidity Previous caregiving experience Proxy-patient relationship Proxy-patient contact frequency Proxy-patient communication about disease and symptoms Proxy perceived knowledge of disease and symptoms Care burden, measured by Caregiver Reaction Assessment (impact on schedule, health, finance; family support; self-esteem) Amount of caregiving required	NB only total scores used in this analysis Significant larger absolute mean differences (worse agreement) if <i>proxies had comorbidities</i> Significant positive correlation with absolute mean difference (worse agreement) and the <i>impact of caregiving on proxy health</i> Significant positive correlation with absolute mean difference (worse agreement) and <i>better proxy-perceived knowledge of patient disease and symptoms</i> In multivariable analyses, only <i>impact of caregiving on proxy health and proxy-perceived knowledge of disease and symptoms</i> were significant (increases in scores for these measures were associated with increased absolute differences, e.g. worse agreement, between proxy and patient scores)
Sneeuw et al. (1998) [42]	Compare proxy-patient differences across various characteristics	Correlation between variables of interest and total QOL score Hierarchical regression analysis with total QOL score as outcome variable Differences measured as both absolute difference and directional difference	Gender Age Education Proxy-patient relationship Proxy global QOL/health Proxy mental health Caregiving intensity Caregiving burden (frequency of feeling burdened) Living with patient Frequency of contact with patient Quality of proxy-patient relationship (Norbeck Questionnaire) Quality of proxy-patient communication (Cancer Rehabilitation Evaluation System)	<i>Male proxies</i> had significantly larger absolute differences with patients <i>Older proxies</i> had significantly larger absolute differences with patients <i>Proxies with poorer QOL</i> had significantly larger absolute differences with patients <i>Proxies with greater caregiving intensity</i> had significantly larger absolute and directional differences with patients <i>Proxies with worse mental health</i> had significantly larger directional differences with patients In multivariable analyses for absolute difference, only <i>proxy QOL</i> remained significant In multivariable analyses for directional difference, only <i>proxy mental health</i> and <i>proxy caregiving intensity</i> remained significant

Study ID	Focus of evaluation	Measurement approach(es)	Factors evaluated	Significant results for QOL, PF, EF domains
Sneeuw et al. (1997) [32, 44]	Evaluate association between number of proxy–patient responses without exact agreement and various characteristic	Number of discrepancies across all questions in the QLQ-C30, per proxy–patient pair ANOVA to compare mean number of discrepancies among relevant groups Test of linear trends in mean number of discrepancies (for multi-level variables only)	Gender Age Proxy–patient relationship Living with patient Length of proxy–patient relationship	No significant results identified for proxy characteristics
Wennman-Larsen et al. (2007) [46]	Compare proxy–patient differences, focusing on situations where proxies underestimated function	Correlation between characteristics and mean proxy–patient differences, if mean differences had effect sizes >0.40 Multiple regression, if mean differences had effect sizes >0.40	Proxy–patient relationship Gender Education Age Care burden, measured by Caregiver Reaction Assessment (impact on schedule, health, finance; family support; self-esteem) Employment status	NB only QOL, EF had effect sizes >0.40; PF thus not considered in these analyses Significantly more disagreement for EF for <i>female proxies</i> <i>Lack of family support for proxy</i> significantly associated with more disagreement for QOL, EF <i>Worse (higher) impact of caregiving on proxy health</i> significantly associated with more disagreement for QOL, EF <i>Higher proxy self-esteem</i> significantly associated with more disagreement for EF In multivariable models, <i>proxy self-esteem</i> was significantly associated with EF concordance (direction unspecified) and <i>lack of family support for proxy</i> was significantly associated with QOL concordance (direction unspecified)
Deschler et al. (1999) [56]	Compare congruence across proxy types	Congruence defined as proxy score within 90% CI of patient score; calculated for each domain Chi square, Fisher's exact test to see if factors significantly associated with differences in congruence	Proxy–patient relationship (spouse, sibling, parent, child) Proxy–patient generational relationship (spouse/sibling vs parent/child)	Significantly better congruence if <i>proxies in same generation (spouse or sibling) as patient (vs parent or child of patient)</i>

* Unlike other included studies, this study defined statistical significance as $p < 0.10$

Table 7

Patient-specific factors evaluated for association with patient-proxy concordance

Study ID	Focus of evaluation	Measurement approach(es)	Factors evaluated	Significant results for QOL, PF, EF domains
Jones et al. (2011) [60]	Evaluate the impact of patient factors on patient-proxy score differences	Linear mixed model with difference in scores as dependent variable	Cognitive function (Short Orientation-Memory-Concentration Test) Symptom burden (Edmonton Symptom Assessment Scale) Performance status (Palliative Performance Scale) Gender Age	Significantly smaller mean differences for psychological scale and total score <i>in patients with poorer cognitive function</i> Significantly smaller mean differences for psychological and physical scales and total score <i>in patients with a higher symptom burden</i>
Rabin et al. (2009) [57]	Compare differences in scores by various characteristics	Hierarchical multiple linear regression	Age Depression (Beck Depression Inventory) Education Stage of disease Treatment Duration of disease	Significantly smaller differences for psychological scale <i>in patients with higher depression scores/more depression</i>
Moinpour et al. (2000) [31]	Evaluate difference in patient QOL by treatment group	Double repeated measures analysis	Assigned treatment (radiotherapy or observation)	Significant proxy-patient difference <i>for radiotherapy</i> over 3-month period (proxies report negative effect of therapy, patients don't)
Sneeuw et al. (1997) [32, 44] (study in patients with range of tumor types)	Compare agreement among groups defined by patient clinical status	Mean of absolute difference in scores; <i>t</i> test to compare groups	Performance status (Eastern Cooperative Oncology Group); good (0/1) versus poor (2/3)	Significantly smaller differences (better agreement) <i>for patients with worse performance status</i> for physical scale, QOL Significantly larger differences (worse agreement) <i>for patients with worse performance status</i> for feelings
Sneeuw et al. (1999) [59]	Compare response agreement by proxy characteristics	Percent large discrepancies (proxy, patient responses are >1 response category from each other) between groups	Performance status (Eastern Cooperative Oncology Group) Age Gender Education	NB: statistical significance not assessed Differences range from 1 to 14% Smallest difference for gender (1%) Largest difference for performance status (14%, ECOG 0 vs ECOG 2; larger % discrepancies seen for ECOG 2)
Tang (2006) [29, 61] *	Identifying predictors of patient-proxy agreement	Multiple regression <i>T</i> test for mean differences Pearson's correlation with mean of absolute difference in scores Pearson's correlation with mean of differences	Age Gender Marital status Education Comorbidity Cancer type Duration of disease Presence and site of metastases DNR order	NB only total score evaluated Significant negative correlation between <i>age</i> and mean absolute difference for total score (e.g. better agreement if patients were older) Significantly smaller absolute mean differences if <i>patients had a comorbidity, DNR order, or brain metastases</i> In multivariable analyses, <i>brain metastases</i> and <i>age</i> were significantly

Study ID	Focus of evaluation	Measurement approach(es)	Factors evaluated	Significant results for QOL, PF, EF domains
Sneeuw et al. (1998) [42]	Compare proxy–patient differences across various characteristics	Scatter plot to visualize proxy–patient agreement by patient QOL levels Correlation between variables of interest and total QOL score Hierarchical regression analysis with total QOL score as outcome variable Differences measured as both absolute difference and directional difference	Performance status (Eastern Cooperative Oncology Group) Weight loss Mental health (Mental Health Inventory-5) Age Gender Education Social desirability (Socially Desirable Response Set-5) Positive appraisal (Utrecht Coping List) Social expressiveness (Utrecht Coping List)	associated with better agreement (smaller mean absolute differences) Scatter plots show better agreement (fewer differences) at either extreme end of patient total QOL score distribution, worse agreement in the middle Significantly larger absolute differences (worse agreement) for patients who were <i>older, female, with worse performance status, more weight loss, worse mental health, and stronger tendencies toward socially desirable responses</i> In multivariable analyses, only <i>socially desirable responses</i> remained significant Significantly larger directional differences (worse agreement) for <i>female patients, patients with positive coping styles, and patients with stronger tendencies toward socially desirable responses</i> In multivariable analyses, only <i>positive coping style</i> remained significant
Sneeuw et al. (1997) [32, 44] (study in brain cancer patients)	Evaluate association between number of proxy–patient responses without exact agreement and various characteristics, particularly patient neurological and physical characteristics	Number of discrepancies across all questions in the QLQ-C30, per proxy–patient pair ANOVA to compare mean number of discrepancies among relevant groups Test of linear trends in mean number of discrepancies (for multi-level variables only) Levels of agreement in the same category (exact) and within one response category (approximate) (for mental confusion only) Comparison of effect size	Performance status (Karnofsky Performance Status) Disease stage (recurrent vs newly diagnosed) Motor deficit Mental confusion Cognitive impairment Gender Age Race/ethnicity Marital status Education Duration of disease Treatment status	Significantly lower proxy scores (vs patient) for PF, EF, QOL among <i>patients with mental confusion</i> , but no significant differences among patients without Significantly more discrepancies in patients with <i>minor mental confusion</i> (vs <i>normal function</i>) Significant linear trend of more discrepancies as <i>performance status worsened</i> and <i>motor deficit increased</i> Worse/lower exact and approximate agreement in patients with mental confusion (vs those without) Moderate effect size (bigger proxy–patient differences) for PF, EF, QOL in patients with mental confusion, vs small effect sizes for patients without confusion
Wennman-Larsen et al. (2007) [46]	Compare proxy–patient differences, focusing on situations where proxies	Correlation between characteristics and mean proxy–patient differences, if mean differences had effect sizes >0.40	Age Gender Time from diagnosis to interview	Significantly worse concordance among <i>male</i> (vs <i>female</i>) patients for EF

Study ID	Focus of evaluation	Measurement approach(es)	Factors evaluated	Significant results for QOL, PF, EF domains
	underestimated function	Multiple regression, if mean differences had effect sizes >0.40	Time from interview to death	
McMillan (1996) [27]	Compare proxy–patient correlation in subgroups of QOL outcome	Patients grouped by score relative to median (above = high, below = low), then proxy–patient correlations conducted within each group	Patient QOL scores	Significant correlation in <i>patients with higher QOL</i> ; this was higher than the non-significant correlation in the lower QOL group
Deschler et al. (1999) [56]	Compare congruence across proxy types	Congruence defined as proxy score within 90% CI of patient score; calculated for each domain Chi square, Fisher's exact test to see if factors significantly associated with differences in congruence	Age Gender Disease stage/status (recurrent vs primary)	Non-significant results for all patient characteristics. Non-significant "tende[ncy]" for better congruence among patients with recurrent disease (vs primary)

* This study defined statistical significance as $p < 0.10$

Table 8

Search terms used

PubMed	PsycINFO	CINAHL
<p>#1 Search ("Proxy" [Mesh] or prox* or "patient agent" or "health care agent" or "healthcare agent" or family or caregiver or "next of kin" or spouse or husband or wife)</p> <p>#2 Search ("Quality of Life" [Mesh] or "quality of life" or "qualityoflife")</p> <p>#3 Search ("EQ5D" or "SF12" or "SF36" or "EORTC" or "FACT")</p> <p>#4 Search ("Neoplasms" [Mesh] OR cancer* OR cancers or cancerous or neoplasm* OR malignan* or "Medical Oncology" [Mesh])</p> <p>#5 Search (#2 or #3)</p> <p>#6 Search (#1 and #4 and #5)</p> <p>#7 Search ("Research" [Mesh] or research* or stud* or trial*)</p> <p>#8 Search (#7 and #6)</p>	<p>S1 "Proxy" or prox* or "patient agent" or "health care agent" or "healthcare agent" or family or caregiver or "next of kin" or spouse or husband or wife</p> <p>S2 "Quality of Life" or "quality of life" or "qualityoflife"</p> <p>S3 "Quality of Life" or "quality of life" or "quality-of-life"</p> <p>S4 Health related quality of life OR hrqol OR quality of life OR qol "Quality of Life" or "quality of life" or "qualityof-life"</p> <p>S5 Health-related quality of life OR hrqol OR quality of life OR qol or "Quality of Life" or "quality of life" or "quality-of-life"</p> <p>S6 "Neoplasms" OR cancer* OR cancers or cancerous or neoplasm* OR malignan* or "Medical Oncology"</p> <p>S7 ("Neoplasms" OR cancer* OR cancers or cancerous or neoplasm* OR malignan* or "Medical Oncology") AND (S1 AND S5 AND S6)</p>	<p>S1 "Proxy" or prox* or "patient agent" or "health care agent" or "healthcare agent" or family or caregiver or "next of kin" or spouse or husband or wives</p> <p>S2 Health-related quality of life OR hrqol OR quality of life OR qol "Quality of Life" or "quality of life" or "quality-of-life"</p> <p>S3 "Quality of Life" or "quality of life" or "quality-of-life"</p> <p>S4 "Neoplasms" OR cancer* OR cancers or cancerous or neoplasm* OR malignan* or "Medical Oncology"</p> <p>S5 ("Neoplasms" OR cancer* OR cancers or cancerous or neoplasm* OR malignan* or "Medical Oncology") AND (S1 AND S3 AND S4)</p>