Bone Marrow Transplantation (2012) 47, 473–482

www.nature.com/bmt

REVIEW

Quality of life in patients before and after haematopoietic stem cell transplantation measured with the European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Core Questionnaire QLQ-C30

N Grulke[1,2], C Albani[3,4] and H Bailer[1,2]

[1]*Luisenklinik—Zentrum für Verhaltensmedizin, Bad Dürrheim, Germany;* [2]*Department of Psychosomatic Medicine and Psychotherapy, Ulm University, Ulm, Germany;* [3]*Schussental-Klinik gGmbH, Aulendorf, Germany and* [4]*Department of Medical Psychology, Leipzig University, Leipzig, Germany*

The EORTC Quality of Life Core Questionnaire QLQ-C30 is widely used, but no reference values are available for patients receiving HSCT. We retrieved data for 38 samples from 33 papers in English and German that provided evaluable information on QLQ-C30 scores (mean, s.d.) covering about 2800 patients. Results are presented as a table that provides reference data that allow QLQ-C30 scores at different points during the disease trajectory to be put in context. With respect to their central tendency and their variance, scores vary over time. Quality of life is lowest during inpatient time. About 1 year after HSCT, the pre-transplant level is reached. Physical functioning is the scale reaching the highest level of all scales. Fatigue, dyspnoea and insomnia are symptoms that remain at an elevated level and should thus be considered as persisting problems after HSCT. For the interpretation of differences between scores, a very conservative recommendation would be to set the s.d. at 30 points. Doing so, one could be quite sure of having found a clinically significant change if the difference of two scores exceeds 15 points. Differences below 5 points should be interpreted with caution.
Bone Marrow Transplantation (2012) **47,** 473–482; doi:10.1038/bmt.2011.107; published online 23 May 2011
Keywords: quality of life; review; QLQ-C30; haematopoietic stem cell transplantation; reference data

Introduction

The concept of health-related quality of life (HRQoL) has gained much interest during the last decades and is considered a second endpoint (besides survival time) for evaluation following cancer treatment. Several instruments measuring HRQoL are applied in the field of oncology, with the 'European Organization for Research and Treatment of Cancer' (EORTC) Quality of Life Core Questionnaire QLQ-C30[1] being the instrument most frequently used. This questionnaire is most often implemented in clinical trials[2] and is the current standard in measuring HRQoL in Europe.[3] This review will focus on the QLQ-C30 in the field of haematopoietic stem cell transplantation (HSCT).

The development of the EORTC questionnaire started in the 1980s,[4] with its current version being 3.0. The QLQ-C30 is a 30-item instrument and has been extensively tested for reliability and validity. It has been translated into 81 languages (http://groups.eortc.be/qol/questionnaires_qlqc30.htm; 15 accessed June 2010). In addition to the generic tool of HRQoL, several disease-specific modules have been added, for example, for lung or breast cancer. The module added for high-dose chemotherapy (QLQ-HDC29) has recently passed the study phase 3.[5]

The QLQ-C30 addresses six multi-item scales of functioning (physical, role, social, emotional and cognitive functioning, global health). Furthermore, three multi-item scales (fatigue, nausea and vomiting, pain) as well as six single items (dyspnoea, constipation, sleeping problems, appetite loss, diarrhoea, financial problems) are related to symptoms or problems. In the current version (for a specimen see http://groups.eortc.be/qol/downloads/modules/specimen_20qlq_c30.pdf; accessed 15 June 2010), 28 of 30 items are measured by a four-point Likert-type scale (1 = not at all, 2 = a little bit, 3 = quite a bit, 4 = very much), with the two remaining items, addressing global physical health and global quality of life, being measured by a seven-point scale. The raw scores for all multi- and single-item scales are linearly transformed to 0-to-100-scales, with '100' reflecting highest functioning or highest symptomatology, respectively.[6] For clinicians and researchers alike, the question is how to interpret these scores gained from the QLQ-C30. Sloan *et al.*[7] have given an overview of different strategies to approach this question.

Correspondence: Dr H Bailer, Klinische Forschung, Luisenklinik—Zentrum für Verhaltensmedizin, Luisenstraße 56, D-78073 Bad Dürrheim, Germany.
E-mail: h.bailer@luisenklinik.de
Received 28 October 2010; revised 17 March 2011; accepted 28 March 2011; published online 23 May 2011

Norms to interpret QLQ-C30 raw scores can be found for a limited number of distinct subgroups, derived from representative surveys. Such norms are available for women in Denmark,[8] as well as for the general population in Norway,[9] Sweden[10] and Germany.[11] However, Fayers[12] remarks that there are considerable discrepancies between normative data from Germany and the Scandinavian countries for different age groups. He concludes that there seem to be intercultural differences between different countries. If no representative, that is, culture-independent national data are available, it could be problematic to compare one's results with normative data from another country. For example, Bieri *et al*.[13] and Hendriks and Schouten[14] compared their data drawn from samples in Switzerland and the Netherlands, respectively, to the Norwegian population,[9] whereas Goodman *et al*.[15] compared their data from the United States of America to the German population.[11] Regarding the study by Bieri *et al*.,[13] the question arises whether the data from German (normative) would have been more suitable for a comparison than the Norwegian data.

The EORTC Quality of Life Group provides reference values for QLQ-C30 items and scales according to the main cancer sites.[16] Unfortunately no reference values for patients treated with high-dose chemotherapy and HSCT have been provided yet. This review summarizes available data on QLQ-C30 scores (mean, s.d.), gathered from published studies on HSCT (German and English publications). Our aim is to provide reference values for researchers in the field who use the QLQ-C30 in the treatment of HSCT patients.

Methods

Several databases (MedLine, PubMed, Psyndex), and Google Scholar beta (http://scholar.google.de/), were scanned by keywords ('QLQ-C30', 'marrow,' or 'stem cell' and 'transplantation') and reference lists were examined for further relevant citations. The identified publications were screened and its data were included if the study published adhered to the following three criteria: the QLQ-C30 was used; the patients examined had received autologous or allogeneic HSCT and results for all or some of the QLQ-C30 scales were reported. We identified 43 papers fulfilling the criteria; 34 papers were subjected to our final review, with nine papers being excluded.

Nine studies were excluded owing to either amalgamation of items from the QLQ-C30 with items from other sources, a change in the wording,[17–19] calculation of scores not following the EORTC recommendations,[17,18,20] insufficient report of scores (as medians or as graphics only),[21–24] and one showed correlations of the QLQ-C30 with other instruments, but no descriptive values for the QLQ-C30 itself.[25]

In 34 papers arithmetic means (and to a lesser extent s.d.s) for QLQ-C30 scales were reported.[13–15,26–56] In most instances, the data used had been published in scientific journals. Data were retrieved from dissertations[29,32] and a monograph[36] for three German papers. We retrieved information regarding the country in which the study took place; the total number of cases included in each sample, and the number of cases split for allogeneic and autologous transplants; the percentage of males in the sample; the mean, median and/or range of age at transplant or at follow-up; diagnoses (categorized into four groups, namely acute leukaemia, chronic myeloid leukaemia, other haematological diseases and solid tumours); the version of the QLQ-C30 in use; and arithmetic means and—if available—s.d.s for each scale of the QLQ-C30 (mean and s.d.s rounded to the next integer).

In 13 of the 34 papers data for repeated measures were provided.[14,26–29,32,34,35,37–40,43] Some researchers, that is, the working groups of Frick[30,31,33] Hjermstad[26,27] and Kopp[42,56] repeatedly reported on their cases over the years. It was not always clear to what extent samples described in different publications were independent of each other. We excluded one study (which reports data for only two QLQ-C30 scales) of the working group of Frick[31] because this sample was pooled with data from two other study centres and published separately.[33] The working group of Hjermstad reported QoL data for their patients before HSCT twice[26,27] We analysed the data for the time before HSCT from the second study,[27] that is, we excluded the four distinguishable samples from the first study[26] for the time before HSCT because these samples were pooled for the report in the second study.[27] Both studies provided data at 1 year[26] and 3–5 years[27] after HSCT. We treated the data provided by the working group of Kopp[42,56] as data stemming from three samples. The two samples described in the first paper[42] are definitely separate. The patients covered in the sample in the second paper[56] are probably a subgroup from the two samples in the first paper (not referenced in the later one) being interviewed 5 years after the first interview. We included this small sample because we did not want to lose the data on this long-term follow-up. Thus, samples for our further analyses were extracted from 33 papers.

The sample sizes we will refer to are not identical with the sizes of the samples covered in the papers in total. Some papers compare transplant patients to patients receiving conventional chemotherapy; for example, Messerer *et al*.[53] report on 419 AML patients of whom 121 received HSCT, 121 being the sample size that will be relevant for our purpose in this case. Other papers report on disparate samples, for example, Kopp *et al*.[42] performed a cross-sectional study and interviewed 56 patients after HSCT. They divided their sample in two subsamples of 15 and 41 patients who had received HSCT in the last 12 months or more than 12 months ago, respectively. QLQ-C30 data were reported for these two groups, so the sample sizes of 15 and 41 will be relevant. Thus, our procedure may foster the impression that sample sizes in the field are small in general, but they are not necessarily so.

We assigned the samples to seven categories, categorized by time of data collection in relation to HSCT: before transplantation, during inpatient stay after HSCT, at discharge after HSCT, up to 6 months after HSCT, 7–12 months after HSCT, more than 1 up to 3 years after HSCT and more than 3 years after HSCT. The assignment of a sample to one of these categories was based on the average time reported in the studies. With respect to these

seven categories we were able to derive 67 samples from the 33 papers providing arithmetic means for at least one of the 15 QLQ-C30 scales. S.d.s are given for 55 of these samples.

The information on QLQ-C30 scores was summarized as described in the following sections. *Results referring to arithmetic means:* For each of the seven measure points and each of the 15 QLQ-C30 scales, we calculated unweighted means and medians, as well as means and medians weighted by sample size for the numbers provided. This resulted in four matrices with 7 by $15=105$ entries. We compared the matrix covering unweighted means with the matrix covering weighted means, and correspondingly for medians; analogously, we compared unweighted means with unweighted medians, and correspondingly for weighted means and medians. These four comparisons were arithmetically performed as the calculation of the mean of the 105 absolute differences. For example, the comparison of unweighted means with unweighted medians resulted in an average absolute difference of 1.1; in three of the 105 cases the difference exceeded 4 points (4.3, 4.5 and 6.2). The mean differences for the other three comparisons were 1.7, 2.1 and 2.7. In the latter case (comparing weighted and unweighted medians) eight differences above 4 points (maximum observed difference 20.5) occurred with the measure point during inpatient time, thus increasing the overall mean difference. The fact that the values are based on only two samples with different sample sizes is decisive for the difference between the weighted and unweighted medians. In view of this, we based our further analyses on the unweighted arithmetic means that were reported for the different samples. Doing so has the further advantage that data to come, that is, arithmetic means, could easily be included in the calculation for updated reference scores.

Results referring to standard deviations. Analogously to the previous step, we analysed the s.d.s provided. We found the following average absolute differences for the four matrixes: 0.9, 1.3., 1.4 and 2.0. As before, the highest average absolute difference (2.0) was associated with weighted and unweighted medians, and the smallest (0.9) with unweighted means and unweighted medians. The results of these analyses motivated us to restrict further evaluations to the unweighted arithmetic means.

Results

Sample characteristics

Considering only the initial sample size, that is, ignoring data for repeated measures, the sample sizes of the 38 samples included varied between 15 and 415 patients. The median sample size was 39.5 (mean 73.8; s.d. 88.6), and the 25th and 75th percentiles were 24 and 83.75, respectively. Eight samples had sizes of $n \geqslant 100$. In total, these 38 samples comprised 2804 patients, 52.6% of whom received an allogeneic HSCT, with the remaining 48.4% receiving an autologous HSCT. In 14 of these 38 samples (covering 1102 patients), allogeneic and autologous HSCT were combined ('mixed samples'); 13 samples (804 patients) focused on allogeneic HSCT and 11 samples (898 patients) on autologous HSCT only.

Within the samples included, patients were either awaiting HSCT or had survived up to 24 years after transplantation. Most patients were examined before the start of HSCT treatment. As not every study included reported on each of the QLQ-C30 scales, for example,[30,38,40] the sample size on which the calculation of the mean of each scale was based varies (see Table 1, bottom part).

At the time of HSCT, patients were between 14 and 70 years old, on average about 40–45 years. The patients were treated for acute leukaemia (28.0%), chronic myeloid leukaemia (15.3%), other haematological diseases (42.1%, in most cases lymphoma) and solid tumours (14.8%, in most cases breast cancer). In total 50.1% of all patients were male.

Most of the samples (12 of 38) were recruited in Germany[29–33,35,36,38,42,47,50,53,56] and comprised 1116 patients, that is, 39.8% of all patients included; 447 patients (15.9%, seven samples) were recruited in Northern European countries,[26,27,37,39,40,46] namely Denmark, Finland, Norway and Sweden, and 644 patients (23.0%, 13 samples) were recruited from other European countries.[13,14,42,44,45,48,49,52,54–56] Data for the remaining 597 patients (21.3%, six samples) were provided from non-European countries, namely Canada,[41] Taiwan[51] and the United States of America.[15,28,43] We found three significant differences when comparing the QLQ-C30 scores from the German samples to those from the non-German samples for all scales and measure points, that is, $P<0.05$ for 3 out of 105 comparisons.

Eleven papers stated that version 3.0 of the QLQ-C30 was used. In three papers QLQ-C30 version 1.0 was mentioned. Version 2.0 was mentioned in two and version +3 in one paper. No explicit information was given in the remaining 16 of the 33 papers.

QLQ-C30 scores

We compared scores from 'mixed samples' to the samples that comprised patients having allogeneic or autologous transplants only, respectively, for all 15 scales and 7 measure points. Some significant ($P<0.05$) differences were found, but considering the α-level accumulation for multiple tests, these findings probably were 'significant' by chance. A visual inspection of the graphical illustrations (Figure 1) suggests that overall the group of patients having allogeneic transplants shows better functioning and less symptomatology than the autologous group for all but the last measure point, and the results of the 'mixed' group may be situated between those two groups. More than 3 years after HSCT there may be a change resulting in a slight advantage for the autologous transplants. But the differences between the groups in most cases are small, and the numbers of samples comprising a group are even smaller (for example, only one sample for the autologous group for measure points inpatient time, at discharge and up to 6 months after HSCT) and are therefore susceptible to random error. Thus, we decided to combine the groups into one for subsequent calculations.

Table 1 Average values (rounded to integers) for means and s.d.s, minimum and maximum (in parentheses) for the QLQ-C30 scales as reported in the reviewed studies by time in relation to HSCT

QLQ-C30 scale		*Before HSCT*	*During hospitalization*	*At discharge*	*Up to 6 months after HSCT*	*7–12 months after HSCT*	*More than 1 up to 3 years after HSCT*	*More than 3 years after HSCT*
GH	Mean (range)	63 (39–76)	38 (32–43)	53 (40–64)	60 (51–73)	67 (59–73)	72 (67–76)	74 (66–82)
	s.d. (range)	20 (12–25)	18 (13–23)	19 (15–22)	21 (16–30)	22 (19–24)	21 (20–22)	21 (14–28)
PF	Mean (range)	80 (66–89)	54 (53–54)	64 (53–75)	72 (61–87)	81 (73–89)	81 (73–91)	86 (78–92)
	s.d. (range)	17 (9–24)	29 (28–29)	21 (17–24)	21 (8–27)	21 (19–22)	20 (15–23)	18 (12–33)
EF	Mean (range)	68 (54–86)	66 (58–74)	65 (54–79)	75 (56–93)	74 (54–84)	74 (64–82)	74 (57–86)
	s.d. (range)	23 (17–31)	27 (24–29)	23 (20–25)	23 (9–31)	22 (19–26)	22 (20–25)	25 (19–32)
CF	Mean (range)	81 (70–89)	68 (63–73)	69 (62–79)	79 (64–89)	82 (73–89)	79 (71–93)	77 (69–87)
	s.d. (range)	21 (13–27)	28 (26–30)	24 (20–27)	24 (14–33)	25 (15–34)	21 (19–23)	26 (17–31)
RF	Mean (range)	62 (38–80)	34 (34–37)	41 (29–52)	54 (32–77)	75 (63–88)	72 (58–81)	78 (60–91)
	s.d. (range)	30 (17–36)	37 (34–40)	36 (28–43)	33 (24–43)	28 (21–33)	30 (29–31)	28 (21–34)
SF	Mean (range)	63 (48–83)	52 (46–57)	52 (42–72)	63 (53–81)	76 (65–87)	73 (64–87)	77 (61–88)
	s.d. (range)	29 (21–36)	32 (31–33)	31 (23–41)	30 (19–36)	24 (20–32)	25 (20–28)	28 (20–35)
FA	Mean (range)	38 (20–54)	70(69–71)	58 (47–71)	49 (30–61)	33 (24–46)	32 (26–41)	30 (21–40)
	s.d. (range)	25 (19–30)	30 (29–31)	24 (20–26)	27 (18–37)	22 (13–29)	25 (23–28)	25 (19–30)
NV	Mean (range)	13 (5–22)	63 (56–69)	31 (15–47)	14 (4–27)	8 (3–15)	7 (3–14)	4 (1–10)
	s.d. (range)	23 (14–33)	32 (27–36)	30 (17–38)	21 (8–31)	18 (9–27)	13 (12–15)	11 (3–19)
PA	Mean (range)	18 (2–36)	47 26–67)	34 (24–46)	21 (4–29)	21 (15–28)	22 (14–28)	16 (11–28)
	s.d. (range)	24 (8–31)	38 (37–38)	29 (20–38)	25 (10–38)	24 (16–27)	26 (26–27)	23 (17–32)
AP	Mean (range)	18 (11–24)	76 (64–87)	53 (39–78)	33 (4–54)	11 (1–17)	10 (6–13)	8 (0–14)
	s.d. (range)	28 (17–35)	34 (28–40)	34 (32–35)	32 (12–41)	16 (7–21)	18 (13–23)	16 (0–27)
CO	Mean (range)	8 (2–22)	19 (18–19)	11 (4–18)	6 (0–24)	7 (2–12)	9 (7–11)	8 (3–16)
	s.d. (range)	16 (8–27)	30 (28–31)	23 (11–32)	15 (0–37)	16 (11–23)	18 (15–20)	19 (9–29)
DI	Mean (range)	15 (8–21)	49 (47–51)	38 (20–57)	16 (9–33)	12 (1–19)	12 (4–16)	10 (4–23)
	s.d. (range)	26 (18–34)	39 (33–44)	33 (27–39)	24 (18–36)	19 (7–29)	20 (11–25)	18 (11–29)
DY	Mean (range)	24 (14–33)	24 (22–26)	29 (24–35)	27 (16–44)	26 (16–35)	19 (6–33)	20 (13–34)
	s.d. (range)	30 (20–38)	35 (33–37)	32 (25–39)	29 (17–43)	32 (27–38)	19 (13–23)	24 (3–36)
SL	Mean (range)	28 (11–42)	50 (49–50)	44 (22–73)	25 (13–33)	23 (14–31)	26 (9–37)	26 (13–40)
	s.d. (range)	31 (22–38)	38 (35–40)	31 (26–36)	29 (17–36)	30 (23–34)	28 (27–30)	30 (23–36)
FI	Mean (range)	28 (15–46)	35 (29–40)	31 (22–40)	36 (21–50)	23 (19–37)	25 (7–31)	23 (11–42)
	s.d. (range)	33 (28–42)	39 (37–40)	33 (31–36)	34 (29–40)	31 (29–32)	28 (18–34)	32 (18–41)
Data were based	Mean n_p	758–1090	49	93–127	134–225	551–753	692–734	674–878
on n_p patients	Mean n_s	13–18	2	4–6	6–11	5–7	5–6	13–16
from n_s samples	s.d. n_p	355–719	49	93–127	134–225	44–661[a]	355	566–735
	s.d. n_s	10–15	2	4–6	6–11	2–5	3	10–13

Abbreviations: AP = appetite loss; CF = cognitive functioning; CO = constipation; DI = diarrhoea; DY = dyspnoea; EF = emotional functioning; FA = fatigue; FI = financial difficulties; GH = global health; HSCT = haematopoietic stem cell transplantation; NV = nausea and vomiting; PA = pain; PF = physical functioning; RF = role functioning; SF = social functioning; SL = insomnia.
[a]Extreme outlier (scale FI); ignoring the outlier the range would be 459–661.

Split by measure points, Table 1 shows the central tendency and the range of the arithmetic means and s.d.s reported in the studies included into this analysis. Overall, the results show a clear decrease in functioning and an increase in symptomatology from the testing time 'before HSCT' to the testing time 'during hospitalization'. Scores 'during hospitalization' and 'at discharge' are the lowest for the functioning scales and the highest for the symptom scales, with two exceptions: for the symptom scales dyspnoea (DY) and financial difficulties (FI) only one of the two scores belongs to the two highest of all seven values.

With regard to the maximum difference between two measure points apart from 'hospitalization' and 'at discharge', we can roughly discriminate four groups with respect to the magnitude of these differences. The first group is characterized by quite stable scores, with a low maximum difference of 4 or 5: cognitive functioning (CF), constipation (CO) and sleeping problems (SL). The second group shows moderate maximum differences ranging from 6 to 10: emotional functioning (EF), nausea and vomiting (NV), pain (PA), dyspnoea (DY), and DI. Large differences (range 13–16) were found for global health (GH), physical functioning (PF), social functioning (SF) and FI, and very large differences (range 19–25) for role functioning (RF), fatigue (FA) and appetite loss (AP).

GH–global health

The clear decrease towards inpatient time is followed by a continuing moderate increase after discharge. One year and more after HSCT the maximum score (about 10 points above the pre-transplant level) is reached.

PF–physical functioning

The pre-transplant level seems to be reached between 7 and 12 months after HSCT. Thereafter only a slight improvement was measured in the sample. This course parallels that of GH albeit on a higher level (difference about 12 points).

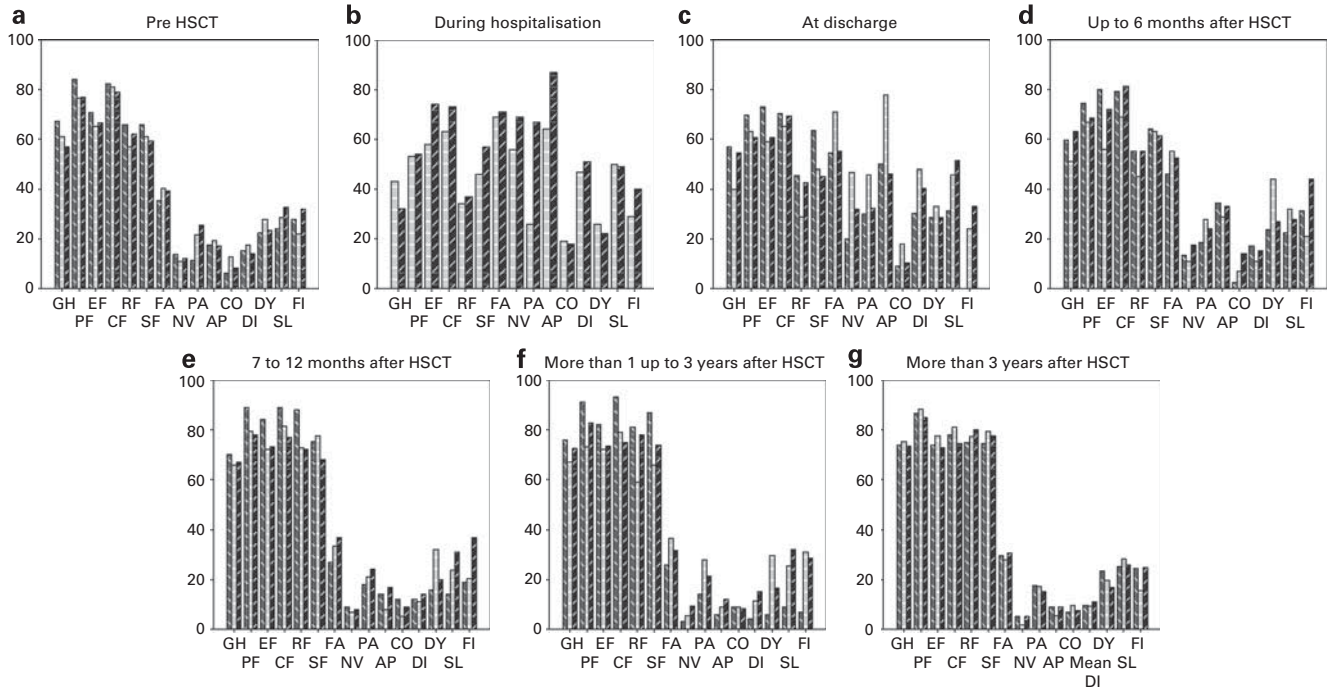

Figure 1 (**a**–**g**) Mean scores for QLQ-C30 scales for allogeneic, autologous and mixed samples by measure points. The bars in medium grey (dashed from upper left to lower right) refer to allogeneic, light grey (dashed horizontally) to autologous and dark grey (dashed from lower left to upper right) to mixed samples. AP, appetite loss; CF, cognitive functioning; CO, constipation; DI, diarrhoea; DY, dyspnoea; EF, emotional functioning; FA, fatigue; FI, financial difficulties; GH, global health; NV, nausea; PA, pain; PF, physical functioning; RF, role functioning; SF, social functioning and vomiting; SL, insomnia.

EF–emotional functioning
A minor decrease from 68 before transplant to 66 and 65 at inpatient time and discharge, respectively, and thereafter an increase by 10 points could be seen. The scores after discharge are practically on the same level (74). All seven scores for EF range from 65 to 75, that is, EF (besides DY) is the scale with minimal variation over time, even though when the two inpatient measure points were considered.

CF–cognitive functioning
The decrease seen towards inpatient time is reversed after discharge. Only slight variations were seen thereafter.

RF–role functioning
RF reflects limitations in doing one's work or daily activities, as well as pursuing one's hobbies or leisure activities. While being an inpatient, opportunities for these activities are naturally limited, thus the decrease in RF seen after admission to hospital is not surprising. Although there is substantial improvement after discharge (13 points), the essential improvement (21 points) happens from the first to the second half year after HSCT. This corresponds to clinical experience and corroborates the practice of physicians in our transplant wards who advise patients that recovery takes up to 1 year.

SF–social functioning
SF addresses the impact of physical condition and medical treatment on family life and social activities. A moderate decrease is followed by an improvement during the first year after discharge. As with GH and RF, after at least 1 year after HSCT, SF exceeds the pre-transplant level by more than 10 points.

FA–fatigue
Of all nine symptom scales, FA has the highest score before HSCT (38). FA increases during inpatient time and remains the prominent symptom throughout, that is, FA scores the highest in relation to all other symptom scales, even though through time it decreases to the (high) pre-transplant level. FA is a major concern persisting over time.

NV–nausea and vomiting; AP–appetite loss
Besides AP, NV is the symptom scale with the most substantial increase from the pre-transplant to the inpatient time (50+ points). After discharge it continuously decreases over time. Three and more years after HSCT, the symptom scale NV has the lowest score overall. After discharge, NV could be considered a minor problem (scores $\leqslant 8$), but AP remains a problem for up to 6 months. Beyond 6 months, it, too, seems to be a minor problem (scores $\leqslant 10$).

PA–pain
Pain increases during inpatient time, returning to its pre-transplant level after discharge. With mean scores of about 20 points, PA seems so be a significant problem for only a few patients, with most being less distressed.

CO–constipation
CO is not a typical problem in HSCT. There is little variation over time, with scores ranging below 20 points for the inpatient time (hospitalization, discharge) and not exceeding 10 points for all other time periods.

DI–diarrhoea
During inpatient time, DI shows a large increase of 34 points. The pre-transplant level is reached after discharge, and further slight improvements could be seen thereafter, reaching a minor problem level (10 points).

DY–dyspnoea
Like EF, DY is another scale with minimal variation over time. For all testing times, mean scores were between 19 and 29 points. No significant impact of treatment during inpatient time could be observed. Dyspnoea must be considered as a moderate, but persistent, problem.

SL–insomnia
Sleeping problems increase during inpatient time and return to the pre-transplant level after discharge (about 25 points). This seems to be a persistent symptom on a rather high level, although not reaching the level of FA.

FI–financial difficulties
The financial impact of the disease and its treatment are relevant at all times (scores ranging from 23 to 36).

With one exception, namely GH, the highest s.d. for each scale is reported for the time during hospitalization. For 10 scales, the s.d. reported for the time of discharge is the second highest. Ignoring the two inpatient periods, the s.d. for the other time periods varies only slightly ($\leqslant$6 points) with the exception of DI (8 points), NV (12 points), DY (13 points) and AP (16 points).

Regarding the last two measure points, the difference for reported values is 5 points for DY and CF, 4 points for FI and $\leqslant$3 points for all other scales. For most of the scales these differences do not change when the period '7–12 months after HSCT' is included.

Discussion

In 33 papers we identified 38 different samples of patients receiving HSCT who filled in the QLQ-C30. In total, our review covers data from about 2800 patients. Table 1 shows these data, differentiated by time span from measurement to HSCT. As an instrument of the EORTC—unsurprisingly—the QLQ-C30 was applied primarily in studies from European countries, mainly Germany.

With respect to the functioning scales, the two more physically based scales PF and CF were initially on a higher level (about 80 points) than the more psychologically based scales GH, EF, RF and SF (about 63 points); out of these six scales, PF reaches the highest level for the period '3 years and more after HSCT' (86 points). The other five scales score at about 75 points. A decrease in symptoms back to pre-transplant level was visible in most cases, in some even an improvement. The impact of treatment during inpatient time is seen on almost all scales, with some exceptions: DY emerged as quite stable over time, and no striking increase from pre-transplant time to inpatient time was seen for CO and FI (9 points). The increase exceeded 20 points for all other scales.

In general, quality of life decreases directly after HSCT, increases towards the discharge time and improves further thereafter. After a year, the level of HRQoL as measured during the weeks before HSCT is reached, and further increases may be seen. PF is the scale reaching the highest level of all scales. Fatigue, dyspnoea and insomnia are symptoms that remain on a level with scores $\geqslant$20 and should thus be considered persisting problems after HSCT, with scores above the general population.

We compared patients undergoing allogeneic and autologous transplantation. Although we could not show a statistically significant difference between the three kinds of samples, an inspection of the graphical comparisons suggests that the allogeneic transplants may show better quality of life than autologous transplants, that is, higher functioning and lower symptomatology, but in the long run the chances may reverse. We therefore suggest that research papers in the future should separate quality-of-life data for allogeneic and autologous transplants if both treatment modalities were studied together. A finding that corroborates this suggestion is the study by Loberiza *et al*.[57] who found a significant difference between autologous and allogeneic transplant patients with respect to depression, depression potentially being a risk factor in the immediate pre-transplant period.[58]

The QLQ-C30 is a developing instrument. It is likely that in some of the studies reviewed in which information about the version of the QLQ-C30 was not provided, version 1.0 was used. The missing statement may be because of the fact that at that time no other version was under discussion and therefore no uncertainty could be anticipated. For our review, this missing information did not present any major problems, but nevertheless, studies to come should explicitly state the version that was used for data collection.

That diagnoses are given should be taken for granted; not doing so, as was the case in one study,[31] should be the exception. With respect to descriptive information about the samples and purpose of our review, we would have welcomed descriptive statistics about the actual samples for which QLQ-C30 data were presented. For example, in one study[52] pre-transplant characteristics (such as gender, age, diagnosis) were shown for 155 patients, but QLQ-C30 data were available for only 106 patients, that is, about one-third of the sample was missing; the question must therefore be raised if descriptive information is representative for QLQ-C30 data. Such an attrition of patients is seen most notably in longitudinal long-term studies with repeated measures, for example, in one study with four measure points[43] the numbers of cases over 4 years were 415–217–117–35, but descriptive data were reported for the sample of 415 patients only.

Most of the samples analysed were small, range 15–40. Only one-quarter of the samples had sizes above 83.

Especially for longitudinal studies with repeated measures, significant drop-out rates are problematic. One of 3 patients must be expected to drop out during the first year after HSCT: the ratio of patients not interviewed at a specific time after HSCT to patients interviewed before HSCT was 36% at 100 days,[38] 33% at 6 months,[37] and 33[39] and 35%[29] at 1 year after HSCT. This ratio increases with time after HSCT, for example, 42% at 3–5 years after HSCT.[27]

If we consider evaluating interventions and comparing, for example, two groups with $\alpha = 5\%$ and power $= 80\%$, we need 64 cases in both groups to detect a significant difference of a 1/2 s.d. (assumed to be equal in both groups),[59] that is, the total sample size should at least be 128. To detect smaller effects the necessary sample size increases and, depending on study purpose, drop-outs must be considered. These considerations highlight the difficulties in clinical studies. The problem of missing values in these instruments exacerbates the situation particularly when patients are requested to fill in questionnaires.

From the studies providing normative data for the general population,[8–11] we learned that differences exist between men and women, and between age groups. Roughly summarized, men report better functioning than women, and with increasing age less functioning and more symptoms are seen. These topics were rarely discussed in the studies reviewed, and data opposing different socio-demographic groups were not provided. Therefore, the distribution of these tendencies in the population of HSCT patients remains unclear. A central registry (for example, run by the EORTC Quality of Life Group) would be helpful in collecting such data and would increase the data basis for reference scores. Further, such a registry could gather data to decide on the hypothesis raised by Fayers[12] about intercultural differences between Germany and Northern European countries. We found no convincing evidence that the scores from German samples differ substantially from those from non-German ones. This fact does not rule out the existence of intercultural differences, but it may be seen as a hint that these differences probably are small. To decide on their clinical relevance remains an unanswered empirical question.

The courses over time, as outlined above, do not necessarily reflect the courses of the individuals. The question of how much patients follow these courses that were developed by comparing group data over time remains unanswered. The group data suggest a statement to the patient like: 'in one year you will probably have reached your pre-transplant level or improved somewhat.' But we do not know for how many patients this forecast will hold true. There is still a proportion of patients who suffer from long-term side effects, such as chronic GVHD,[60] and do not reach their pre-transplant level of HRQoL. Syrjala *et al*.[61] concluded from their long-term study that full recovery is a 3- to 5-year process; that is, the 1-year perspective might be too optimistic. For the individual patient it would be helpful if we could provide a more tailored prognosis on the course of recovery and the HRQoL to be expected.

There are several rules of thumb about how to interpret differences between two scores on the QLQ-C30. One paper states that a difference of about 10–15 points should be considered clinically significant.[3] A more differentiated view is provided by Osoba *et al*.[62] who quantified changes as reported by patients: a 'little' change corresponded to differences in the range of 5–10 points on the QLQ-C30; a 'moderate' change corresponded to differences of 10–20 points; and 'very much' change corresponded to differences above 20 points. King[63] pointed out that in patient groups with decreased health status, higher s.d.s were found than in groups with better health. This means that the minimal difference significant in the clinical setting depends on the clinical context, and therefore no general rule can be given. Fayers *et al*.[4] summarized other research in that patient reports of a moderate change were associated with an absolute change of 10% on the respective quality-of-life scale, and a small change was associated with an absolute change of 5–10% on the scale, irrespective of the instrument used. These numbers resemble those reported by Osoba *et al*.[62] Furthermore, Fayers *et al*.[4] see these results as consistent with the concept of Cohen's effect size.[64]

King's[63] observation that higher s.d.s were found with decreased health status finds its replication in our analysis, too, when comparing the values of inpatient time and at discharge with all other measure points. Disregarding the values for inpatient time measures, the s.d.s for the scales are on quite different levels. Therefore, denoting a single number for all scales, for example, 15, seems not to be appropriate. For GH and PF the s.d. could be set at about 20 and for CO at about 17, irrespective of the measure point. By contrast, scores for RF show a considerably higher variation; therefore, calculations should be based on a s.d. of about 30. The s.d.s for the other three functioning scales should be set between these numbers, roughly 25. For the symptom scales we see a similar picture. SL and FI have high s.d.s that could be set about 31. FA and PA have s.d.s about 25. For the other symptom scales, the picture is less clear. For NV, AP, DI and DY the mean score for symptomatology as well as the corresponding s.d.s decrease over time. In these cases a single number might be appropriate for measurements 1 or more years after HSCT, but not for the periods before.

Conclusion

Table 1 can be considered as the essence of our review and provides reference data that allow putting in context QLQ-C30 scores of HSCT patients at different points during their disease trajectory. No simple rule of thumb for all situations can be made as central tendency as well as variation of QLQ-C30 scores vary over time (and from study to study as can be seen from the given ranges). But if such a simplification were needed, a very conservative recommendation would be to set the s.d. at 30 points. Doing so, one could be quite sure of having found a clinically significant change if the difference of two scores exceeds 15 points. With respect to the variation of the reported numbers and the different sample characteristics we would not put too much weight on differences below 5 points.

For reports on QLQ-C30 data in the future we recommend to explicitly state the version of the QLQ-C30 that was used, to report information about the diagnoses, to differentiate between men and women, and to report scores separately for autologous and allogeneic transplants.

Conflict of interest

The authors declare no conflict of interest.

References

1 Aaronson NK, Ahmedzai S, Bergmann 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 trials in oncology. *J Natl Cancer Inst* 1993; **85**: 365–376.
2 Efficace F, Bottomley A, Coens C. Which is the most frequently used health-related quality of life (HRQOL) measure in cancer clinical trials? (abstract). *Psychooncology* 2006; **15**(2 (Suppl)): S135–S136.
3 Küchler T, Behrend M. Der onkologische Patient—Lebensqualität und supportive Therapie. *Im Focus Onkologie* 2001; **4**: 49–52.
4 Fayers P, Bottomley A. Quality of life research within the EORTC—the EORTC QLQ-C30. *Eur J Cancer* 2002; **38**(Suppl 4): S125–S133.
5 Velikova G, Weis J, Hjermstad MJ, Kopp M, Morris P, Watson M *et al*. The EORTC QLQ-HDC29: a supplementary module assessing the quality of life during and after high-dose chemotherapy and stem cell transplantation. *Eur J Cancer* 2007; **43**: 87–94.
6 Fayers PM, Aaronson NK, Bjordal K, Curran D, Groenvold N. *The EORTC QLQ-C30 Scoring Manual*. 2 edn. European Organization for Research and Treatment of Cancer: Brussels, Belgium, 1999.
7 Sloan JA, Frost MH, Berzon R, Dueck A, Guyatt G, Moinpour C *et al*. The clinical significance of quality of life assessments in oncology: a summary for clinicians. *Support Care Cancer* 2006; **14**: 988–998.
8 Klee M, Groenvold M, Machin D. Quality of life of Danish women: population-based norms for the EORTC QLQ-C30. *Qual Life Res* 1997; **6**: 27–34.
9 Hjermstad MJ, Fayers PM, Bjordal K, Kaasa S. Health-related quality of life in the general Norwegian population assessed by the European Organization for Research and Treatment of Cancer Core Quality-of-Life Questionnaire: the QLQ = 30 (+3). *J Clin Oncol* 1998; **16**: 1188–1196.
10 Michelson H, Bolund C, Nilsson B, Brandberg Y. Health-related quality of life measured by the EORTC QLQ-C30: reference values from a large sample of the Swedish population. *Acta Oncol* 2000; **39**: 477–484.
11 Schwarz R, Hinz A. Reference data for the quality of life questionnaire EORTC QLQ-C30 in the general German population. *Eur J Cancer* 2001; **37**: 1345–1351.
12 Fayers PM. Interpreting quality of life data: population-based reference data for the EORTC QLQ-C30. *Eur J Cancer* 2001; **37**: 1331–1334.
13 Bieri S, Roosnek E, Helg C, Verholen F, Robert D, Chapuis B *et al*. Quality of life and social integration after allogeneic hematopoietic SCT. *Bone Marrow Transplant* 2008; **42**: 819–827.
14 Hendriks MGJ, Schouten HC. Quality of life after stem cell transplantation: a patient, partner and physician perspective. *Eur J Int Med* 2002; **13**: 52–56.
15 Goodman KA, Riedel E, Serrano V, Gulati S, Moskowitz C, Yahalom J. 2584: Health-related quality of life in long-term survivors of autologous hematopoietic-cell transplantation for refractory and relapsed Hodgkin's lymphoma. *Int J Radiat Oncol Biol Phys* 2006; **66**(3, Supplement 1) S534–S535 (abstract 2584).
16 Scott NW, Fayers PM, Aaronson NK, Bottomley A, de Graeff A, Gundy C *et al*. *EORTC QLQ-C30 Reference Values* (online: http://groups.eortc.be/qol/downloads/reference_values_manual 2008.pdf,accessed Jan 14, 2009) European Organization for Research and Treatment of Cancer—Quality of Life Department: Brussels, Belgium, 2008.
17 Bush NE, Haberman M, Donaldson G, Sullivan KM. Quality of life of 125 adults surviving 6–18 years after bone marrow transplantation. *Soc Sci Med* 1995; **40**: 479–490.
18 Molassiotis A, Morris PJ. Quality of life in patients with chronic myeloid leukemia after unrelated donor bone marrow transplantation. *Cancer Nurs* 1999; **22**: 340–349.
19 Zittoun R, Achard S, Ruszniewski M. Assessment of quality of life during intensive chemotherapy or bone marrow transplantation. *Psychooncology* 1999; **8**: 64–73.
20 Watson M, Buck G, Wheatley K, Homewood JR, Goldstone AH, Rees JKH *et al*. Adverse impact of bone marrow transplantation on quality of life in acute myeloid leukaemia patients; analysis of the UK Medical Research Council AML 10 Trial. *Eur J Cancer* 2004; **40**: 971–978.
21 Wettergren L, Langius A, Björkholm M, Björvell H. Physical and psychosocial functioning in patients undergoing autologous bone marrow transplantation—a prospective study. *Bone Marrow Transplant* 1997; **20**: 497–502.
22 Olivieri A, Capelli D, Montanari M, Brunori M, Massidda D, Poloni A *et al*. Very low toxicity and good quality of life in 48 elderly patients autotransplanted for hematological malignancies: a single center experience. *Bone Marrow Transplant* 2001; **27**: 1189–1195.
23 Gulbrandsen N, Hjermstad MJ, Wisløff F. Interpretation of quality of life scores in multiple myeloma by comparison with a reference population and assessment of the clinical importance of score differences. *Eur J Haematol* 2004; **72**: 172–180.
24 Poppelreuter M, Weis J, Mumm A, Orth HB, Bartsch HH. Rehabilitation of therapy-related cognitive deficits in patients after hematopoietic stem cell transplantation. *Bone Marrow Transplant* 2008; **41**: 79–90.
25 Harder H, Duivenvoorden HJ, van Gool AR, Cornelissen JJ, van den Bent MJ. Neurocognitive functions and quality of life in haematological patients receiving haematopoietic stem cell grafts: a one-year follow-up pilot study. *J Clin Exp Neuropsychol* 2006; **28**: 283–293.
26 Hjermstad MJ, Evensen SA, Kvaloy SO, Fayers PM, Kaasa S. Health-related quality of life 1 year after allogeneic or autologous stem-cell transplantation: a prospective study. *J Clin Oncol* 1999; **17**: 706–718.
27 Hjermstad MJ, Knobel H, Brinch L, Fayers PM, Loge JH, Holte H *et al*. A prospective study of health-related quality of life, fatigue, anxiety and depression 3–5 years after stem cell transplantation. *Bone Marrow Transplant* 2004; **34**: 257–266.
28 Hacker ED, Ferrans CE. Quality of life immediately after peripheral blood stem cell transplantation. *Cancer Nurs* 2003; **26**: 312–322.
29 Kamm M. *Veränderungen von Indikatoren der Lebensqualität' nach Hochdosis-Chemotherapie* (Dissertation; accessible online: http://webdoc.sub.gwdg.de/diss/2004/kamm/kamm.pdf) Universität Göttingen: Göttingen, 2004.

30 Frick E, Borasio GD, Zehentner H, Fischer N, Bumeder I. Individual quality of life of patients undergoing autologous peripheral blood stem cell transplantation. *Psychooncology* 2004; **13**: 116–124.
31 Frick E, Rieg-Appleson C, Tyroller M, Bumeder I. Social support, affectivity, and the quality of life of patients and their support-givers prior to stem cell transplantation. *J Psychosoc Oncol* 2005; **23**: 15–34.
32 Baumann FT. *Auswirkungen von Bewegungstherapie auf die physische und psychische Verfassung von Patienten im Rahmen einer hämatopoetischen Stammzelltransplantation (Dissertation)*. Institut für Rehabilitation und Behindertensport der Deutschen Sporthochschule Köln: Köln, 2005.
33 Frick E, Ramm G, Bumeder I, Schulz-Kindermann F, Tyroller M, Fischer N *et al*. Social support and quality of life of patients prior to stem cell or bone marrow transplantation. *Br J Health Psychol* 2006; **11**: 451–462.
34 Hacker ED, Ferrans C, Verlen E, Ravandi F, van Besien K, Gelms J *et al*. Fatigue and physical activity in patients undergoing hematopoietic stem cell transplant. *Oncol Nurs Forum* 2006; **33**: 614–624.
35 Schulz-Kindermann F, Mehnert A, Scherwath A, Schirmer L, Schleimer B, Zander AR *et al*. Cognitive function in the acute course of allogeneic hematopoietic stem cell transplantation for hematological malignancies. *Bone Marrow Transplant* 2007; **39**: 789–799.
36 Grulke N. *Psychosoziale Prädiktoren. Eine Studie zur Überlebenszeit von Patienten unter allogener hämatopoietischer Stammzelltransplantation*. Ulmer Textbank: Ulm, 2007.
37 Jarden M, Baadsgaard MT, Hovgaard DJ, Boesen E, Adamsen L. A randomized trial on the effect of a multimodal intervention on physical capacity, functional performance and quality of life in adult patients undergoing allogeneic SCT. *Bone Marrow Transplant* 2009; **43**: 725–737.
38 Rischer J, Scherwath A, Zander AR, Koch U, Schulz-Kindermann F. Sleep disturbances and emotional distress in the acute course of hematopoietic stem cell transplantation. *Bone Marrow Transplant* 2009; **44**: 121–128.
39 Wettergren L, Sprangers M, Björkholm M, Langius-Eklöf A. Quality of life before and one year following stem cell transplantation using an individualized and a standardized instrument. *Psychooncology* 2008; **17**: 338–346.
40 Brandberg Y, Michelson H, Nilsson B, Bolund C, Erikstein B, Hietanen P *et al*. Quality of life in women with breast cancer during the first year after random assignment to adjuvant treatment with marrow-supported high-dose chemotherapy with cyclophosphamide, thiotepa, and carboplatin or tailored therapy with fluorouracil, epirubicin, and cyclophosphamide: Scandinavian Breast Group Study 9401. *J Clin Oncol* 2003; **21**: 3659–3664.
41 Epstein JB, Phillips N, Parry J, Epstein MS, Nevill T, Stevenson-Moore P. Quality of life, taste, olfactory and oral function following high-dose chemotherapy and allogeneic hematopoietic cell transplantation. *Bone Marrow Transplant* 2002; **30**: 785–792.
42 Kopp M, Schweigkofler H, Holzner B, Nachbaur D, Niederwieser D, Fleischhacker WW *et al*. Time after bone marrow transplantation as an important variable for quality of life: results of a cross-sectional investigation using two different instruments for quality-of-life assessment. *Ann Hematol* 1998; **77**: 27–32.
43 Bush NE, Donaldson GW, Haberman MH, Dacanay R, Sullivan KM. Conditional and unconditional estimation of multidimensional quality of life after hematopoietic stem cell transplantation: a longitudinal follow-up of 415 patients. *Biol Blood Marrow Transplant* 2000; **6**: 576–591.
44 Uyl-de Groot CA, Buijt I, Gloudemans IJM, Ossenkoppele GJ, Berg HP, Huijgens PC. Health related quality of life in patients with multiple myeloma undergoing a double transplantation. *Eur J Haematol* 2005; **74**: 136–143.
45 Caocci G, Pisu S, Argiolu F, Giardini C, Locatelli F, Vacca A *et al*. Decision-making in adult thalassemia patients undergoing unrelated bone marrow transplantation: quality of life, communication and ethical issues. *Bone Marrow Transplant* 2006; **37**: 165–169.
46 Gulbrandsen N, Wisløff F, Nord E, Lenhoff S, Hjorth M, Westin J *et al*. Cost–utility analysis of high-dose melphalan with autologous blood stem cell support vs melphalan plus prednisone in patients younger than 60 years with multiple myeloma. *Eur J Haematol* 2001; **66**: 328–336.
47 Hensel M, Egerer G, Schneeweiss A, Goldschmidt H, Ho AD. Quality of life and rehabilitation in social and professional life after autologous stem cell transplantation. *Ann Oncol* 2002; **13**: 209–217.
48 Harder H, Cornelissen JJ, Van-Gool AR, Duivenvoorden HJ, Eijkenboom WMH, Van den Bent MJ. Cognitive functioning and quality of life in long-term adult survivors of bone marrow transplantation. *Cancer* 2002; **95**: 183–192.
49 Zittoun R, Suciu S, Watson M, Solbu G, Muus P, Mandelli F *et al*. Quality of life in patients with acute myelogenous leukemia in prolonged first complete remission after bone marrow transplantation (allogeneic or autologous) or chemotherapy: a cross-sectional study of the EORTC-GIMEMA AML 8A trial. *Bone Marrow Transplant* 1997; **20**: 307–315.
50 Mehnert A, Scherwath A, Schirmer L, Schleimer B, Petersen C, Schulz-Kindermann F *et al*. The association between neuropsychological impairment, self-perceived cognitive deficits, fatigue and health related quality of life in breast cancer survivors following standard adjuvant versus high-dose chemotherapy. *Patient Educ Couns* 2007; **66**: 108–118.
51 Hsu C, Wang J-D, Hwang J-S, Tien H-F, Chang S-M, Cheng A-L *et al*. Survival-weighted health profile for long-term survivors of acute myelogenous leukemia. *Qual Life Res* 2003; **12**: 503–517.
52 Worel N, Biener D, Kalhs P, Mitterbauer M, Keil F, Schulenburg A *et al*. Long-term outcome and quality of life of patients who are alive and in complete remission more than two years after allogeneic and syngeneic stem cell transplantation. *Bone Marrow Transplant* 2002; **30**: 619–626.
53 Messerer D, Engel J, Hasford J, Schaich M, Ehninger G, Sauerland C *et al*. Impact of different post-remission strategies on quality of life in patients with acute myeloid leukemia. *Haematologica* 2008; **93**: 826–833.
54 Hayden PJ, Keogh F, Conghaile MN, Carroll M, Crowley M, Fitzsimon N *et al*. A single-centre assessment of long-term quality-of-life status after sibling allogeneic stem cell transplantation for chronic myeloid leukemia in first chronic phase. *Bone Marrow Transplant* 2004; **34**: 545–556.
55 Gielissen MFM, Schattenberg AVM, Verhagen CAHHVM, Rinkes MJ, Bremmers MEJ, Bleijenberg G. Experience of severe fatigue in long-term survivors of stem cell transplantation. *Bone Marrow Transplant* 2007; **39**: 595–603.
56 Kopp M, Holzner B, Meraner V, Sperner-Unterweger B, Kemmler G, Nguyen-Van-Tam DP *et al*. Quality of life in adult hematopoietic cell transplant patients at least 5 yr after treatment: a comparison with healthy controls. *Eur J Haematol* 2005; **74**: 304–308.
57 Loberiza FR, Rizzo JD, Bredeson CN, Antin JH, Horowitz MM, Weeks JC *et al*. Associations of depressive syndrome and early deaths among patients after stem-cell transplantation for malignant diseases. *J Clin Oncol* 2002; **20**: 2118–2126.
58 Grulke N, Larbig W, Kächele H, Bailer H. Pre-transplant depression as risk factor for survival of patients undergoing

allogeneic haematopoietic stem cell transplantation. *Psychooncology* 2008; **17**: 480–487.
59 van Belle G. *Statistical Rules of Thumb*. Wiley: New York, 2002.
60 Baker KS, Gurney JG, Ness KK, Bhatia R, Forman SJ, Francisco L *et al*. Late effects in survivors of chronic myeloid leukemia treated with hematopoietic cell transplantation: results from the Bone Marrow Transplant Survivor Study. *Blood* 2004; **104**: 1898–1906.
61 Syrjala KL, Langer SL, Abrams JR, Storer B, Sanders JE, Flowers MED *et al*. Recovery and long-term function after hematopoietic cell transplantation for leukemia or lymphoma. *J Am Med Assoc* 2004; **291**: 2335–2343.
62 Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol* 1998; **16**: 139–144.
63 King MT. The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. *Qual Life Res* 1996; **5**: 555–567.
64 Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2 edn. Lawrence Erlbaum Associates: Hillsdale, New Jersey, 1988.