Pediatric acute lymphoblastic leukemia: Quality of life and cost-effectiveness of treatment

van Litsenburg, R.R.L.

2012

document version
Publisher's PDF, also known as Version of record

Chapter 7

Health status utilities in pediatric acute lymphoblastic leukemia: a systematic review

Raphaële R.L. van Litsenburg*, Annemieke Kunst*, Jaap Huisman, Johannes C.F. Ket, Gertjan J.L. Kaspers, Reinoud J.B.J. Gemke
* Both authors contributed equally

Submitted

Abstract

Background: Significant progress in the treatment of pediatric acute lymphoblastic leukemia (ALL) has increased the survival rates to 80-85%. This has led to a stronger emphasis on health status and cost-effectiveness of treatment.

Purpose: This systematic review summarized the available evidence on utility scores and aimed to differentiate between utility scores in children on treatment and in survivors, scores associated with specific treatment modalities and scores gathered from different respondents in order to facilitate future cost-utility analyses.

Data Sources: Searches of data sources Pubmed, Embase, Wiley/Cochrane Library, EBSCO/Cinahl and CSA/PsycInfo were performed from inception to January 1st 2011. Studies had to report on utility scores in pediatric ALL, either on or after treatment, to be included.

Data synthesis: Thirteen studies were included. All studies used the Health Utility Index (HUI) Mark 2 or 3. The majority of studies had methodological shortcomings, which mainly concerned study design, and definition and representativeness of the study group. Utility scores seem to be dependent on treatment variables and generally there is an improvement in HRQL as treatment or suvivorship advances. In general, proxy respondents are less reliable for subjective phenomena than for observable conditions. HUI2 and HUI3 scores are not interchangable.

Limitations: All reviewed studies had methodological shortcomings, which can lead to a biased presentation of the results. Sample sizes were often small.

Conclusions: The available evidence for utility scores in pediatric ALL is sparse and methodological suboptimal. Use of these utilities in economic evaluations would warrant extensive sensitivity analyses. Therefore, new longitudinal studies are neseccary to adequately assess utility scores in pediatric ALL patients during and after treatment.

Introduction

Over the last decades there has been significant progress in the treatment of pediatric acute lymphoblastic leukemia (ALL), the most common type of pediatric malignancies. Survival rates have increased to 80-90% [2, 63]. This success has led to a stronger emphasis on other aspects of treatment, such as health status or health related quality of life (HRQL) and cost-effectiveness of treatment. The attention to the cost-effectiveness of pediatric (oncology) interventions is rising [153-155]. It is likely that economic evaluations will become even more necessary in the future as expensive and time-consuming health-care technology evolves while the strain on healthcare budgets increases.

Most guidelines on economic evaluations recommend cost-utility analyses, i.e. the calculation of costs per quality adjusted life years (QALY) to account for the effect of quality on the number of life years saved [59]. Preference-based HRQL scores are based on community preferences concerning health states or HRQL. The more preferable an outcome, the more utility is associated with it. Utility scores are used to calculate QALY, and can be essential in determining which alternative is most cost-effective. Previous research has shown that HRQL, and thus utility scores, is influenced by several variables. Survivors enjoy a better HRQL than children during treatment [90, 92], although on certain domains HRQL in survivors is often still impaired compared to the norm [25]. Treatment modalities such as irradiation and glucocorticoid treatment have been associated with a lower HRQL [77, 92, 102]. Finally, an important aspect of HRQL research in pediatric oncology concerns the choice of respondent. During treatment, children are often too young or too ill to act as respondent. Parents, nurses and physicians can serve as proxy-respodents. In general, proxy-respondents are likely to be reliable for readily observable conditions or events, but less reliable for subjective phenomena. There is a tendency for proxy respondents who are not highly familiar with a patient to underestimate health problems [66, 156, 157].

Since utility scores are necessary for the calculation of QALY but systematic longitudinal assessments of utility scores are time-consuming and therefore often not readily available, we performed a systematic review of utility scores associated with pediatric ALL. The aim was to summarize the available evidence on utility scores and to differentiate between utility scores in children on treatment and in survivors, scores associated with specific treatment modalities and scores gathered from different respondents in order to facilitate future cost-utility analyses without the need for time-consuming assessments.

Methods

Search strategy

This study is part of a comprehensive review of HRQL in children with ALL, the results on non-preference based HRQL scores will be presented elsewhere. The following databases were searched by RvL and JCFK for relevant studies from inception to January 1st 2011: PubMed, Embase, Wiley/Cochrane Library, EBSCO/Cinahl and CSA/PsycInfo. For this literature search the following terms were used: "acute lymphoblastic leukemia", "quality of life", "child", "survivors". The complete search strategy is presented in the appendix. Reference lists of included studies and relevant reviews were screened for missed articles. All abstracts identified by the literature search were independently screened by two reviewers (RL screened all identified abstracts and JH/GK/RG each a third of all abstracts).

Inclusion criteria

To be included in this systematic review, studies had to report on utility scores in children diagnosed with ALL before the age of 19 years, either on or after treatment. All research designs were considered. Full texts had to be available, non-peer reviewed dissertations were excluded. Questionnaire scores had to be presented separately for pediatric ALL patients. Languages were restricted to English, French and Dutch.

Data extraction and assessment of methodological quality

Three independent reviewers (RL and AK each all included studies, and JH/GK/RG each a third of all included studies) abstracted information from each included study using standardized data collection forms. The methodological quality of each paper was also assessed independently by three reviewers, according to criteria based on the papers by Tsimicalis et al.[70] and Eiser et al.[71], see Table 7.1.

Results

Study characteristics

The larger search yielded 2609 hits, the details of the study selection process is presented in figure 7.1. Thirteen studies provided utility scores and were included in this systematic review. All included studies used the Health Utilities Index (HUI). The HUI Mark 2 (HUI2) consists of seven attributes (sensation, mobility, emotion, cognition, self-care, pain and fertility). The HUI Mark 3 (HUI3) is comprised of eight attributes (vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain). Multi-Attribute Utility (MAU) scores can be determined using published utility functions, in which scores

Table 7.1 Criteria used to define the methodological qualities of the selected studies and a summary of the assessment of the methodological qualities

Parameter	Criteria	Points
1. Study design	1) Longitudinal prospective 2) Retrospective or mixed design 3) Cross-sectional 4) Survey or not explicitely stated	- longitudinal prospective (3) - retrospective or mixed (2) - cross-sectional (1) - survey/not explicitely stated (0)
2. Definition of study group	1) Age at diagnosis or measurement 2) Gender 3) Moment of measurement (during treatment, survivors) 4) Type of intervention described 5) In- and exclusion criteria are adequately described	- all 5 criteria described (3) - 4 criteria described (2) - 3 criteria described (1) - > 2 criteria missing (0)
3. Study group is representive	1) The study sample is a random sample of the source population (>70% of eligible patients) 2) Response rate is adequate (>80%) 3) Attempts to collect information on participants who dropped out of the study are described 4) Reasons for loss to follow up are provided	- all 4 criteria described (3) - 3 criteria described (2) - 2 criteria described (1) - > 2 criteria missing (0)
4. Definition of outcomes	1) The questionnaires are adequately valid and reliable 2) Respondents are described including reasons for excluding self-reports 3) There is a well-matched control group or results are compared to measurement norms 4) Study results are represented seperately for ALL, including mean and SD scores	- all 4 criteria described (3) - 3 criteria described (2) - 2 criteria described (1) - > 2 criteria missing (0)

Author	Year of publication	Study design	Definition of study group	Study group is representive	Definition of outcomes
Feeny [167]	1992	0	0	0	1
Feeny [168]	1993	0	3	0	2
Barr [166]	1993	0	0	0	2
Barr [106]	1997	0	3	3	3
Barr [164]	2001	1	0	0	2
Wright [165]	2003	0	3	3	3
Wright [119]	2005	1	3	3	3
Cox [169]	2005	0	2	2	2
Shimoda [161]	2005	0	2	0	3
Fu [163]	2006	0	0	0	3
Hinds [170]	2007	3	2	1	2
Fluchel [156]	2008	0	0	2	3
Shimoda [162]	2008	0	0	2	3

of 0.00 represent being dead and 1.00 living in perfect health. Mean MAU differences greater than 0.03 are considered clinically important. There are two major formats available, one is designed for self-completion (self-assessment or proxy-assessment), one is designed for interviewer administration. The HUI has been proven to be valid, reliable and responsive to change.[158-160]

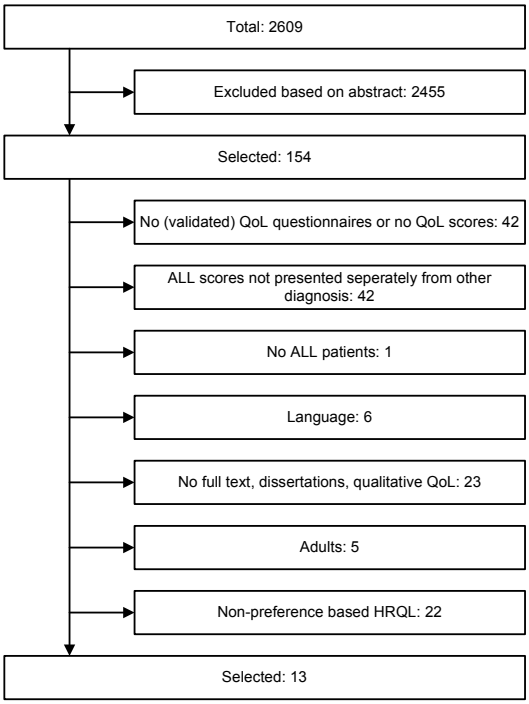

Figure 7.1 The literature search process

An overview of the study characteristics is provided in table 7.2. Seven studies employed both the HUI2 and the HUI3 [106, 119, 161-165], although only four studies presented both utility scores [161, 163-165]. The majority of studies did not report the study design (77%). Most studies were considered to be cross-sectional (62%), although this was explicitly stated in only two studies [119, 156, 161-166]. Eight studies (62%) were performed in Northern America or Western Europe [106, 119, 165-170], 38% in Central or Latin America [156, 161-164]. Most studies (69%) involved survivors [119, 156, 161-166, 168], three (23%) included children during treatment [106, 169, 170] and one included both [167]. In patients during treatment only proxy-reports were collected. Four survivor studies collected both proxy and self reports [156, 161, 163, 165], four used proxy-reports only [119, 164, 166, 168] and one study collected self-reports only [162]. The number of included ALL patients ranged between 8 and 106. The age of participants ranged between 0.9 and 40.2 years at the time of enrollment, three studies did not report the age of their participants. Four studies (31%) did not specify the treatment modality [162, 164, 167, 169], six (46%) included irradiated patients in the cohort [119, 156, 161, 165, 166, 168].

The majority of studies had the intention to describe the health status and HRQL of children during treatment for ALL or of survivors of childhood cancer. Some of the studies were specifically interested in the differences among types of assessors [106,

Table 7.2 Details of the included studies

Author, Year	Country	Data collection	Study design*	N (ALL)	Age at enrollment Mean ± SD (years)	Risk group	Treatment	Phase of care	HUI§	Respondent‖	Controls
Feeny [167] 1992	Canada	Not reported	Not reported (Prospective)	15	Not reported	High risk	Not reported	On treatment (n=11) Survivor <1 year off therapy (n=4)	2‡	Physician	Other childhood cancer
Feeny [168] 1993	England	Not reported	Not reported (Retrospective)	69	- (range: 8 – 25)	High risk	Chemotherapy Irradiation	Survivors (mean : 9.33 years since diagnosis)	2‡	Physician (chart review)	National population based sample
Barr [166] 1993	Canada	Not reported	Not reported (Cross sectional)	65	Not reported	Standard risk High risk	Chemotherapy Irradiation (n=55)	Survivors (< 15 years and >15-20 years off therapy)	2‡	Physician and Nurse consortium	Population norm
Barr [106] 1997	Canada	Not reported	Not reported (Prospective longitudinal)	18	Median: 3.9 (range: 0.9 – 14)	Standard risk High risk	Chemotherapy Corticosteroids	Maintenance	2 (3)	Nurse Physician Parent	Patients act as own controls
Barr [164] 2001	Cuba Honduras Colombia Uruguay	2000	Cross sectional	95	Not reported	Not reported	Not reported	Survivors (> 2 years off therapy)	2 3	Parent (Physician)	Other childhood cancer and equivalent survivor group in Canada
Wright [165] 2003	Canada	Not reported	Not reported (Cross sectional)	62	12.1 ± 3.6	Standard risk High risk	Irradiation	Survivors (mean: 5.4 ± 3.5 years off therapy)	2 3	Parent and self combination	Healthy subjects
Wright [119] 2005	Canada	3 year period	Cross sectional	77	12.1 ± 4.9	Standard risk High risk	Chemotherapy Corticosteroids Irradiation (some)	Survivors (mean: 4.7 ± 3.5 years off therapy)	(2) 3	Parent	Healthy subjects

Table 7.2 Details of the included studies (continued)

Author, Year	Country	Data collection	Study design*	N (ALL)	Age at enrollment Mean ± SD (years)	Risk group	Treatment	Phase of care	HUI§	Respondent‖	Controls
Cox [169] 2005	USA	Not reported	Not reported (Prospective longitudinal)	27	Median: 10 (range: 6.0 – 18.0)	Not reported	Not reported	Induction Continuation	3	Nurse	Patients act as own controls
Shimoda [161] 2005	Brazil	2002	Not reported (Cross sectional)	8	23.4 ± 5.51†	Not reported	Chemotherapy Irradiation	Survivors (not specified)	2 3	Self (Physician) Nurse	ALL survivors in simular studies
Fu [163] 2006	El Salvador Honduras Panama Nicaragua	2001	Not reported (Cross sectional)	91	Median: 12.8† (range: 3.4-25.8)	Not reported	Chemotherapy	Survivors (>2 years off therapy)	2 3	Self (Parent) (Physician)	Other childhood cancer
Hinds [170] 2007	USA	Not reported	Prospective	106	Median: 8.6 (range: 4.9 – 18.8)	Not reported	Chemotherapy	Induction Maintenance	3	Nurse	Three time points during treatment
Fluchel [156] 2008	Uruguay	1998-1999	Not reported (Cross sectional)	49	13.6 ± 4.26†	Not reported	Chemotherapy Irradiation	Survivors (>2 years off therapy)	3	Self Parent Physician Teacher	Healthy subjects
Shimoda [162] 2008	Brazil	2002-2003	Not reported (Cross sectional)	31	22.8† (range:13.4-40.2)	Not reported	Not reported	Survivors (>8 years off therapy)	(2) 3	Self	Other childhood cancer and ALL survivors in Central America

* Study designs between parenthesis are the authors' interpretations of the study designs where this is not literally reported in the manuscript. † Age refers to the total study-group, including other childhood cancers. ‡ Mutli-attribute scores were computed based on the reported attribute levels. On the fertility attribute level 1 morbidity was assumed (i.e. no morbidity) since no scores were reported. § Questionnaires between parenthesis indicate that ALL utility scores were not reported for the specified questionnaire. ‖ Respondents between parenthesis indicate that ALL utility scores were not reported for the specified responder.

156, 161, 163, 164]. The studies performed in Central and Latin America also aimed at determing the feasibility, reliability and validity of the translation [156, 161, 163, 164].

Methodological qualities of the studies

A summary of the assessment of the methodological qualities is shown in table 7.1. None of the thirteen studies were judged to have a perfect score on all four criteria. Methodological shortcomings mainly concerned the study design and also the definition and representativeness of the study group. Study design was explicitly stated in only three studies [119, 164, 170]. Study group characteristics were complete in four studies [106, 119, 165, 168], studies that did not report characteristics seperately for the ALL population did not receive the maximum amount of points [156, 161-163]. The study group was found to be representative (i.e. three points) in three studies [106, 119, 165]. The quality of the included studies was better for the definition of outcomes. Seven studies met all criteria [106, 119, 156, 161-163, 165] and none received 0 points. Two studies did not report MAU scores [166-168]. For the purpose of this review, lacking MAU scores were computed based on the reported attribute levels. On the fertility attribute level 1 morbidity was assumed (i.e. no morbidity) since no scores were reported in these two studies.

Utility scores during treatment

Four studies investigated children on active treatment, two with use of the HUI2 [106, 167] and two using the HUI3 [169, 170], Table 7.3. All studies employed proxy-respondents. Feeny et al.[167] included both children on treatment and survivors. The calculated mean MAU score for 11 patients on treatment was 0.78. No information on treatment variables was available and it was unclear during which period the patients were assessed. It needs to be noted that this study took place over 20 years ago and it is likely that treatment would have included irradiation in at least a part of the patients, which is known to be associated with a lower HRQL [102, 171]. Two more recent longitudinal prospective studies included children at three different time points during the induction and continuation phase [169, 170]. HUI3-proxy assessors were nurses in both studies. A decline in MAU scores with advancement of therapy was reported, although a lack of variation in scores was found in both studies, indicating according to the authors that the HUI3 might not be sensitive enough to detect (small) shifts in HRQL during treatment. Due to a large number of missing values, MAU scores could only be calculated for a part of the participants (n=27[169], n=106[170]). MAU scores varied between 0.90 and 0.94 during induction therapy and 0.87-0.92 during consolodation/ continuation therapy. Barr et al.[106] investigated the influence of glucocorticosteroid therapy on HUI2 scores in 18 patients. Proxy respondents were a nurse, a physician and a parent. The assessments were made at three different time points: at the start of a

Table 7.3 Health Utilities Index Multi Attribute Utility scores during treatment

Study (year)	HUI	Phase of care	MAU score Mean (SD)			
			Parent	Nurse	Physician	Self
Cox (2005) [169]	3	Week 6	-	0.90 (0.03) †	-	-
		Week 7	-	0.91 (0.06) †	-	-
		Week 31	-	0.87 (0.06) †	-	-
Hinds (2007) [170]	3	Week 6	-	0.94 (0.01)	-	-
		Week 7	-	0.92 (0.03)	-	-
		Week 48	-	0.91 (0.03)	-	-
Barr (1997) [106]	2	Start of 5-day GCS cycle: day 1	0.86 (0.17)	0.96 (0.05)	0.90 (0.10)	-
		Shortly after GCS cycle: day 7	0.83 (0.21)	0.86 (0.12)	0.83 (0.20)	-
		After GCS cycle: day 14	0.89 (0.09)	0.91 (0.13)	0.89 (0.12)	-
Feeny (1992)* [167]	2	Not specified	-	-	0.78 (0.28)	-

GCS=glucocorticosteroid therapy. * MAU score calculated based on the reported attribute levels. Fertility was assumed normal (i.e. level 1) in patients with an unknown status. † standard error scores.

five day glucocorticorticoid cycle, on day 7 and day 14. The lowest utility scores were found on day 7, shortly after glucocorticoid therapy. Nurses scored the highest MAU-scores for their patients (0.86-0.96), followed by physicians (0.83-0.90) and parents (0.83-0.89). No significant differences in scores were found between standard risk and high risk children, the latter group receiving a threefold corticosteroid dose. According to the authors this was possibly because of a lack of power or because the steroid dose was already high in the standard risk group.

Utility scores in survivors

Ten studies evaluated utility scores in ALL survivors [119, 156, 161-168], three reported HUI2 scores [166-168], three studies reported HUI3 scores [119, 156, 162] and four reported both [161, 163-165], summarized in Table 7.4. Proxy-respondents were employed in seven studies [119, 161, 164-168] and self-reports in four [156, 161-163]. Feeny et al.[167] report on physician-proxy assessments in during the first year off treatment. No information on treatment variables was available, the (calculated) mean MAU score for four ALL survivors was 0.71. Another study from the same author reported physician-proxy scores in longer term survivors (mean 9.3 years), although it has to be noted that these scores were constructed through chart-review [168]. MAU score for these irradiated patients (n=69) was 0.83. Barr et al.[166] reported physician-nurse consortium HUI2 assessments in a cohort of largely irradiated patients (55/65). MAU score was 0.90 for standard risk patients and 0.78 for high risk patients.

Parent-proxy assessments were reported in three studies. Barr et al.[164] reported MAU scores of 0.93 (HUI2) and 0.92 (HUI3) in a cohort of 100 ALL patients that were at least two years off therapy. No information on treatment variables was available.

Table 7.4 Health Utilities Index Multi Attribute Utility scores in survivors

Study (year)	Phase of care	HUI	MAU score mean (SD)			
			Parent	Nurse	Physician	Self
Feeny (1992)* [167]	<1 yr off therapy	2	-	-	0.71 (0.30)	-
Feeny (1993)* [168]	Mean 9.3 yrs off therapy	2	-	-	0.83 (0.24)	-
Barr (1993)* [166]	Not specified (Standard risk ALL)	2	-	-	0.90 (0.16)	-
	Not specifief (-Very- High risk ALL)		-	-	0.78 (0.26)	-
Barr (2001) [164]	>2 yrs off therapy	2	0.93 (0.10)	-	-	-
		3	0.92 (0.13)	-	-	-
Wright (2005) [119]	>4.7 yrs off therapy	3	0.86 (0.18)	-	-	-
Wright (2003) [165]	mean 5.4 yrs off therapy	2	0.91 (0.084)	-	-	-
		3	0.86 (0.166)	-	-	-
Shimoda (2005) [161]	Not specified	2	-	0.85 (0.16)	-	0.88 (0.19)
		3	-	0.85 (0.17)	-	0.85 (0.25)
Fu (2006) [163]	>2 yrs off therapy	2	-	-	-	0.871 (0.18)
		3	-	-	-	0.725 (0.30)
Fluchel (2008) [156]	>2 yrs off therapy	3	-	-	-	0.722 (0.25)
Shimoda (2008) [162]	>8 yrs off therapy	3	-	-	-	0.83 (0.225)

* MAU score calculated based on the reported attribute levels. Fertility was assumed normal (i.e. level 1) in patients with an unknown status.

Wright et al.[165] studied a cohort including irradiated survivors (49/62) with a mean time off treatment of 5.4 years. Results from parent-proxy (n=48) and self-assessments (n=14) were combined, MAU scores were 0.94 (HUI2) and 0.86 (HUI3). Another cohort reported by Wright et al.[119] also included some irradiated patients (number not specified). The parent-reported HUI3 MAU score of 77 ALL patients was similar to the previous study (0.86).

Shimoda et al.[161] reported on nurse and self-reports of eight Brazillian ALL patients, seven of whom were irradiated. Time since diagnosis was not stated. Nurse-assessed utility scores were 0.88 (HUI2) and 0.85 (HUI3), self-assessed scores were 0.85 on both questionnaires. The other three studies including self-reports were also performed in Central or South America [156, 162, 163]. Fu et al.[163] included chemotherapy-only treated survivors that were off therapy for at least two years. MAU scores were 0.87 (HUI2) and 0.73 (HUI3). Fluchel et al.[156] included survivors at least two years off therapy. It was unclear how many ALL patients had been irradiated, the mean MAU score was 0.72 (HUI3). A second study by Shimoda et al.[162] included survivors who were more than eight years off therapy. The mean MAU score was 0.83 (HUI3). No information on treatment variables was avialable.

Differences between HUI2 and HUI3

Four studies reported both HUI2 and HUI3 results [161, 163-165]. In all but one of the assessments the HUI2 MAU-scores were higher compared to HUI3 MAU-scores, with differences ranging from to 0.01 to 0.15. Nurse proxy assessments as found by Shimoda et al.[161] were simular for both HUI2 and HUI3.

Agreement between respondents

In seven studies agreement between the different respondents was investigated. No significant effect of the type of respondent was found in three studies for both HUI2 and HUI3 scores [106, 161, 165]. Feeny et al.[167] reported 61% agreement between six clinicians. In the study by Barr et al.[164] significant positive correlations between parents and physicians for both HUI2 (r=0.33) and HUI3 (r=0.61) were found, although parental HUI2 scores were significantly lower than physician scores (HUI2 mean score difference 0.04, p=<0.0005 and HUI3 mean score difference 0.02, p=0.04). Fu et al.[163] reported important size differences in mean HRQL scores between patients, parents and physicians, in particular for the more subjective attributes emotion and pain. The largest differences were found between patients and physicians (HUI2 mean score difference 0.07, p<0.001 and HUI3 mean score difference 0.16, p<0.001). Self-reports scored the highest level of morbidity in the emotional functioning, while the parent-proxies scored the highest level of morbidity in pain. In the study by Fluchel et al.[156] there was a high inter-rater agreement between the different types of assessors for readily observable attributes and less agreement for attributes not directly observable. Agreement was the largest between survivors and parents (89%) as compared to survivor and physician (56%), i.e. the agreement was higher when the assessor knew the subject well.

Discussion

This systematic review aimed at summarizing the available evidence on utility scores in pediatric ALL, to serve as a reference for future cost-utility analyses. Up to January 1st 2011, studies reporting utility scores in pediatric ALL have all employed the HUI. To the best of our knowledge, no studies have been published that reported ALL scores using other preference based instruments. Large differences in HUI scores between the included studies were found. During treatment scores varied between 0.78-0.96 and in survivors between 0.71-0.93. Considering that score differences of 0.03 or greater can be considered clinically important, the variation in scores is remarkable. The variation in scores can be explained by several mechanisms, such as the choice of respondent, patient and treatment variables and differences in the HUI2 and the HUI3.

In pediatrics patients are often too young or too ill to participate in HRQL research and therefore proxy-respondents are employed. It is well known that children, parents and physicians do not always agree on HRQL [36, 37, 157, 172, 173], as was illustrated by several of the included studies [106, 161]. In general, proxy-respondents are likely to be reliable for observable conditions or events, but less reliable for subjective phenomena [157]. There also seems to be a tendency for proxy-respondents who are not highly familiar with a patient, such as healthcare professionals, to underestimate health problems [157] or to be unable to adequately complete the HRQL questionnaire [174]. This was illustrated by two nurse-proxy respondent studies in this review. The HUI was completed by nurses independent of the patients' or parental imput, which lead to a high rate of missing values [169, 170]. On the other hand, parental and self reports can also be influenced. In parents, elevated levels of psychological distress have been reported [101], which can affect HRQL assessment. In cancer patients, a phenomenon called response shift has been described, which means that a patient's perception of QoL can improve even when functional abilities or health status do not, due to the adaptation to the disease process [38]. In general, self-reports are deemed most informative [34, 66, 67] but both views on HRQL are considered valuable [172].

Utility scores are dependent on treatment variables. The negative effect of glucocorticosteroids was clear in the study by Barr et al.[106] and has been reported by other authors as well [77, 92]. Irradiation has also been associated with a lower HRQL[102, 171]. In the included studies, however, utility scores were usually not presented seperately for irradiated patients [119, 161, 165, 166] although Barr et al.[166] reported that children who were not irradiated had no emotional or cognitive deficits as opposed to those that were irradiated. Also, many studies did not explicitely state the treatment modalities. The study by Feeny et al.[168] is the only one that reports utility scores for a cohort of irradiated survivors. These scores were indeed low, although they were computed through chart review and the validity of this method is unclear. Finally, longitudinal HRQL studies have generally shown an improvement in HRQL as treatment or survivorship advances [77, 78, 85]. The longitudinal studies in this systematic review generally reported a decline in utility scores, although this may be related to methodological issues [169, 170]. Unfortunately, most included studies were cross-sectional. Comparing parent-proxy scores in patients during (0.83-0.89) and after (0.86-0.93) treatment, there seems to be trend of improvement similar to previous studies.

Finally, there appear to be systematic differences between HUI2 and HUI3 scores. A difference larger than the minimally clinical important difference of 0.03 [158] was found in three of the four studies that employed both the HUI2 and the HUI3 [161, 163-165]. This finding implies that utility scores collected with the HUI2 and the HUI3 scores are not interchangable.

The strength of this systematic review includes the extensive literature search, offering a comprehensive report on health status utilities in pediatric ALL. It systematically summarizes utility scores associated with ALL to facilitate future cost-utility analyses and it identifies issues that remain unsolved. However, some limitations need to be mentioned. Our search strategy cannot rule out the presence of language bias. All of the studies reviewed had methodological shortcomings, which can lead to a biased presentation of the results. For example, patient and treatment characteristics were often not adequately described while these characteristics can influence utility scores, as described in the previous section. Sample sizes were often small, reducing the robustness of the data. For the calculation of MAU scores in three studies [166-168], our assumption of a normal fertility in view of lacking data may have induced bias.

In conclusion, the available evidence for utility scores in pediatric ALL is sparse and methodologically suboptimal. The use of these existing utility scores in economic evaluations would warrant extensive sensitivity analyses. Therefore, new longitudinal studies are neseccary to adequately assess utility scores in pediatric ALL patients during and after treatment. These new studies should include large cohorts in order to have enough power to evaluate the effect of specific therapies. Ideally, self-reports should be included, as well as the complementary and sometime indespensible parental-reports. Finally, new studies should include a careful description of the study methods and the population studied in order to facilitate the interpretation of results and provide the possibilty for future meta-analysis.

Acknowledgements

The authors would like to thank dr. L. Kremer and dr. E. van Dalen for assisting in the methodological setup of this review.