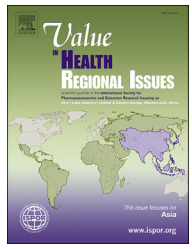
VALUE IN HEALTH REGIONAL ISSUES 18 (2019) 97–105

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/vhri



Methodology

Utility Measures in Pediatric Temporary Health States: Comparison of Prone Positioning Valuation Through 5 Assessment Tools

Shima Shahjouei, MD, MPH [1,2,3], *Alireza Vafaei Sadr, PhD* [4,5], *Soheila Khorasani, MSc* [2], *Farideh Nejat, MD, MPH* [1], *Zohreh Habibi, MD* [1], *Ali Akbari Sari, MD, PhD* [2,*]

[1]*Department of Neurosurgery, Children's Hospital Medical Centre, Tehran University of Medical Sciences, Tehran, Iran;* [2]*Department of Health Management and Economics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran;* [3]*Universal Scientific Education and Research Network, Tehran, Iran;* [4]*Department of Physics, Shahid Beheshti University, Tehran, Iran;* [5]*Department of Theoretical Physics and Center for Astroparticle Physics, University of Geneva, Geneva, Switzerland*

ABSTRACT

Background: Assessment of health-related quality of life (HRQOL or utility) is a complex issue especially in children with temporary health states. **Objectives:** To assess the utility of prone positioning as a prophylactic postsurgical approach with the aid of 5 frequently used general instruments. **Methods:** Visual analogue scale (VAS), time trade-off (TTO), modified TTO (m-TTO), standard gamble (SG), and chain of gambles (ChGs) instruments and interview with the parent caregivers were used to measure the HRQOL (utility value) of patients who were admitted in the surgical wards of Children's Medical Center Hospital between July and November 2015. **Results:** A total of 74 parent caregivers with a mean age of 30.48 ± 6.66 years were enrolled. On the basis of the Gaussian model of the repeated VAS measures, we classified the behavior of the participants into 4 clusters. Cumulative study of all these clusters demonstrated that TTO has the highest utility measure for prone positioning (0.682 ± 0.359), whereas the lowest utility value was measured by VAS2 (0.132 ± 0.569). In addition, all VAS measures underestimated the preferences. Overall, values of TTO, m-TTO, and ChGs remained consistent through each of these 4 clusters (intracluster consistency) and within each cluster (intercluster consistency). The adopted utility value of prone positioning based on these 3 instruments was estimated as 0.68 ± 0.21. **Conclusions:** We recommended a model for assessment of HRQOL in children with temporary health states to overcome the challenges of each isolated instrument and used this model to measure the utility value of prone positioning in pediatric patients.

Keywords: chain of gambles, modified time trade-off, prone positioning, standard gamble (SG), temporary health state, time trade-off (TTO), visual analogue scale (VAS)

Introduction

In a randomized clinical trial on pediatric patients who underwent untethering surgeries, we studied the effectiveness of 2 postoperative modalities of care, acetazolamide administration and prone positioning (positioning the child on his or her abdomen), in reducing operation-site complications[1] We demonstrated that keeping the patient in prone position for 10 days mitigates the rate of cerebrospinal fluid leak and collection at the operation site. Although this prophylactic approach is effective with no cost (cost-beneficial), it is unfavorable. Hence, we desired to assess the health-related quality of life (HRQOL) of prone positioning to conduct the cost-utility analysis of our randomized clinical trial.

HRQOL is the value of health state (utility) that can be measured by different methodological instruments. These measures are categorized either as direct/indirect methods or as generic/disease-specific instruments.[2,3] Indirect and questionnaire-based measures

This study was conducted as the second phase of an MD-MPH thesis, a shared project by the Department of Neurosurgery and the Department of Health Management and Economics, Tehran University of Medical Sciences.

Conflicts of interest: The authors have no conflicts of interest or financial relationships relevant to this article to disclose.

Source of financial support: This study was funded by Tehran University of Medical Sciences (ID: 2137).

* *Address correspondence to:* Ali Akbari Sari, MD, PhD, Department of Health Management and Economics, School of Public Health, Tehran University of Medical Sciences, Pour Sina St, Keshavarz Blvd, Tehran 14194, Iran.

Email: akbarisari@tums.ac.ir

2212-1099/$36.00 - see front matter © 2019 ISPOR–The professional society for health economics and outcomes research. Published by Elsevier Inc.

https://doi.org/10.1016/j.vhri.2019.01.003

estimate health state preference by the aid of implicit questions.[4–7] In contrast, in direct methods such as standard gamble (SG), time trade-off (TTO), and visual analogue scale (VAS), the desirability of the health condition is explicitly valuated by the individuals.[8,9] Most generic instruments are in the form of a standardized questionnaire.[10] The Childhood Health Assessment Questionnaire, Child Health Questionnaire, Pediatric Quality of Life Inventory Generic Core Module, parent/patient global assessment of well-being, and Disabkids Chronic Generic Measure are some of the generic questionnaires specifically designed for children.[6,11–14] Most of these instruments are multidimensional and evaluate difficulties in motor, emotional, social, and school/daycare domains. Nevertheless, they could provide the overall quality of life and not the specific utility measure for further analysis.

HRQOL assessment in pediatric neurological disorders is scarce. Some authors used an adopted version of pediatric generic instruments for this.[15] The National Institute of Neurological Disorders and Stroke suggested an HRQOL measurement specific for neurological conditions (Neuro-QOL).[16–18] The pediatric version of the Neuro-QOL includes both generic and targeted item banks. Stigma, depression, anxiety, anger, interaction with peers, general concerns, fatigue, and pain are the main domains of the Neuro-QOL Pediatric Short Form Domains version of this questionnaire.[19]

Choosing the most appropriate HRQOL instrument for measuring the utility of prone positioning was challenging. First, the strong emotional connection between the parents and the patients, high life expectancy of the child, and short duration of health condition in our study (only 10 days) make it difficult for the proxy to easily valuate the health state by tools trading the life-years (TTO) or gambling against death (SG). In addition, each of these tools have limitations and considerations (see Discussion section). Thus, a single direct utility instrument was not appropriate for our study. Second, none of the commonly used questionnaire-based instruments contain the domains of our interest. Third, in case the available questionnaires in the current literature evaluate similar characteristics of our interest, these measures have moderate correlation with quality of life.[12,20] Fourth, even those questionnaires with child self-report versions (eg, the Pediatric Quality of Life Inventory Generic Core Module) cannot be used in children younger than 5 years,[3,21] and proxies (parents, guardians, or healthcare providers) should be asked to consider the condition of the child.[22,23] This in turn limits the validity of the results, because the caregiver's status, beliefs, preferences, and convenience effect a change in the measured utility.[24–26] Accordingly, we designed this study to determine the best pediatric utility instrument in temporary health conditions adapted for caregivers. We also explored the utility value of prone positioning in pediatric patients for future cost-utility analysis.

Methods

This study was conducted at Children's Medical Center Hospital, Tehran, Iran, in 2015. By the aid of medical file numbers, we randomly selected admitted pediatric patients who needed surgical intervention and invited their parent caregivers (as a proxy) to this study. Participants were free to quit the interview anytime they desired. Patients older than 12 years, poor in Persian or English, and with emergent or urgent medical condition threatening them at the time of interview were excluded. This study was registered in the ClinicalTrials.gov database (registration ID: NCT01867268) and in the Iranian Registry of Clinical Trials (No.: IRCT2012100611023N).

Pilot Study

To adopt and improve the scenarios and minimize the effect of time horizon[27] on the utility values, we ran a pilot phase with 20 participants. We assumed different intervals for prone positioning (10 days, 1 month, 1 year, 10 years, and lifelong). The participants mentioned that in spite of difficulties of prone positioning, they were unwilling to gamble life or trade life-years to evade this position in short intervals. One way to reduce the influence of the health state temporality on utility value is considering the health state to be chronic and irreversible.[28] On the basis of this assumption and pattern of responses to different intervals of prone positioning, we considered that keeping the position is necessary through lifetime. To quantify the life-years that participants expect for a condition to be lifelong, we asked them to determine the promising life expectancy of their children. The mean of the responses was equal to 70 years, which is close to the life expectancy of Iranians. Accordingly, in all scenarios we proposed 70 years to ask about a lifelong condition.

The other purpose of the pilot study was to determine the order of scenarios. We realized when 5 VAS scenarios were presented consecutively, respondents devoted less attention to the last VAS scenario. Accordingly, the VAS scenarios that tested whether individuals are responding particularly to prone positioning and not just bed rest and can distinguish between different positions were proposed at the end of the interview.

Utility Assessment and Model Description

To eliminate bias, all interviews were performed by 1 reviewer. All interviews were performed in the staff resting room or children playroom on the same floor where the child was admitted. The introduction and interviews took no longer than 20 minutes. We asked about the demographic characteristics of the caregivers after explaining the purpose of the study and the independence of a child's process of diagnosis or treatment from the responses. Then, we provided a brief description about the scaling spectrum on VAS and how the software works. To prevent bias due to method order,[29] we kept the order of the scenarios the same in all interviews—VAS1-2, TTO, modified TTO (m-TTO), chain of gambles (ChGs1-3), SG, and VAS3-5. The rationale is described in the Discussion section.

Visual analogue scale

VAS was assessed with visual scale painometer application by Miguel Martin on an iPad horizontal screen, with Wong Baker schematic faces and a scale maximum of 10.

As a trial of the software, parents determined the utility of their children at the time of interview (VAS1). Then the interviewer proposed other scenarios with simple words. In VAS2, parents were asked to imagine an unknown illness in which the child had to be kept in the prone position. The interviewer emphasized that (1) the illness has no pain or any other signs or symptoms; (2) the motion of the head and limbs is not restricted; (3) the patient has to keep the position while feeding, bathing, and doing all other daily activities; and (4) the patient is intellectually normal and has no difficulty in learning, speaking, or communication. After proposing other scenarios (TTO, m-TTO, ChGs1-3, and SG), we continued VAS scenarios. The scenario in VAS3 was identical to that in VAS2. Comparing VAS2 and VAS3 enabled us to interpret the behavior of the participants and see whether responses were inducible by proposing other scenarios.

To check whether the VAS scenarios are sensitive to prone positioning and not to restriction in bed rest, we assessed VAS4 (simultaneous prone positioning, quadriplegic, and limitation of cervical range of motion) and VAS5 (supine position [lying on

back]). All VAS values less than 1 on the VAS (0-10 scale) were clinically equal to no preference and were replaced by 0.005 to minimize calculation error. Similarly, VAS values greater than 9 (0-10 scale) were considered equal to 9.995.[22] All results on the 0 to 10 scale were transformed to the 0 to 1 scale before analysis. To prevent assigning anchors, the interviewer did not mention "death" while describing the lowest utility state. If the participant selected a value less than 1 (0-10 scale) or implicitly mentioned that the condition was out of bearing, a complementary VAS was conducted to estimate preference of death over questioned health state. The utility of worse than death health conditions is commonly calculated as follows[30]:

$$\text{Utility} = (x - d)/(1 - d), \tag{1a}$$

where x is the scale placement of prone positioning and d is the state of death, both values after transforming to the 0 to 1 interval. This equation gives rise to the values in a scale of $-\infty$ to 0. To obtain values in a $(-1, 0)$ scale and keep the mathematical characteristic of Equation 1a, we calculated the worse than death values as follows:

$$\text{Utility} = (x - d)/d, \tag{1b}$$

where the parameters are the same as in Equation 1a.

Because all other utility instruments produced values in the (0, 1) scale, we used the simplest mathematical approximation to equalize the range of VAS measures with other tools:

$$\text{Utility} = \frac{\text{Utility Equation 1b} + 1}{2}, \tag{1c}$$

where the utility measures are in the (0, 1) scale.

Time trade-off

The principle of TTO is traded years of life to avoid disability.[28] For this purpose, the interviewer asked the subjects to choose between 2 alternatives: to leave their child in prone position for 70 years (t) or accept their child to live shorter but in full health (x). The traded years of life ($t - x$) were being changed until the respondent became indifferent about the 2 alternatives. The commonly used function to assess the utility is as follows[31]:

$$\text{Utility} = x/t, \tag{2}$$

where x is the desired years in full health and t is the life expectancy.

Modified TTO

In m-TTO, a temporary health condition is compared with an alternative of shorter duration of worse health condition, where both conditions are followed by full health. In our study, lifelong prone positioning was compared with an alternative of an imaginary treatment that aggravates the condition of the patient to prone positioning and quadriplegic and complete limitation of cervical range of motion for a limited time, but complete health would be achieved after then. The duration of worse condition was changed until the parent became indifferent between the 2 alternatives. The preference value was calculated as follows:

$$\text{Utility} = 1 - \frac{t1}{t2}, \tag{3}$$

where $t1$ is the duration of the temporary aggravated condition and $t2$ is the duration of the prone positioning, which is 70 years.

Standard gamble

In SG, the subjects are offered a scenario of a particular number of life-years in a constant impaired health state and are asked a gamble between perfect health and death.[28] We requested the participants to consider the prone positioning scenario, and then a new imaginary modality of treatment was introduced to them with either of these 2 outcomes: complete health with probability of p or death with probability of $1 - p$. To prevent anchoring effect,[32] the probabilities for both the perfect health and death were repeated in each turn until the respondent became indifferent between the 2 alternatives. At this point, the probability of the health state was considered as the preferred value of the health condition.[30]

Chain of gambles

Chain methods overcome ceiling effect, while individuals should compare death with health conditions in which utility approaches 0 or 1.[23,28] In ChGs, each health condition is compared with perfect health state and an alternative with immediately dispreferred health state instead of death.[30] To consider worse than death options, we asked the participants to sort the utility of a pediatric patient who should spend the whole life in (1) supine position, (2) prone position, (3) prone position and quadriplegic and complete restriction of cervical range of motion, and (4) death. Utility value of health state in each turn was calculated as follows[23]:

$$\begin{aligned}\text{Utility}(\text{state i}) = {} & p(\text{perfecthealth state i}) \\ & + [1 - p(\text{perfecthealth state i})] \times \text{Utility}(\text{state i} + 1),\end{aligned} \tag{4}$$

where state i is the study position (repeat for $i = 1$-3) and p is the probability of the assumed condition. By considering the utility of the worst preferred state (state 4) equal to 0, all utilities would be derived.

Sample Size and Effective Size Calculation

Sample size for comparing the mean of the utilities among different utility measures was calculated as follows:

$$N = \frac{(1/q1 + 1/q0)(Z\alpha + Z\beta)^2}{(E/S)}, \tag{5}$$

where N is the desired sample size, $q1$ is the proportion of the subjects in 1 utility measure, $q0$ is the proportion of the subjects in another utility measure, $Z\alpha$ is the z score corresponding to the desired level of confidence, $Z\beta$ is the z score corresponding to the desired level of statistical power, α is the threshold probability for rejecting the null hypothesis (type I error rate), β is the probability of failing to reject the null hypothesis under the alternative hypothesis (type II error rate), E is the effect size, and S is the standard deviation (SD) of the outcome in the population. Considering $\alpha = 0.01$, $\beta = 0.2$, $E = 0.5$, and $S = 0.75$ (a maximum approximation of previous studies[33]) and equal proportions of subjects among groups give rise to 53 subjects in each group.

By considering the number of the subjects, the effect size of the study was calculated as follows:

$$\text{Effect size of the study} = S\sqrt{\frac{(1/q1 + 1/q0)(Z\alpha + Z\beta)^2}{N}}, \tag{6}$$

where the parameters are the same as in Equation 5.

Data Analysis

We used the Python packages (Scipy, Pandas, and Scikit) for Linux. Demographic data were presented as mean ± SD. The distribution and normality of each of the utility measure tools (VAS, TTO, m-TTO, SG, and ChGs) were studied using the Kolmogorov–Smirnov test. The differences between instruments were studied by analysis of variance. To identify the pairs with

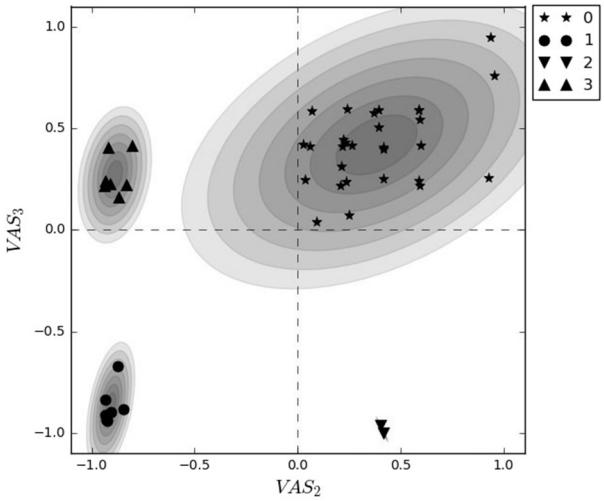


Fig. 1 – The unsupervised Gaussian mixture model fitting of VAS2 vs VAS3. Each point in the diagram represents the responses of 1 interviewee to 2 identical scenarios (VAS2 and VAS3). As demonstrated, responders can be split into 4 clusters on the basis of their behavior. Each shadowed ellipse demonstrates confident levels of fitted Gaussian function equal to 1σ difference. According to the fitted Gaussian function, the difference of clusters is significant up to 7σ. VAS indicates visual analogue scale.

significant difference, we conducted post hoc analysis with the Student's *t* test and the Wilcoxon rank sum test. Because this study compares 6 different tools (VAS2, VAS3, TTO, m-TTO, ChGs, and SG), *P* value of less than .003 $\left(0.05/\binom{6}{2}\right)$ was considered significant in all tests according to Bonferroni correction.[34] To assess the size of the difference between utility instruments, the effect size was evaluated by Cohen's relation.[35] On the basis of Cohen's assumptions, the thresholds of 0.2, 0.5, and 0.8 were interpreted as small, medium, and large effect size. By this means, the result of 2 utility tools with *P* value of more than .003 and small effect size can be considered similar.

Clustering the Respondents

To assess the validity and reproducibility of the utilities, we studied the behavior of the respondents through VAS2 and VAS3 (2 identical scenarios separated by TTO, m-TTO, ChGs, and SG assessments). We have plotted VAS3 values against VAS2 values for each interviewee to see whether individuals changed their responses to these 2 identical scenarios (Fig. 1). We classified the behavior of the interviewees by Gaussian mixture model (GMM), an unsupervised clustering fitting method. A GMM is a parametric probabilistic model to fit a Gaussian density function to a data set. According to the central limit theorem, normal distribution assumption is usually applicable for each data set. GMM assumes that data points are produced by a combination of a finite number of normal distributions. As an unsupervised clustering model, GMM can extract information about the covariance and cluster centers.[27] Accordingly, we could categorize the individuals to separate clusters. The significance of the predicted clustering was computed by covariance matrix of classes. A σ value of more than 3 was considered as statistically different.

Selection Between VAS2 and VAS3

To present the generalization and representativeness power of the VAS2 versus VAS3, we plotted the mean of the responses (VAS2 and VAS3) against different validate margins (*m*) (Fig. 2). The ideal condition is met when VAS2 = VAS3 (variance equal to 0). Validate margin determines the accepted aversion from this ideal. By increasing the *m*, more cases with greater VAS2 and VAS3 variances are included (increase in size and heterogeneity of the sample). A favorable model is the one that presents a more constant mean and predictable behavior regarding different *m* values.

Results

A total of 74 female individuals (pilot study, 20; included, 50; excluded, 4) with a mean age of 30.48 ± 6.66 years were interviewed. One of the respondents was illiterate, 15 (30%) completed primary school, 5 (10%) did not complete high school, 21 (42%) were high school graduates, and 8 (16%) were university graduates. Most (88%) of the respondents were homemakers. One caregiver was a university student, and 5 (10%) of the participants were employed. The mean age of index patients was 31.66 ± 37.62 years and 23 of them were boys. Sex ratio and age distribution of

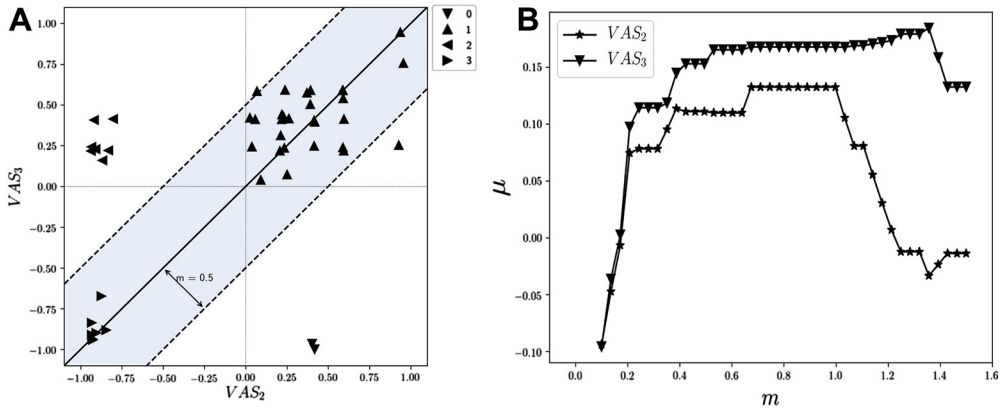


Fig. 2 – The behavior of the VAS according to valid margin. (A) The valid margin (*m*) limits the deviated cases from the ideal responses (VAS2 = VAS3). By increasing the *m* (eg, from 0 to 0.5 as shown), more heterogeneous cases are included. (B) The mean of VAS values in response to *m*. VAS3 shows almost steady and predictable behavior, whereas VAS2 fails to keep the constant trend. VAS indicates visual analogue scale.

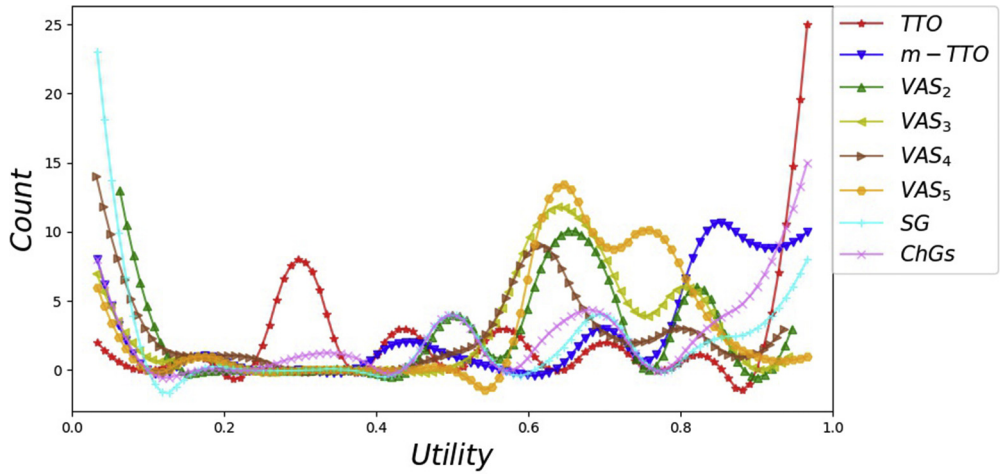


Fig. 3 – The distribution of responses by each utility instrument. (A) VAS2, (B) VAS3, (C) VAS4, (D) VAS5, (E) SG, (F) ChGs, (G) TTO, and (H) m-TTO. ChGs indicates chain of gambles; m-TTO, modified time trade-off; SG, standard gamble; TTO, time trade-off; VAS, visual analogue scale.

girls and boys were not significantly different in pediatric patients. In 54% of the cases, index child was the first child. Most (67%) of the patients were admitted because of spinal intervention. Three patients had multiple admissions (1 needed reoperation).

Two parents quit the interviews, because they became anxious and "they even did not bear to imagine their child in those health conditions." One parent could not continue the interview because of the emergency health condition of the patient and another because of inadequate statistical knowledge.

Figure 3 demonstrates the distribution of the responses by each utility tool. The values of utilities obtained by each instrument are presented in Table 1. Overall, the values obtained by VAS scenarios were lower than those obtained by other utility instruments. The highest utility measure was reported with TTO (0.68 ± 0.36), whereas VAS2 resulted in the lowest utility value (0.57 ± 0.28). Except in VAS, there was no significant intercluster difference.

The values of VAS4 (prone positioning + quadriplegic + limited cervical range of motion) and VAS5 (supine positioning) were different from those of VAS2 and VAS3 (prone positioning; all $P<.003$). This finding demonstrated that VAS2 and VAS3 scenarios are sensitive to prone positioning and respondents had a different attitude toward distinct positions and not just bed rest.

Behavior of the Respondents

To optimize the validity of the responses and reduce the effect of method's order,[29] we analyzed the behavior of the interviewees through repeated VAS measures for 1 identical scenario (VAS2 and VAS3). VAS3 was measured after the introduction of TTO, m-TTO, ChGs1-3, and SG. GMM fitting of the VAS2 and VAS3 responses demonstrated that the participants were categorized into 4 clusters (Fig. 1). Most of the respondents were located in cluster 0 (cluster 0, 29; cluster 1, 7; cluster 2, 2; and cluster 3, 7). Participants were consistent in their responses in cluster 0 (both VAS2 and VAS3 were better than death) and cluster 1 (both VAS2 and VAS3 were worse than death). We considered these participants as valid cases. The respondents changed their opinions in cluster 3 (VAS2, worse than death; VAS3, better than death) and cluster 4 (VAS2, better than death; VAS3, worse than

Table 1 – Utility values obtained by different instruments for each cluster and total valid data (clusters 0 and 1).

Instrument	Valid data	Cluster				P value*
		0	1	2	3	
TTO	0.68 ± 0.36	0.64 ± 0.38	0.86 ± 0.22	0.79 ± 0.30	0.94 ± 0.16	.130
m-TTO	0.66 ± 0.39	0.66 ± 0.38	0.67 ± 0.46	0.57 ± 0.55	0.89 ± 0.04	.440
VAS2 (−1, 1)†	0.13 ± 0.57	0.38 ± 0.26	−0.91 ± 0.04	0.41 ± 0.01	−0.88 ± 0.05	.000
VAS2 (0, 1)‡	0.57 ± 0.28	0.69 ± 0.13	0.05 ± 0.02	0.75 ± 0.01	0.06 ± 0.03	.000
VAS3 (−1, 1)†	0.17 ± 0.55	0.42 ± 0.19	−0.87 ± 0.09	−0.98 ± 0.03	0.27 ± 0.01	.000
VAS3 (0, 1)‡	0.58 ± 0.27	0.71 ± 0.09	0.07 ± 0.04	0.01 ± 0.01	0.63 ± 0.05	.000
SG	0.38 ± 0.43	0.45 ± 0.44	0.07 ± 0.19	0.85 ± 0.21	0.34 ± 0.43	.070
ChGs	0.68 ± 0.36	0.71 ± 0.36	0.57 ± 0.40	0.92 ± 0.11	0.54 ± 0.39	.450

ChGs indicates chain of gambles; m-TTO, modified time trade-off; SG, standard gamble; TTO, time trade-off; VAS, visual analogue scale.
* Intercluster difference.
† Values in (−1, 1) scale.
‡ Values in (0, 1) scale, as described in Equation 1b.

death). These participants were considered as outliers and invalid cases. There was no intersection between the clusters up to 7σ according to GMM covariance matrices.

To better interpret the behavior of participants with invalid responses (clusters 2 and 3), we evaluated the trend of VAS2 and VAS3 while sample size and heterogeneity were increased. As evident in Figure 2, VAS3 is more valid than VAS2, because the mean value tested by VAS3 remains steady, whereas VAS2 loses the stable trend. In valid cases (clusters 0 and 1), the preferences obtained by TTO, m-TTO, and ChGs were similar (all $P>.003$ and small effect size), whereas they were significantly different from VAS2 or VAS3 (very large effect size). Two respondents in cluster 2 changed their opinion (VAS2 > VAS3). All other tools were in consistence with each other in this cluster (for all pairs $P>.003$). In cluster 3, the interviewees showed similar responses between each 2 pairs of scenarios coming after each other (for all pairs $P>.003$).

Estimate of Prone Positioning Utility

Because of limitations and considerations regarding each utility instrument (see Discussion section), the authors decided to develop an adopted model to combine more than one utility tool. The candidate instruments should have similar results ($P>.003$ and small effect size). Analysis of variance (all data) gave rise to a dramatically small P value, demonstrating the utility difference between some pairs. Results from post hoc analysis and pairwise comparison of the utility instruments indicate the similarities of TTO, m-TTO, and ChGs (low effect size and high P values in all pairs). As demonstrated in Appendix Figure 1 in Supplemental Materials found at https://doi.org/10.1016/j.vhri.2019.01.003, other pairs were significantly different (low P values and high effect size). Further analyses (Table 1) showed that utility values measured by each of these 3 instruments are similar when comparing the 4 clusters (intercluster consistency; $P_{TTO} = 0.13$, $P_{m\text{-}TTO} = 0.44$, and $P_{ChGs} = 0.45$). Interestingly, pairwise comparison of the utility values obtained by TTO, m-TTO, and ChGs was similar within each of the 4 clusters (intracluster consistency) (see Appendix Table 1 in Supplemental Materials found at https://doi.org/10.1016/j.vhri.2019.01.003). These findings indicate that TTO, m-TTO, and ChGs are not sensitive to the behavioral characteristics of the respondents (see Discussion section). Accordingly, we calculated the utility of the prone positioning by considering these 3 instruments:

$$\text{Utility} = \frac{\sum_{i=1}^{n}(Xi\sigma_i^{-2})}{\sum_{i=1}^{n}\sigma_i^{-2}} \pm \frac{1}{\sum_{i=1}^{n}\sigma_i^{-2}}, \quad (7)$$

where X_i and σ_i^{-2} are mean and SD of each instrument, respectively, and n is equal to 3.

Appendix Table 2 in Supplemental Materials found at https://doi.org/10.1016/j.vhri.2019.01.003 shows the effect size and P values regarding the comparison of this utility with values of each instrument—under 4 clusters, valid cases, invalid cases, and cumulative data. As shown in Figure 4, the adopted model has significant difference with the solitary measured values of VAS and SG ($P<.003$ and effect size > 0.8). To further improve the homogeneity of the results, we conducted a cluster analysis of the data. When outliers (9 respondents in clusters 2 and 3, invalid cases) are omitted, the utility values of our adopted model differ only from SG ($P<.003$ and effect size = 4.05). Cluster analysis of comparing our model to other utility tools is shown in Appendix Figure 2 in Supplemental Materials found at https://doi.org/10.1016/j.vhri.2019.01.003. On the basis of Equation 6, our study has a medium effect size (0.513).

Omitting the outliers yield in similarities of TTO, m-TTO, and ChGs with small to medium effect sizes (TTO−m-TTO: effect size, 0.32, $P=.640$; TTO−ChGs: effect size, 0.06, $P=.970$; m-TTO−ChGs: effect size, 0.26, $P=.850$) (Fig. 5). Calculated utility for prone positioning in valid cases according to our model is 0.68 ± 0.21.

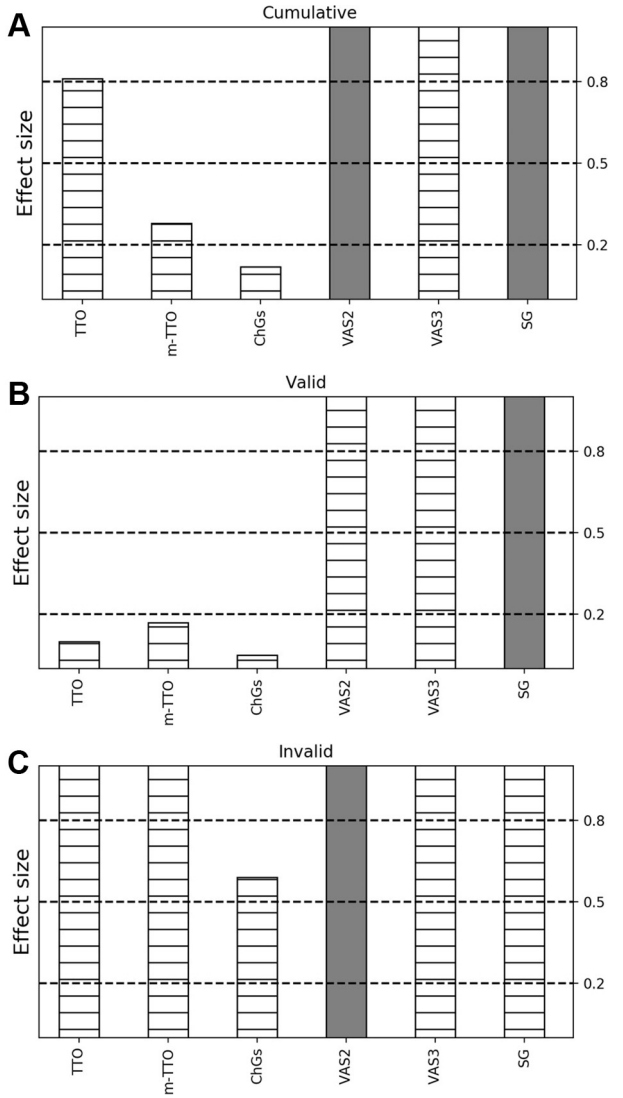


Fig. 4 – The effect size while comparing our proposed model with each utility instrument. Results are presented as (A) cumulative data (all clusters), (B) valid cases (clusters 0 and 1), and (C) invalid cases (clusters 2 and 3). Thresholds for small (0.2), medium (0.5), and large (0.8) effect size are demonstrated as dashed lines. Solid bars demonstrate $P<.003$. ChGs indicates chain of gambles; m-TTO, modified time trade-off; SG, standard gamble; TTO, time trade-off; VAS, visual analogue scale.

The measured utility of prone positioning via our model is equal to 0.68 ± 0.21. We recommend using TTO, m-TTO, and ChGs simultaneously to obtain the valid utility value for temporary health states when proxies are involved.

Discussion

We designed this study to evaluate the preference of prone positioning, one of the postoperational prophylactic modalities of

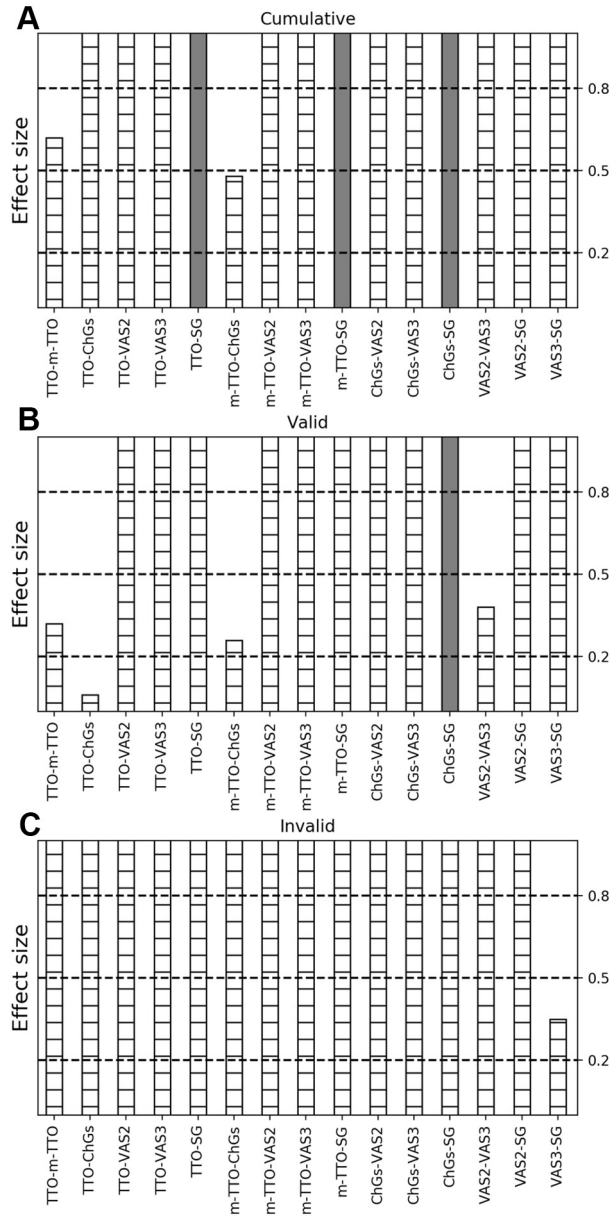


Fig. 5 – The effect size while comparing each pair of scenarios. (A) Cumulative data, (B) valid cases in clusters 0 and 1, and (C) invalid cases in clusters 2 and 3. Solid bars demonstrate P<.003. ChGs indicates chain of gambles; m-TTO, modified time trade-off; SG, standard gamble; TTO, time trade-off; VAS, visual analogue scale.

care, in patients who underwent untethering surgeries. We introduced an adopted utility instrument on the basis of weighted mean of TTO, m-TTO, and ChGs to measure the utility of temporal health state among pediatric patients. The utility values obtained by TTO, m-TTO, and ChGs have intracluster consistency (similar values when comparing the results of these 3 instruments within each cluster) and intercluster consistency (similar values when comparing the result of each instrument between the 4 clusters). These 3 utility tools also provide similar results when all responders are considered cumulatively (indifference about clustering the interviewee by repeated VAS measures). Other instruments did not provide such stable results. We offered Equation 7 as an adopted model to measure the temporary health state preference in pediatrics. On the basis of this equation, we calculated prone positioning utility equal to 0.68 ± 0.21. Although TTO, m-TTO, and SG had similar outcomes, there are some considerations and limitations regarding the validity of the values obtained by each instrument per se.

Considerations Regarding TTO

First, different years of life do not have the same value.[26,36] The greatest segregation between the utility of a person in prone position in comparison with a person in perfect health of the identical age is derived in the productive years of life, followed by a milder slope of disutility. Our participants frequently stated that they would like to deal many years of life to save school age and age of marriage. TTO trades the years at the end of life, which is associated with disabilities, illness, and lower productivity due to the normal aging process. TTO or any other method that assumes health state to be linearly related to time should be used with caution.[37] Second, for a subject in perfect health, the disutility of additional life-years due to the physiologic process of aging has a mild slope, which starts from years near the end of life (diminishing marginal utility).[28,36] Nevertheless, when prone positioning (or any other temporary health state) is considered to be chronic, the condition is judged to be tough and each additional day of living has a great aversion. Thereby, the degree of disutility regarding additional years in temporary health state surpasses the slope of disutility by aging.[37] Third, the key assumption for TTO is "constant proportionality."[31] When the marginal utility functions of additional years in both perfect health and health states of interest have similar characteristics, different durations of health status with similar proportions—in comparison with perfect health—have identical TTO values. Accordingly, even if the utilities of time in health states of interest and perfect health are not constant, but rather matched by nonlinear properties, constant proportionality will result.[31] Fourth, when proxies are considered to determine the utility of the health state (as in pediatrics studies), the response may be under influence of their desire and comfort rather than the difficulty of the health condition for the child. In tough conditions that the proxy should support the child, some individuals bear the health state only for a limited time (maximum endurable time) and the estimated utilities would vary significantly from real estimate for the child.[31,37]

Considerations Regarding m-TTO

To overcome the mentioned challenges about TTO, we devised m-TTO in which the unfavorable outcome precedes the healthy life-years. In m-TTO, parents are asked to declare their maximum tolerance for taking care of a child with severe disability (quadriplegic and in prone position) to meet the years with full health. Although this method weighs the preference of the child over the caregiver, the tolerance of the proxy to support the child and the emotional connection between her and the child affect the results. In addition, the participants show a bipolar pattern of response—either accept the condition or deny dealing with it (data not presented). Because the TTO postpones the traded years to the end of life, the pattern of its responses is almost complementary to m-TTO.

Considerations Regarding ChGs

ChGs was the other utility tool with consistent results. It has been stated that respondents find it difficult to gamble death with the intermediate outcomes and especially conditions with similar preference as perfect health. ChGs keeps the probability and

risk-taking nature of SG, but softens the choices in each step.[23,28] In our study, when participants interpreted the prone positioning to be worse than death (cluster 0), VAS values were close to −1 and SG values were close to 0—representing that respondents accept more than 97% risk of death to get rid of the disability. In ChGs, the scenarios are narrated in a way that step by step the interviewee should deal with a more severe health condition. In the last set of the test, the respondents should gamble health condition with death. Interestingly, we observed an agreement of SG and VAS with the last ChGs. Accordingly, addition of the ChGs to Equation 7 implicitly contributes to the risk taking and mortality acceptance of the respondents.

To address all these concerns, we developed an adopted model based on TTO, m-TTO, and ChGs (Equation 7). Besides considering the strengths and limitations of these 3 instruments together, our adopted utility instrument has smoother distribution of the responses through all possible utility values. In comparison, each of these tools produce limited domain of the utility values (Fig. 3).

Behavioral Classification of Interviewees

We classified our data on the basis of the behavior of the respondents through 2 identical VAS measures (Fig. 2). All participants of cluster 0 interpreted prone positioning as being worse than death and had a consistent opinion between VAS2 and VAS3 (VAS2 = −0.907 ± 0.035 and VAS3 = −0.868 ± 0.093; P=.470). The preference obtained by SG was 0.071 ± 0.188, demonstrating that the respondents in cluster 0 were willing to accept even up to 93% risk of death to get rid of the prone positioning and achieve full health. The results of VAS and SG were consistent (VAS2−SG: effect size, 1.79, P=.189; VAS3−SG: effect size, 3.26, P=.033). These findings might show the risk-taking behavior of the participants, as well as low tolerance or underlying mood disorder, which made death more favorable for them. We could not present the association between VAS and SG in cluster 1, in which the participants had more positive opinions about the prone positioning. Two participants of cluster 2 gave a lower preference in VAS3 than in VAS2. Because all the other pairs (except VAS3−ChGs) were in consistence with each other, it is possible that the VAS results were just an error to work with the software. In cluster 3, the result of each instrument was consistent with the previous assessment—a possible trend in which the response would be under influence of the previous scenario. These respondents changed their opinion from worse than death in VAS2 to more positive values in VAS3. Because VAS3 was more stable than VAS2 (Fig. 3), the response change in this cluster might be due to a more realistic understanding of the situation after introducing more severe conditions (such as being paralyzed).

Preference Assessment of Temporary Health States

Constant proportional trade-off, risk neutrality across life-years, and independence of preferences for health regardless of life expectancy are 3 major assumptions regarding the utility values for generating quality-of-life assessments.[28] Temporality of a health state produces bias in measuring the utility value. When parents are asked to trade life-years or gamble death to evade the temporary health state in our pilot study, almost all the participants avoid taking the deal because of very short duration of the health state, high life expectancy of a child, and emotional engagement with the child. One strategy to deal with these limitations was considering the temporary prone positioning as if it is chronic and irreversible.[28] Accordingly, and on the basis of the result of our pilot study on various durations of prone positioning (10 days, 1 month, 1 year, 10 years, and lifelong), we considered all scenarios to be lifelong (70 years, similar to life expectancy in Iran) and irreversible. Because none of the existing instruments fully matched the characteristics of our study, we used different measures (VAS, TTO, m-TTO, SG, and ChGs) to determine the most appropriate tool for assessing the utility of prone positioning, and similar pediatric temporary health impairments.

Strengths and Limitations

Use of different utility instruments and validating the results on the basis of the behavior of the respondents are strengths of our study. Nevertheless, worse than death scenarios could not be applied in all utility measures. In addition, illiterates and people with low sense of statistics cannot engage with our scenarios. Further studies with larger sample sizes including both males and females and random order of utility measures are required to determine the power of our recommended utility function.

One concern about our study is keeping the order of the scenarios the same through all interviews. When TTO scenarios are described first, more risk-aversion results are obtained than when SG scenarios are primarily introduced.[29] In addition, health states with higher aversion and lower utility—such as prone positioning in our study—are more affected by the assessment order.[29] Participants can remember the prominent component of each scenario through the interview. To reduce the major impact of assessment order on outcome,[29] we introduced all interviews in the same order. As described earlier, we repeated 1 of the VAS scenarios (VAS2 and VAS3) and analyzed the behavior of the respondents accordingly. We incorporated only the values obtained from clusters 0 and 1 into the model—those that the VAS3 responses remained consistent with VAS2 even after introduction of TTO, m-TTO, ChGs, and SG. By this strategy we (1) reduced the invalid responses of the individuals who had inconsistent opinion through the interview and (2) homogenized the impact of each utility tool on consecutive scenarios. Nevertheless, to better minimize the test order effect, a random set of scenarios should be introduced in a large sample size.

Conclusions

By considering the utility values obtained from TTO, m-TTO, and ChGs, we recommended an adopted model for assessment of HRQOL in children with temporary health states to overcome the challenges of each isolated instrument. Based on this model, the utility value of prone positioning in pediatric patients was estimated as 0.68 ± 0.21.

Supplemental Materials

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.vhri.2019.01.003.

REFERENCES

1. Shahjouei S, Hanaei S, Habibi Z, Hoseini M, Ansari S, Nejat F. Randomized clinical trial of acetazolamide administration and/or prone positioning in mitigating wound complications following untethering surgeries. *J Neurosurg Pediatr*. 2016;17(6):1–8.
2. Hawthornel G, Richardson J, Day NA. A comparison of the Assessment of Quality of Life (AQoL) with four other generic utility instruments. *Ann Med*. 2010;33(5):358–380.
3. Brunner HI, Klein-gitelman MS, Miller MJ, et al. Health of children with chronic arthritis: relationship of different measures and the quality of parent proxy reporting. *Arthritis Rheum*. 2004;51(5):763–773.
4. Marra CA, Woolcott JC, Kopec JA, et al. A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. *Soc Sci Med*. 2005;60(7):1571–1582.

5. Kontodimopoulos N, Pappa E, Papadopoulos AA, Tountas Y, Niakas D. Comparing SF-6D and EQ-5D utilities across groups differing in health status. *Qual Life Res*. 2009;18(1):87–97.
6. Hullmann SE, Ryan JL, Ramsey RR, Chaney JM, Mullins LL. Measures of general pediatric quality of life: Child Health Questionnaire (CHQ), DISABKIDS Chronic Generic Measure (DCGM), KINDL-R, Pediatric Quality of Life Inventory (PedsQL) 4.0 Generic Core Scales, and Quality of My Life Questionnaire (QoML). *Arthritis Care Res (Hoboken)*. 2011;63(S11):420–430.
7. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Econ*. 2004;13(9):873–884.
8. Finnell SME, Carroll AE, Downs SM. Application of classic utilities to published pediatric cost-utility studies. *Acad Pediatr*. 2012;12(3):219–228.
9. Heintz E, Wiréhn AB, Peebo BB, Rosenqvist U, Levin LÅ. QALY weights for diabetic retinopathy—a comparison of health state valuations with HUI-3, EQ-5D, EQ-VAS, and TTO. *Value Health*. 2012;15(3):475–484.
10. Haverman L, Limperg PF, Young NL, Grootenhuis MA, Klaassen RJ. Paediatric health-related quality of life: what is it and why should we measure it? *Arch Dis Child*. 2017;102(5):393–400.
11. Parmar A, Vandriel SM, Ng VL. Health-related quality of life after pediatric liver transplantation: a systematic review. *Liver Transpl*. 2017;23(3):361–374.
12. van Mater HA, Williams JW, Coeytaux RR, Sanders GD, Kemper AR. Psychometric characteristics of outcome measures in juvenile idiopathic arthritis: a systematic review. *Arthritis Care Res (Hoboken)*. 2012;64(4):554–562.
13. Eiser C, Morse R. Quality-of-life measures in chronic diseases of childhood. *Health Technol Assess*. 2001;5(4):1–157.
14. Hersh A. Measures of health-related quality of life in pediatric systemic lupus erythematosus: Childhood Health Assessment Questionnaire (C-HAQ), Child Health Questionnaire (CHQ), Pediatric Quality of Life Inventory Generic Core Module (PedsQL-GC), Pediatric Quality of Life Inventory Rheumatology Module (PedsQL-RM), and Simple Measure of Impact of Lupus Erythematosus in Youngsters (SMILEY). *Arthritis Care Res (Hoboken)*. 2011;63(suppl 11):446–453.
15. Gigi M, Roth J, Eshel R, Constantini S, Bassan H. Health-related quality of life after post-haemorrhagic hydrocephalus in children born preterm. *Dev Med Child Neurol*. 2018;61(3):343–349.
16. Gershon RC, Lai JS, Bode R, et al. Neuro-QOL: quality of life item banks for adults with neurological disorders: item development and calibrations based upon clinical and general population testing. *Qual Life Res*. 2012;21(3):475–486.
17. Carlozzi NE, Hahn EA, Goodnight SM, et al. Patient-reported outcome measures in Huntington disease: quality of life in neurological disorders (Neuro-QoL) social functioning measures. *Psychol Assess*. 2018;30(4):450–458.
18. Cella D, Lai J-S, Nowinski CJ, et al. Neuro-QOL: brief measures of health-related quality of life for clinical research in neurology. *Neurology*. 2012;78(23):1860–1867.
19. Lai J-S, Nowinski C, Victorson D, et al. Quality-of-life measures in children with neurological conditions: pediatric Neuro-QOL. *Neurorehabil Neural Repair*. 2012;26(1):36–47.
20. Kemper A, Coeytaux R, Sanders G, et al. *Disease-Modifying Antirheumatic Drugs (DMARDs) in Children With Juvenile Idiopathic Arthritis (JIA). Report No.: 11-EHC039-EF. AHRQ Comparative Effectiveness Reviews*. Rockville, MD: Agency for Healthcare Research and Quality; 2011.
21. Griebsch I, Coast J, Brown J. Quality-adjusted life-years lack quality in pediatric care: a critical review of published cost-utility studies in child health. *Pediatrics*. 2005;115(5):e600–e614.
22. Bennett JE, Sumner W, Downs SM, Jaffe DM. Parents' utilities for outcomes of occult bacteremia. *Arch Pediatr Adolesc Med*. 2000;154(1):43–48.
23. Carroll AE, Downs SM. Improving decision analyses: parent preferences (utility values) for pediatric health outcomes. *J Pediatr*. 2009;155(1):21–25.
24. Matza LS, Boye KS, Feeny DH, Johnston JA, Bowman L, Jordan JB. Impact of caregiver and parenting status on time trade-off and standard gamble utility scores for health state descriptions. *Health Qual Life Outcomes*. 2014;12(1):48.
25. Jakubczyk M, Golicki D, Niewada M. The impact of a belief in life after death on health-state preferences: true difference or artifact? *Qual Life Res*. 2016;25(12):2997–3008.
26. Arnesen TM, Norheim OF. Quantifying quality of life for economic analysis: time out for time tradeoff. *Med Humanit*. 2003;29(2):81–86.
27. Matza LS, Boye KS, Feeny DH, et al. The time horizon matters: results of an exploratory study varying the timeframe in time trade-off and standard gamble utility elicitation. *Eur J Health Econ*. 2016;17(8):979–990.
28. Wright DR, Wittenberg E, Swan JS, Miksad RA, Prosser LA. Methods for measuring temporary health states for cost-utility analyses. *Pharmacoeconomics*. 2009;27(9):713–723.
29. Finnell SME, Carroll AE, Downs SM. The utility assessment method order influences measurement of parents' risk attitude. *Value Health*. 2012;15(6):926–932.
30. Torrance GW. Measurement of health state utilities for economic appraisal. *J Health Econ*. 1986;5(1):1–30.
31. Buckingham KJ, Joy N. A note on the nature of utility in time and health and implications for cost utility analysis. *Soc Sci Med*. 2009;68(2):362–367.
32. Nord E. Methods for quality adjustment of life years. *Soc Sci Med*. 1992;34(5):559–569.
33. Franic DM, Pathak DS. Effect of including (versus excluding) fates worse than death on utility measurement. *Int J Technol Assess Health Care*. 2003;19(2). S026646230300031X.
34. Bonferroni CE. Teoria statistica delle classi e calcolo delle probabilita, Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze. 1936;8(2):3–62.
35. Sullivan GM, Feinn R. Using effect size—or why the P value is not enough. *J Grad Med Educ*. 2012;4(3):279–282.
36. Attema AE, Brouwer WBF. The correction of TTO-scores for utility curvature using a risk-free utility elicitation method. *J Health Econ*. 2009;28(1):234–243.
37. Dolan P. Modelling valuations for health states: the effect of duration. *Health Policy*. 1996;38(3):189–203.