



Measuring Health Spillover Effects in Caregivers of Children with Autism Spectrum Disorder: A Comparison of the EQ-5D-3L and SF-6D

Clare C. Brown¹ · J. Mick Tilford¹ · Nalin Payakachat² · D. Keith Williams³ · Karen A. Kuhlthau^{4,5} · Jeffrey M. Pyne⁶ · Renske J. Hoefman⁷ · Werner B. F. Brouwer⁷

Published online: 13 March 2019
© The Author(s) 2019

Abstract

Background and Objective Healthcare interventions that improve the health of children with autism spectrum disorder (ASD) have the potential to affect the health of caregivers. This study compares the three-level EuroQoL-5 Dimension (EQ-5D-3L) and the Short Form-6 Dimension (SF-6D) in their ability to value such spillover effects in caregivers.

Methods Clinical data collected from two Autism Treatment Network (ATN) sites was combined with survey data of caregivers of children diagnosed with ASD. Caregivers completed instruments by proxy describing child health and completed the EQ-5D-3L and SF-6D preference-weighted instruments to describe their own health.

Results There was a strong correlation between the health utility scores of the two preference-weighted instruments ($\rho = 0.6172$, $p < 0.001$) measuring caregiver health-related quality of life. There was a similar correlation between both the SF-6D and EQ-5D-3L scores with a previously validated care-related quality of life measure (Care-related Quality of Life instrument [CarerQoL-7D]) ($\rho = 0.569$, $p < 0.001$ and $\rho = 0.541$, $p < 0.001$, respectively). The mean SF-6D scores for caregivers differed significantly in relation to four of the five child health or behavior measures whereas the EQ-5D-3L differed for only two of them.

Conclusions Health utility values of caregivers for children with ASD vary by the health characteristics of the child, suggesting significant potential for spillover effects. The comparison of the EQ-5D-3L and SF-6D demonstrated that both instruments can be used to estimate spillover effects of interventions to improve child health, but the SF-6D exhibited greater sensitivity to child health among children with ASD.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s40273-019-00789-2>) contains supplementary material, which is available to authorized users.

✉ J. Mick Tilford
tilfordmickj@uams.edu

¹ Department of Health Policy and Management, University of Arkansas for Medical Sciences, 4301 W. Markham, Slot 820, Little Rock, AR 72205, USA

² Division of Pharmaceutical Evaluation and Policy, University of Arkansas for Medical Sciences, Little Rock, AR, USA

³ Department of Biostatistics, University of Arkansas for Medical Sciences, Little Rock, AR, USA

⁴ Department of Pediatrics, Harvard Medical School, Boston, MA, USA

⁵ Center for Adolescent Health Policy, Massachusetts General Hospital, Boston, MA, USA

⁶ Center for Mental Healthcare and Outcomes Research, Central Arkansas Veterans Healthcare System and Psychiatric Research Institute, University of Arkansas for Medical Sciences, Little Rock, AR, USA

⁷ Erasmus School of Health Policy and Management, Erasmus University, Rotterdam, The Netherlands

Key Points for Decision Makers

Health interventions that benefit patients can positively affect family members.

Accurate measurement of these spillover effects is necessary to appropriately value health programs and technologies.

Both the Short Form-6 Dimension (SF-6D) and the three-level EuroQoL-5 Dimension (EQ-5D-3L) can be used to estimate potential spillover effects associated with interventions for children with ASD, but the former performed slightly better in this population.

1 Introduction

A common metric used to quantify the effectiveness of healthcare interventions in cost-effectiveness evaluations is the quality-adjusted life-year (QALY), which incorporates both the quantity and quality of life gained [1, 2]. The QALY has a number of useful properties that led to it being the standard for conducting cost-effectiveness analysis [2]. One area of concern, however, is the fact that, in practice, QALY measurement for cost-effectiveness analysis typically focuses solely on the health effects accruing to patients, as if these were isolated individuals [3]. By now, it has been shown that health effects in patients are typically associated with substantial spillover effects on the health and well-being of caregivers and family members [4–6]. Failure to include such spillover effects in economic evaluations can lead to a misrepresentation of the burden of disease and the benefits of health interventions [7]. This, in turn, may lead to suboptimal decisions, both from a healthcare and a societal perspective [8].

Regulatory agencies now recognize the need to incorporate spillover effects in economic evaluations. Both the National Institute for Health and Care Excellence (NICE) and the Second US Panel on Cost-Effectiveness in Health and Medicine recognized the potential for spillover effects to influence estimated cost-effectiveness ratios and recommend including them in a reference case analysis [9–12]. The Second US Panel also emphasized the importance of increasing research efforts on clarifying *how* to incorporate family and caregiver spillover effects in economic evaluations [10].

Despite the recognition for the need to include health spillover effects when valuing health interventions, little guidance exists for including spillover effects in cost-effectiveness analysis [8]. For example, there is no guidance for incorporating spillover effects into a cost-effectiveness analysis in the context of clinical trials that could inform

regulatory agencies. One option would be to capture these effects by measuring health states across trial arms for patients, caregivers, and family members. This results in a focus on health (rather than well-being), which has the advantage of being the most relevant outcome in most studies and decision-making contexts, making effects comparable across groups and able to be aggregated. In the design of such clinical trials, decisions need to be made about which instruments are able to capture spillover effects in QALY terms. Early research on spillovers in Alzheimer's disease attempted to estimate effects by comparing caregiver outcomes across clinical characteristics such as stage of disease and setting [13], but likely failed because the instrument was not sensitive or did not discriminate well [7, 14]. Subsequent analysis showed that traditional measures of burden and health changed in the expected direction, but the Health Utilities Index Mark 2 (HUI-2) did not capture these changes [14]. While a large literature has emerged that allows us to understand whether a given instrument is valid for measuring QALYs for different conditions affecting patient populations [15–17], research identifying instruments that are valid and responsive in measuring spillover effects in caregivers and family members remains understudied. Indeed, we are aware of only two studies that have compared different generic preference-weighted instruments to measure spillover effects. Payakachat et al. [18] compared three preference-weighted health instruments to measure spillover effects among caregivers of children with craniofacial malformations. Bhadhuri et al. [19] compared two preference-weighted instruments to measure spillover effects among family members of meningitis survivors.

Family and caregiver spillover effects, in terms of health and well-being, may be particularly pronounced in child health interventions [20–22] and for mental health conditions where social support systems may be lacking [23, 24]. Interventions for children with autism spectrum disorder (ASD) have the potential for substantial spillover effects in caregivers and family members due to an increased prevalence of psychiatric and medical co-morbidities such as anxiety, behavioral problems, sleep disturbance, and cognitive issues [25–27]. Preventing symptoms of ASD in the child is thus likely to improve family and caregiver health and reduce burden [22].

Given the potential for interventions such as medications or applied behavioral therapy to improve the health of children with ASD, a comparison of generic preference-weighted instruments to determine whether they capture spillover effects in QALY terms of health associated with treatment for children with ASD appears warranted. Therefore, the purpose of this study was to assess the ability of two commonly used generic preference-weighted instruments, the three-level EuroQoL-5 Dimension instrument (EQ-5D-3L) and the Short Form-6 Dimension (SF-6D)

[28, 29], derived from the 12-item Short Form survey version 2.0 (SF-12 v2.0), to value spillover effects in caregivers of children with ASD in order to provide guidance about their use, especially in the context of clinical trials and other approaches where indirect elicitation techniques are warranted.

2 Methods and Participants

2.1 Data Collection

This study is a secondary analysis of data we previously collected from two Autism Treatment Network (ATN) sites (a developmental center in Little Rock, AR, USA and an outpatient psychiatric clinic at Columbia University Medical Center in New York, NY, USA). The dataset consists of clinical registry data and data collected via a postal survey of caregivers of children aged 4–17 years old with an ASD diagnosis, which was clinically determined by *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition* (DSM-IV) criteria [30]. The ATN sites collected clinical registry data of children with ASD that included diagnostic, cognitive, behavioral, and physical assessments. Caregivers of children with ASD registered at the ATN sites who had agreed to be contacted for future research were mailed a postal survey and asked to report on instruments describing their child with ASD and themselves. Data were collected from 2010 to 2012. Around 10% and 5% of families in Little Rock and Columbia opted out of being contacted for future research. The study protocol was approved by all of the institutions involved in the study. A more detailed description of the data collection procedures is outlined in Hoefman et al. [31] and Tilford et al. [32].

2.2 Instruments

2.2.1 Information on Children with Autism Spectrum Disorder

A number of clinical and health-related quality of life (HR-QOL) measures for children with ASD were selected from the ATN assessments, including autism severity scores (Autism Diagnostic Observation Schedule [ADOS] [33]), adaptive behavior scores (Vineland Adaptive Behavior Scales Second Edition [Vineland-II] [34]), cognitive ability (IQ; the Stanford-Binet Intelligence Scales [35], the Mullen Scales of Early Learning [36], or the Bayley Scales of Infant and Toddler Development [37], depending on the child's age), emotional and behavioral problems (Child Behavior Checklist [CBCL] [38]), sleep behavior (Children's Sleep Habits Questionnaire [CSHQ] [39]), and pediatric quality of life measures (the Pediatric Quality of Life Inventory™ 4.0 [PedsQL™] [40] and the

Health Utilities Index Mark 3 (HUI-3) [41]). Higher Vineland-II and IQ scores indicate better child adaptive behavior and cognitive ability, respectively. Higher ADOS, CBCL, or CSHQ scores indicate increased autism severity, maladaptive behavior, or worse sleep behaviors, respectively. Higher scores on the PedsQL™ and HUI-3 indicate better child HR-QOL. Responses for the PedsQL™, HUI-3, Vineland-II, CBCL, and CSHQ were reported by the caregiver; other child data came from clinical assessment. Further details on the measures can be found in the Electronic Supplementary Material.

We additionally obtained information on the child's age and gender. The child's age was included in the analysis given the evidence that maladaptive behaviors among individuals with ASD improve with age, suggesting that younger children with ASD may require increased caregiving relative to older children with ASD [42, 43]. Child behavior and health conditions can be expected to influence caregiver burden and ultimately caregiver HR-QOL following the logic in previous studies [13, 18, 31, 44, 45].

2.2.2 Information on Caregivers

Caregivers reported information on their demographic characteristics, depressive symptoms (Center for Epidemiologic Studies Depression Scale [CES-D] [46]), care-related quality of life (Care-related Quality of Life instrument [CarerQol-7D] [47, 48]), family-related quality of life (Family Quality of Life Scale [FQLS] [49]), and the EQ-5D-3L and SF-12 v2.0 instruments measuring HR-QOL. All of the caregiver information was provided through the postal survey.

The EQ-5D-3L (range – 0.109 to 1) measures health utility using five dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) with three response options each [28, 50]. The SF-12 v2.0 was used to derive the SF-6D (range 0.3–1), which contains six dimensions of health (physical functioning, role limitations, social functioning, pain, energy, and mental health) [29]. Both the EQ-5D-3L and SF-6D measures provide a health utility score where 0 equals the score related to the health state of dead, 1 represents the state of perfect health, and scores less than 0 represent states worse than death.

Higher scores on the CarerQol-7D, FQLS, EQ-5D-3L, and SF-6D indicate higher levels of quality of life. Higher scores on the CES-D indicate worse depressive symptoms. Details on each of the measures can be found in the Electronic Supplementary Material.

3 Statistical Analysis

3.1 Descriptive Statistics

Domain and health utility scores were calculated using weights from the US general population for the EQ-5D-3L and SF-6D [51, 52]. Descriptive statistics were calculated for child and caregiver demographics, health, and HR-QOL.

3.2 Validation of the Three-Level EuroQoL-5 Dimension (EQ-5D-3L) and the Short Form-6 Dimension (SF-6D) to Assess Spillover Effects

To study the ability of the two generic preference-weighted instruments to assess spillover effects in caregivers of children with ASD, we first compared the convergent validity and clinical validity of the SF-6D and EQ-5D-3L in measuring caregiver HR-QOL. Failure to meet the criteria for convergent validity and clinical validity for measuring caregiver HR-QOL would discredit the ability of the SF-6D and EQ-5D-3L for measuring spillover effects. We then assessed the discriminative power and clinical validity of the SF-6D and EQ-5D-3L in specifically measuring spillover effects related to caring for a child with ASD. In the discriminative power and clinical validity analyses, comparisons were made using both caregiver and child measures to assess the two preference-weighted instruments in their ability to measure spillover effects. Statistical significance was assumed at $p < 0.05$, using Bonferroni correction when multiple comparisons were performed simultaneously.

3.2.1 Convergent Validity of EQ-5D-3L and SF-6D

Convergent validity is the agreement between two instruments that are measuring the same theoretical construct, namely caregiver health effects [53]. In this study, we analyzed the agreement between the EQ-5D-3L and the SF-6D and agreement between these HR-QOL measures with validated measures of outcomes in caregivers (i.e., CES-D and CarerQoL-7D) using Spearman's ρ correlation. Both the CES-D and CarerQoL-7D measure quality-of-life outcomes in caregivers; thus, a strong correlation between the EQ-5D-3L and SF-6D with each other and with other caregiver measures will indicate the ability of the EQ-5D-3L and SF-6D to capture HR-QOL effects. Following convention, Spearman's ρ between 0.10 and 0.29, 0.30 and 0.49, and > 0.50 were classified as weak, moderate, and strong effect sizes, respectively [54, 55].

Further, to evaluate the agreement between the EQ-5D-3L and SF-6D, a Bland–Altman plot was created, which uses a scatter plot to display the difference between

the two health utility scores for a given caregiver and the average of the two scores for that caregiver [56]. This allows a comparison of whether the scores are similar across the range of HR-QOL.

Based on findings in other patient populations, we hypothesized that positive agreement would exist between the health utility scores of the two preference-weighted measures of caregiver health [15, 57]. Similarly, we hypothesized that both of the health utility measures will be positively correlated with the CarerQoL-7D and negatively correlated with the CES-D [18, 58].

3.2.2 Clinical Validity of EQ-5D-3L and SF-6D

Clinical validity describes how differences in clinical- or behavioral-related characteristics are reflected in an individual's instrument score [59]. Clinical validity was assessed using a one-way analysis of variance (ANOVA) to test differences in the average health utility scores of caregivers (EQ-5D-3L and SF-6D) with different characteristics. First, we compared mean scores on the EQ-5D-3L and SF-6D among subgroups of caregivers based on the caregiver's number of hours sleep per night, depressive symptoms (CES-D), care-related quality of life (Carer-QoL-7D), and family-related quality of life (FQLS) to assess the clinical validity of the EQ-5D-3L and SF-6D. In this analysis, the criteria for clinical validity was met if higher health utility scores measured by the EQ-5D-3L and SF-6D were associated with increased hours of sleep, caregiver quality of life (Carer-QoL-7D), and family-related quality of life (FQLS), and decreased depressive symptoms (CES-D).

Next, we compared the average health utility scores for caregivers in relation to the child's IQ level, autism severity (ADOS score), HR-QOL (PedsQL™ and HUI-3), behavioral characteristics (Vineland-II and CBCL), and sleep behavior (CSHQ) to evaluate the ability of the EQ-5D-3L and SF-6D to capture differences in caregiver HR-QOL in relation to measures of child health and behavioral problems. In this analysis, the criteria for clinical validity was met if higher health utility scores measured by the EQ-5D-3L and SF-6D were associated with lower levels of autism severity (ADOS scores) and maladaptive behavior (Vineland-II and CBCL), better sleep habits (CSHQ), and higher levels of child IQ and HR-QOL (PedsQL™ and HUI-3).

It was expected that caregivers who were caring for children with higher ASD severity or worse behavioral problems would have lower scores on both preference-weighted instruments, suggesting spillover effects of caring for a child with ASD [44]. It was also expected that caregivers of children who were younger in age would have lower HR-QOL, given the increased burden of caring for younger children with ASD [42, 43].

3.2.3 Discriminative Power of EQ-5D-3L and SF-6D

Discriminative power quantifies whether an instrument is sensitive to differences in comparator outcomes across different response levels of the preference-weighted instruments [60, 61]. Discriminative power was tested by comparing the mean values of outcomes in children for each level of the EQ-5D-3L and SF-6D dimensions in caregivers using two-way *t* tests. To quantify differences in discriminative power between the SF-6D and the EQ-5D-3L, we compared the percentage of statistically significant *t* tests of the total number of *t* tests calculated for each of the two preference-weighted instruments. This differs from our analysis of clinical validity, which compared average caregiver health utility scores among different classifications of health- or behavior-related characteristics of the child or caregiver. Responses to each domain were categorized as having “no problems” related to the given domain or as having “at least some” of the indicated problem, with the exception of the vitality domain of the SF-6D, which was categorized as “no or some problems” and “moderate or severe problems” given the limited number of responses of “no problems” for this domain.

It was anticipated that caregivers who reported “no problems” on a given domain would be associated with better child outcomes, including higher health utility scores (PedsQL™ and HUI-3), better adaptive behavior scores (Vineland-II), and fewer behavioral and emotional problems (CBCL).

4 Results

The study sample contained 224 caregivers of children with ASD. The response rate for the completion of the postal survey components was 115 of 220 (52%) at one ATN site and 109 of 179 (61%) at the second ATN site. In our sample, 90% of caregivers were female, around 60% had at least a college education, and 76% were married (Table 1). The average caregiver age was 39.4 years, and the average child age was 8.4 years. Of the children included in the study sample, 87% were male (Table 1). A complete description of the demographic characteristics of the study sample can be found in Hoefman et al. [31].

4.1 SF-6D and EQ-5D-3L Scores

The EQ-5D-3L had a higher average health utility index score than the SF-6D (mean 0.847 vs. 0.741) and greater variation (standard deviation 0.139 vs. 0.119). EQ-5D-3L

Table 1 Characteristics of children with autism spectrum disorder and their caregivers, *n* = 224^a

Characteristic	Child ^b	Caregiver ^b
Demographic characteristics		
Age (years)	8.4 (3.5)	39.4 (8.3)
Gender		
Male	86.6	10.5
Female	13.5	89.5
IQ	75.7 (24.3)	
DSM-IV diagnosis		
Autism	73.4	
Asperger's syndrome	8.7	
PDD-NOS	17.9	
ADOS severity	7.2 (1.8)	
Educational level		
≤ High school		11.4
Some college		27.3
College		30.9
Professional/graduate degree		30.5
Married		75.6
Quality of life measures		
PedsQL™ total score	63.0 (15.8)	
HUI-3 index	0.66 (0.23)	
EQ-5D-3L		0.85 (0.14)
SF-6D		0.74 (0.12)
Other Health and Behavioral Characteristics		
Vineland-II composite	68.4 (11.1)	
CBCL total score	63.7 (9.4)	
CSHQ	49.1 (8.3)	
Nightly hours of sleep		
≤ 5		24.9
6–7		59.7
≥ 8		15.4
CES-D	13.8 (10.7)	
CarerQoL-7D	75.5 (21.3)	
FQLS	100.9 (16.1)	

Higher scores on the ADOS severity, CBCL, and CES-D indicate worse problems; higher scores on the PedsQL™, HUI-3, Vineland-II, CSHQ, CarerQoL-7D, and FQLS indicate better functioning/health, SF-6D Short Form-6 Dimension

ADOS Autism Diagnostic Observation Schedule, CarerQoL-7D Care-related Quality of Life instrument, CBCL Child Behavior Checklist, CES-D Center for Epidemiologic Studies Depression Scale, CSHQ Children's Sleep Habits Questionnaire, DSM-IV Diagnostic and Statistical Manual of Mental Disorders 4th Edition, EQ-5D-3L three-level EuroQoL-5 Dimension, FQLS Family Quality of Life Survey, HUI-3 Health Utilities Index Mark 3, PDD-NOS Pervasive Developmental Disorder Not Otherwise Specified, PedsQL™ Pediatric Quality of Life Inventory™, Vineland-II Vineland Adaptive Behavior Scales Second Edition

^aThe analysis includes 224 caregivers and 224 children. Each characteristic was calculated to include the maximum number of respondents for that characteristic. The number of included individuals ranged from 203 to 221 for caregivers and from 187 to 224 children

^bMean and standard deviation are given for each continuous variable. Percentage is given for each categorical variable

Table 2 Convergent validity: Spearman's ρ correlation between preference-weighted instruments and other caregiver measures for caregivers of children with autism spectrum disorder, $n=224^a$

Caregiver measure	Caregiver health-related quality of life measure	
	SF-6D index	EQ-5D-3L index
SF-6D index		0.617***
EQ-5D-3L index	0.617***	
CarerQoL-7D	0.569***	0.541***
CES-D	-0.742***	-0.501***

Spearman's ρ : 0.10–0.29 = weak, 0.30–0.49 = moderate, and >0.50 = strong. Higher scores on the CES-D indicate worse problems; higher scores on the CarerQoL-7D indicate better care-related quality of life

CarerQoL-7D Care-related Quality of Life instrument, *CES-D* Center for Epidemiologic Studies Depression Scale, *EQ-5D-3L* three-level EuroQoL-5 Dimension, *SF-6D* Short Form-6 Dimension

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

^aSpearman's ρ correlations included the maximum number of observations per correlation, ranging from 197 to 213

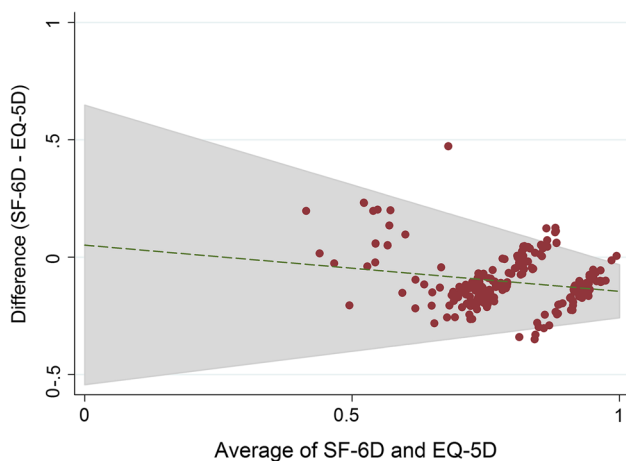


Fig. 1 Bland Altman plot demonstrating the relationship between the difference between each health utility score with the average between the two scores for each parent (Combination Art, created in STATA® and edited in Adobe Photoshop). *EQ-5D* EuroQoL-5 Dimension, *SF-6D* Short Form-6 Dimension

scores ranged from 0.308 to 1.000, and SF-6D scores ranged from 0.378 to 1.000.

4.2 Validity of the EQ-5D-3L and SF-6D in Measuring Health-Related Quality of Life of Caregivers

4.2.1 Convergent Validity

Health utility index scores of the EQ-5D-3L and the SF-6D were strongly correlated ($\rho = 0.617$, $p < 0.001$) (Table 2).

The Bland–Altman plot (Fig. 1) illustrates that the difference in a caregiver's EQ-5D-3L and SF-6D scores decreased as the average of the two scores increased, suggesting a higher level of agreement (i.e., more similar scores) for caregivers with better health than in caregivers with poorer health.

4.2.2 Clinical Validity

Both instruments demonstrated clinical validity with respect to caregiver HR-QOL. As expected, health utility scores differed significantly among caregivers with fewer hours of sleep per night, more depressive symptoms, lower caregiver-related quality of life, and lower family-related quality of life, indicating that both the SF-6D and EQ-5D-3L were sensitive to differences among caregivers (Table 3).

4.3 Ability of the SF-6D and EQ-5D-3L to Measure Spillover Effects in Caregivers

4.3.1 Discriminative Power

Caregivers who responded as having “no problems” on all six of the SF-6D domains and on two of the EQ-5D-3L domains (usual activities and anxiety/depression) had children with higher quality of life measured by the PedsQL™ (Table 4). For example, caregivers who reported “at least some” mental health problems on the SF-6D or “at least some” anxiety/depression problems the EQ-5D-3L had children with lower average HR-QOL (PedsQL™) than parents with “no problems” (61.7 vs. 69.8, $p = 0.01$, and 59.5 vs. 66.3, $p < 0.001$, respectively). Overall, there was a greater percentage of significant t tests for the SF-6D (63%) than for the EQ-5D-3L (25%). There was only one domain (SF-6D role limitations) with significant differences in child adaptive behavior scores (Vineland-II) (Table 4).

4.3.2 Clinical Validity

Results related to clinical validity favored the SF-6D. For caregivers of children with increased behavior problems (indicated by higher CBCL or lower Vineland-II scores) or with lower quality of life (indicated by lower PedsQL™ or HUI-3 scores), the SF-6D captured significantly different caregiver HR-QOL scores for two of the child measures (PedsQL™ and CBCL). In addition, the EQ-5D-3L and SF-6D both captured significantly different caregiver quality of life for one of the child measures (HUI-3) (Table 5). For example, caregivers with a child whose HUI-3 score was below the sample average of 0.659 (indicating worse child HR-QOL) had a significantly lower SF-6D score (0.712 vs. 0.762, $p = 0.003$) and a significantly lower EQ-5D-3L score (0.819 vs. 0.867, $p = 0.013$) than caregivers with a child whose HUI-3 score was above the sample average. Although

Table 3 Clinical validity: one-way analysis of variance comparing mean EQ-5D-3L and SF-6D health utility index scores of caregivers with different demographic characteristics, $n = 224$

Caregiver measure	n (%) ^a	Caregiver quality of life measures [mean (SD)]	
		SF-6D index	EQ-5D-3L index
Nightly hours of sleep			
≤5	55 (25)	0.698 (0.13)*	0.777 (0.16)*
6–7	132 (60)	0.749 (0.11)*	0.861 (0.13)*
≥8	34 (15)	0.780 (0.11)*	0.907 (0.10)*
CES-D			
≥13.8	94 (46)	0.660 (0.08)*	0.778 (0.14)*
<13.8	109 (54)	0.812 (0.09)*	0.912 (0.10)*
CarerQoL-7D			
<75.5	78 (36)	0.676 (0.10)*	0.770 (0.14)*
≥75.5	137 (64)	0.779 (0.11)*	0.891 (0.12)*
FQLS			
<100.9	80 (39)	0.689 (0.11)*	0.801 (0.13)*
≥100.9	124 (61)	0.774 (0.11)*	0.881 (0.13)*

Higher scores on the CES-D indicate worse problems; higher scores on the FQLS and CarerQoL-7D indicate better quality of life

CarerQoL-7D Care-related Quality of Life instrument, CES-D Center for Epidemiologic Studies Depression Scale, EQ-5D-3L three-level EuroQoL-5 Dimension, FQLS Family Quality of Life Scale, SF-6D Short Form-6 Dimension

* $p < 0.05$ using Bonferroni correction factor such that groups with two, three, or four levels must exhibit a p value of 0.0500, 0.0167, and 0.0083, respectively

^a n values may not total to 224 due to missing responses

the average SF-6D and EQ-5D-3L scores differed, the difference between health utility scores for caregivers of children with above- or below-average HR-QOL (HUI-3) were of similar magnitude for the SF-6D and EQ-5D-3L. There was not a significant difference in HR-QOL among caregivers of children with an above- or below-average age, IQ, or autism severity (ADOS) score or for children with different autism diagnoses using either caregiver preference-weighted instrument (Table 5).

5 Discussion

The Second US Panel on Cost-Effectiveness Analysis in Health and Medicine and other government agencies around the world have emphasized that research on valuing spillover effects in family and caregivers is warranted [9–12]. While spillover effects can also be studied in a broader context, increasing knowledge of the health effects among caregivers and family members in QALY terms is highly relevant and consistent with a societal perspective [4] and a healthcare perspective [8]. A number of approaches have been adopted to value spillover effects in QALY terms, including direct and indirect elicitation [62–65]. In the context of clinical trials and other intervention studies, indirect elicitation techniques are likely to be favored, as is the case with measuring patient QALYs, and require guidance about appropriate

instruments consistent with the large literature devoted to identifying the most appropriate instrument for patient conditions. Despite the obvious appeal of indirect elicitation techniques for capturing spillover QALYs, research on comparing different instruments is lacking.

The need for guidance on different approaches for measuring spillover effects in QALY terms is especially important given the recent change in recommendations by the Second US Panel. Traditionally, spillover effects were included in cost-effectiveness analyses using monetary costs, often measured by the additional time devoted by a caregiver to caring for the patient. Inclusion of non-monetary values, or QALYs of caregivers and other family members, along with monetary costs raised concerns about double-counting [65]. Indeed, a recent review of methods for valuing informal care offered guidance for including spillover effects in either monetary or non-monetary terms because of issues with double counting and other concerns [63]. The Second US Panel now recommends the inclusion of both monetary and non-monetary spillover effects in economic evaluations [10].

This study compared the EQ-5D-3L and SF-6D with respect to their ability to capture spillover effects in caregivers of a child with ASD. To compare the instruments, we first assessed whether they would provide similar results for similar caregivers. In particular, if the two instruments were correlated with each other and with

Table 4 Discriminative power: *t* tests comparing the mean scores of child outcomes (PedsQL™, HUI-3, Vineland-II, CBCL) for domain responses on caregiver preference-weighted instruments, *n* = 224

Caregiver quality of life measures	<i>n</i> (%) ^a	Child outcomes [mean (SD)]			
		PedsQL™ ^b	HUI-3 ^b	Vineland-II ^b	CBCL ^b
SF-6D domains					
Physical functioning					
No problems	193 (87)	64.2 (15.6)*	0.664 (0.24)	68.6 (11.5)	63.3 (9.3)
At least some problems	28 (13)	55.3 (15.5)*	0.616 (0.20)	66.8 (7.9)	67.2 (9.6)
Role limitations					
No limitations	112 (51)	66.5 (16.4)*	0.699 (0.20)*	70.4 (10.9)*	62.3 (10.0)*
At least some limitations	106 (49)	59.6 (14.6)*	0.613 (0.26)*	66.2 (11.3)*	65.3 (8.7)*
Social functioning					
No problems	91 (42)	67.3 (16.1)*	0.691 (0.23)	68.9 (11.4)	61.4 (9.7)*
At least some problems	128 (58)	59.8 (14.7)*	0.632 (0.24)	68.0 (11.1)	65.8 (8.8)*
Pain					
No pain	109 (50)	66.9 (16.4)*	0.693 (0.23)*	69.4 (11.2)	62.2 (9.6)*
At least some pain	111 (50)	59.3 (14.3)*	0.621 (0.24)*	67.4 (11.1)	65.5 (9.07)*
Mental health					
No problems	35 (16)	69.8 (13.4)*	0.754 (0.18)*	69.5 (9.8)	59.3 (9.5)*
At least some problems	183 (84)	61.7 (15.8)*	0.641 (0.24)*	68.2 (11.5)	64.6 (9.2)*
Vitality					
No or some problems	76 (35)	69.4 (15.6)*	0.697 (0.23)	69.6 (12.1)	59.4 (10.4)*
Moderate or severe problems	143 (65)	60.0 (14.8)*	0.636 (0.23)	67.7 (10.7)	66.1 (8.0)*
EQ-5D-3L domains					
Mobility					
No problems	205 (93)	63.2 (15.6)	0.657 (0.24)	68.3 (11.3)	63.7 (9.6)
At least some problems	15 (7)	61.0 (16.8)	0.666 (0.18)	68.9 (10.0)	64.3 (7.8)
Self-care					
No problems	215 (98)	63.0 (15.5)	0.657 (0.24)	68.3 (11.2)	63.9 (9.4)
At least some problems	5 (2)	63.0 (25.4)	0.701 (0.16)	75.0 (14.0)	59.0 (12.1)
Usual activities					
No problems	184 (84)	64.3 (15.7)*	0.673 (0.23)*	68.7 (11.3)	63.5 (9.4)
At least some problems	36 (16)	55.8 (14.1)*	0.581 (0.24)*	66.7 (10.8)	65.7 (9.5)
Pain/discomfort					
No problems	121 (55)	64.5 (15.8)	0.670 (0.24)	68.6 (11.2)	63.7 (10.1)
At least some problems	99 (45)	61.3 (15.4)	0.643 (0.23)	67.9 (11.2)	63.9 (8.7)
Anxiety/depression					
No problems	107 (49)	66.3 (16.1)*	0.718 (0.21)*	70.0 (11.3)	61.4 (10.3)*
At least some problems	112 (51)	59.5 (14.5)*	0.601 (0.25)*	66.8 (10.9)	66.1 (7.9)*

Higher scores on the CBCL indicate worse problems; higher scores on the PedsQL™, HUI-3, and Vineland-II indicate better functioning/health

CBCL Child Behavior Checklist, EQ-5D-3L three-level EuroQoL-5 Dimension, HUI-3 Health Utilities Index Mark 3, PedsQL™ Pediatric Quality of Life Inventory™, SF-6D Short Form-6 Dimension, Vineland-II Vineland Adaptive Behavior Scales Second Edition

**p* < 0.05 using a Bonferroni correction factor such that groups with two, three, or four levels must exhibit a *p* value of 0.0500, 0.0167, and 0.0083, respectively

^a*n* values may not total to 224 due to missing responses

^bCaregiver-reported child measure

other measures of caregiving quality of life or health, it would suggest the measures were valid instruments. Both measures demonstrated convergent validity as they were strongly correlated with each other, the CarerQoL, and the CES-D. While the SF-6D exhibited a stronger correlation

with the CES-D, the EQ-5D-3L exceeded criteria for a strong correlation.

Second, we assessed whether the instruments would provide similar results in relation to the characteristics of the child with ASD. In particular, scores on measures of child

Table 5 Clinical validity: one-way analyses of variance comparing mean EQ-5D-3L and SF-6D health utility index scores of caregivers with children who have different demographic characteristics, $n = 224$

Child measure	n (%) ^a	Caregiver quality of life measures [mean (SD)]	
		SF-6D index	EQ-5D-3L index
Child age (years) ^b			
< 8.4	135 (60)	0.736 (0.12)	0.848 (0.15)
≥ 8.4	89 (40)	0.749 (0.12)	0.846 (0.12)
Child IQ ^b			
Low IQ (≤ 70)	105 (53)	0.739 (0.12)	0.848 (0.14)
High IQ (> 70)	92 (47)	0.740 (0.11)	0.853 (0.12)
Child DSM-IV diagnosis ^b			
Autism	160 (73)	0.747 (0.12)	0.855 (0.13)
Asperger's syndrome	19 (9)	0.694 (0.13)	0.779 (0.16)
PDD-NOS	39 (18)	0.741 (0.11)	0.854 (0.14)
Child ADOS severity ^b			
≥ 7.2	88 (43)	0.742 (0.12)	0.843 (0.12)
< 7.2	117 (57)	0.745 (0.11)	0.863 (0.14)
Child PedsQL™ total score ^c			
< 63.0	91 (47)	0.705 (0.11)*	0.833 (0.13)
≥ 63.0	103 (53)	0.770 (0.11)*	0.870 (0.13)
Child HUI-3 index ^c			
< 0.659	90 (41)	0.712 (0.12)*	0.819 (0.13)*
≥ 0.659	128 (59)	0.762 (0.11)*	0.867 (0.14)*
Child Vineland-II composite ^c			
< 68.4	100 (52)	0.736 (0.12)	0.852 (0.12)
≥ 68.4	94 (48)	0.748 (0.11)	0.865 (0.13)
Child CBCL total score ^c			
≥ 63.7	93 (50)	0.710 (0.12)*	0.837 (0.13)
< 63.7	94 (50)	0.771 (0.12)*	0.871 (0.12)
Child CSHQ ^c			
≤ 49.1	113 (56)	0.777 (0.11)*	0.887 (0.11)*
> 49.1	89 (44)	0.698 (0.11)*	0.808 (0.15)*

Higher scores on the ADOS severity and CBCL indicate worse problems; higher scores on the PedsQL™, HUI-3, Vineland-II, and CSHQ indicate better functioning/health

ADOS Autism Diagnostic Observation Schedule, CBCL Child Behavior Checklist, CSHQ Children's Sleep Habits Questionnaire, DSM-IV Diagnostic and Statistical Manual of Mental Disorders 4th Edition, EQ-5D-3L three-level EuroQoL-5 Dimension, HUI-3 Health Utilities Index Mark 3, PDD-NOS Pervasive Developmental Disorder Not Otherwise Specified, PedsQL™ Pediatric Quality of Life Inventory™, SF-6D Short Form-6 Dimension, Vineland-II Vineland Adaptive Behavior Scales Second Edition

* $p < 0.05$ a using Bonferroni correction factor such that groups with two, three, or four levels must exhibit a p -value of 0.0500, 0.0167, and 0.0083, respectively

^a n values may not total to 224 due to missing responses

^bRegistry data

^cCaregiver-reported child measure

health were compared in relation to the top and bottom of the distributions for the two instruments as well as the difference in instrument scores in response to changes in the child health measures. Significant differences in child health scores in relation to the distributions of the two instruments demonstrates discriminative power while differences in instrument scores in relation to differences in child health measures demonstrates clinical validity. Both instruments demonstrated discriminative ability; however, the SF-6D had a greater percentage of significant findings than the EQ-5D-3L. With respect to clinical validity, the SF-6D similarly performed slightly better on the measures of child health, with significant differences in average health utility scores relative to scores on four of the five child measures (PedsQL™, HUI-3, CBCL, and CSHQ) compared with significant differences for two child measures (HUI-3 and CSHQ) for the EQ-5D-3L. Neither measure was associated with the child's age, IQ, autism severity, or diagnosis. Significant differences in average scores across the child health measures, but not the child's age, IQ, or autism diagnosis, indicates that it is differences in child health and behavior that drive spillover effects among caregivers.

Based on the comparison of the two instruments in this study, some guidance can be offered for those interested in developing clinical studies to measure caregiver spillover effects associated with caring for a child with autism. Either the SF-6D or the EQ-5D-3L are likely to capture health effects among caregivers in QALY terms for interventions or changes in the clinical characteristics of children with autism that are associated with measurable health effects for the child. Interventions such as new molecules for the treatment of behavior problems that produce meaningful changes in the CBCL are likely to have spillover effects for the caregiver that can be captured by standard preference-weighted instruments such as the SF-6D or the EQ-5D-3L, and we recommend their inclusion in clinical trials and other research designs that can identify causal effects.

Researchers such as Hoefman et al. [48] suggest that the effects of caregiving on caregivers can be measured with the same preference-weighted instruments used to measure HR-QOL in patients. Surprisingly, few studies have assessed preference-weighted instruments to determine whether they are sensitive or responsive for measuring caregiver or family spillover effects. Given that regulatory agencies recommend indirect elicitation with preference-weighted instruments to measure patient QALYs [9–12], more research appears warranted to compare preference-weighted instruments for measuring spillover effects in other contexts, such as adult children caring for their parents, and other conditions in children, including somatic and mental health conditions.

Several limitations to the study should be noted. First, we limited the caregiver quality-of-life and health measures to two previously validated instruments: the CarerQoL-7D

and CES-D. It can be argued that the CarerQoL-7D captures a different construct (burden of caregiving) than the health utility instruments (HR-QOL) and the CES-D is limited to mental health problems. Still, both of these measures should be correlated with health utility measures as greater caregiver burden translates into worse HR-QOL. This was the case in this study and, more importantly, the comparison of the EQ-5D-3L and the SF-6D demonstrated strong correlations with our measures of caregiver burden. The information produced with caregiver-specific instruments such as the CarerQoL-7D may be more appropriate in evaluations of interventions targeted at caregivers specifically given that the CarerQoL-7D measures the impact of caregiving beyond health effects [31, 48].

Second, we relied on caregiver self-reports regarding health states for themselves and their children. This approach may lead to problems of endogeneity, especially in study designs where treatment effects cannot be identified. Alternative designs, where the child is rated by a family member other than the primary caregiver, may provide an indication of the extent to which caregivers project their own health states onto the rating of their children [63]. Clinical studies based on exogenous instruments, such as randomization or disease states, are likely to limit problems with endogeneity and can identify spillover effects using indirect elicitation techniques. In addition, there have been considerable methodological advances in using direct elicitation techniques to measure spillover effects [64]. Direct elicitation techniques may be particularly sensitive to a given population and expanding research on direct elicitation techniques to include caregivers of children with ASD could supplement the findings from our study.

Third, the comparisons were based on the EQ-5D-3L, which has only three response levels per construct, rather than the EQ-5D-5L which has five response levels. The EQ-5D-5L may have increased validity and discriminative power [66] and has been suggested to have greater responsiveness than SF-6D in other caregiver contexts [19]. Our finding that both the SF-6D and EQ-5D-3L can be used to capture spillover effects for interventions involving children with autism remains and likely translates to the use of the EQ-5D-5L. Finally, we limited this investigation to health-related spillover effects in primary caregivers. Broader investigations, including observing effects in other family members and effects beyond health in an extra-welfarist context, remain important as well [8].

6 Conclusion

Capturing spillover effects in cost-effectiveness analyses is necessary to ensure accurate valuations of healthcare interventions and programs. Our comparison of the SF-6D

and EQ-5D-3L health utility instruments indicated that both can capture health-related spillover effects in terms of health among caregivers of children with ASD, although the SF-6D had stronger discriminative power and clinical validity in this context. The findings provide useful information for researchers and practitioners interested in developing protocols for measuring caregiver spillover effects in QALY terms. It is feasible to use indirect assessment in clinical studies to measure caregiver spillover effects associated with interventions to improve the health of children with autism. Regulatory agencies recommend indirect assessment to measure *patient* QALYs, and this study provides evidence for recommending a similar approach to capture *caregiver* spillover effects associated with child health conditions such as ASD.

Author contributions CCB planned the statistical analysis and wrote the initial draft of the manuscript. She edited the revised versions of the manuscript. JMT served as the Principal Investigator for funding the data collection, collaborated in the statistical analysis, and helped draft and edit the manuscript. NP served as a Co-Investigator on the grant that funded the data collection. She collaborated on the statistical analysis and the drafting and editing of the manuscript. DKW assisted with the statistical analysis and drafting of the manuscript. KAK served as a Principal Investigator on the grant that funded the data collection. She collaborated on the drafting and editing of the manuscript. JMP was a Co-Investigator on the grant that funded the data collection. He collaborated on drafting and editing the manuscript. RJH collaborated on the interpretation of the study findings and assisted with drafting and editing the manuscript. WBFB served as a consultant on the grant that funded data collection. He assisted with drafting and editing the manuscript.

Compliance with Ethical Standards

Funding The project was funded by a grant from the National Institute of Mental Health (R01MH089466) with JMT and KAK serving as Principal Investigators. NP acknowledges support from the National Institute of Mental Health (R03MH102495). This Autism Treatment Network was supported by Autism Speaks and cooperative agreement UA3 MC11054 through the US Department of Health and Human Services, Health Resources and Services Administration, Maternal and Child Health Research Program to the Massachusetts General Hospital.

Conflict of interest Clare C. Brown, J. Mick Tilford, Nalin Payakachat, D. Keith Williams, Karen A. Kuhlthau, Jeffrey M. Pyne, Renske J. Hoefman, and Werner B.F. Brouwer have no conflicts of interest to disclose.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. For this type of study, formal consent is not required. This article does not contain any studies with animals performed by any of the authors.

Informed consent Informed consent was obtained from all individual participants included in the study.

Data availability The datasets utilized and/or analyzed during the current study are available from the corresponding author on reasonable request.

Open Access This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Neumann PJ, Goldie SJ, Weinstein MC. Preference-based measures in economic evaluation in health care. *Annu Rev Public Health*. 2000;21:587–611.
2. Lipscomb J, Drummond M, Fryback D, Gold M, Revicki D. Retaining, and enhancing, the QALY. *Value Health*. 2009;12(Suppl 1):S18–26.
3. Krol M, Papenburg J, van Exel J. Does including informal care in economic evaluations matter? A systematic review of inclusion and impact of informal care in cost-effectiveness studies. *Pharmacoeconomics*. 2015;32(2):123–35.
4. Basu A, Meltzer D. Implications of spillover effects within the family for medical cost-effectiveness analysis. *J Health Econ*. 2005;24(4):751–73.
5. Al-Janabi H, Flynn TN, Coast J. QALYs and carers. *Pharmacoeconomics*. 2011;29(12):1015–23.
6. Wittenberg E, Ritter GA, Prosser LA. Evidence of spillover of illness among household members: EQ-5D scores from a US sample. *Med Decis Mak*. 2013;33(2):235–43.
7. Tilford JM, Payakachat N. Progress in measuring family spillover effects for economic evaluations. *Expert Rev Pharmacoecon Outcomes Res*. 2015;15(2):195–8.
8. Al-Janabi H, Van Exel J, Brouwer W, Coast J. A framework for including family health spillovers in economic evaluation. *Med Decis Mak*. 2016;36(2):176–86.
9. Sanders GD, Neumann PJ, Basu A, Brock DW, Feeny D, Krahn M, et al. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: second panel on cost-effectiveness in health and medicine. *JAMA*. 2016;316(10):1093–103.
10. Neumann P, Sanders G, Russell L, Siegel J, Ganiats T, editors. *Cost-effectiveness in health and medicine*. 2nd ed. New York: Oxford University Press; 2016.
11. The National Institute for Health and Care Excellence. *Guide to the methods of technology appraisal 2013*. London: The National Institute for Health and Care Excellence; 2013.
12. Brazier J, Longworth L. NICE DSU Technical Support Document 8: an introduction to the measurement and valuation of health for NICE submissions. London: NICE Decision Support Unit; 2011.
13. Neumann PJ, Kuntz KM, Leon J, Araki SS, Hermann RC, Hsu M, et al. Health utilities in Alzheimer's disease: a cross-sectional study of patients and caregivers. *Med Care*. 1999;37(1):27–32.
14. Bell CM, Araki SS, Neumann PJ. The association between caregiver burden and caregiver health-related quality of life in Alzheimer disease. *Alzheimer Dis Assoc Disord*. 2001;15(3):129–36.
15. 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):873–84.
16. Akehurst RL, Brazier JE, Mathers N, O'Keefe C, Kaltenthaler E, Morgan A, et al. Health-related quality of life and cost impact of irritable bowel syndrome in a UK primary care setting. *Pharmacoeconomics*. 2002;20(7):455–62.
17. Clayton DJ, Wild DJ, Quarterman P, Duprat-Lomon I, Kubin M, Coons SJ. A comparative review of health-related quality-of-life measures for use in HIV/AIDS clinical trials. *Pharmacoeconomics*. 2006;24(8):751–65.
18. Payakachat N, Tilford JM, Brouwer WB, van Exel NJ, Grosse SD. Measuring health and well-being effects in family caregivers of children with craniofacial malformations. *Qual Life Res*. 2011;20(9):1487–95.
19. Bhadhuri A, Jowett S, Jolly K, Al-Janabi H. A comparison of the validity and responsiveness of the EQ-5D-5L and SF-6D for measuring health spillovers: a study of the family impact of meningitis. *Med Decis Mak*. 2017;37(8):882–93.
20. Brouwer WBF, van Exel NJA, Tilford MJ. Incorporating caregiver and family effects in economic evaluations of child health. In: Ungar WJ, editor. *Economic evaluation in child health*. New York: Oxford University Press; 2009. p. 55–76.
21. Meltzer DO, Smith PC. Theoretical issues relevant to the economic evaluation of health technologies. In: Pauly MV, McGuire TG, Barros PP, editors. *Handbook of health economics*. North Holland: Oxford; 2012. p. 433–70.
22. Lavelle TA, Wittenberg E, Lamarand K, Prosser LA. Variation in the spillover effects of illness on parents, spouses, and children of the chronically ill. *Appl Health Econ Health Policy*. 2014;12(2):117–24.
23. Magliano L, Fiorillo A, De Rosa C, Malangone C, Maj M, National Mental Health Project Working Group. Family burden in long-term diseases: a comparative study in schizophrenia vs. physical disorders. *Soc Sci Med*. 2005;61(2):313–22.
24. Magliano L, Fiorillo A, Malangone C, De Rosa C, Maj M, National Mental Health Project Working Group. Social network in long-term diseases: a comparative study in relatives of persons with schizophrenia and physical illnesses versus a sample from the general population. *Soc Sci Med*. 2006;62(6):1392–402.
25. Simonoff E, Pickles A, Charman T, Chandler S, Loucas T, Baird G. Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. *J Am Acad Child Adolesc Psychiatry*. 2008;47(8):921–9.
26. Khanna R, Jariwala K, Bentley JP. Psychometric properties of the EuroQol Five Dimensional Questionnaire (EQ-5D-3L) in caregivers of autistic children. *Qual Life Res*. 2013;22(10):2909–20.
27. Khanna R, Jariwala K, Bentley JP. Health utility assessment using EQ-5D among caregivers of children with autism. *Value Health*. 2013;16(5):778–88.
28. Brooks R, EuroQol Group. EuroQol: the current state of play. *Health Policy*. 1996;37(1):53–72.
29. Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Med Care*. 2004;42(9):851–9.
30. American Psychiatric Association, American Psychiatric Association. *Diagnostic and statistical manual-text revision (DSM-IV-TR, 2000)*; 2000.
31. Hoefman R, Payakachat N, van Exel J, Kuhlthau K, Kovacs E, Pyne J, et al. Caring for a child with autism spectrum disorder and parents' quality of life: application of the CarerQol. *J Autism Dev Disord*. 2014;44(8):1933–45.
32. Tilford JM, Payakachat N, Kovacs E, Pyne JM, Brouwer W, Nick TG, et al. Preference-based health-related quality-of-life outcomes in children with autism spectrum disorders. *Pharmacoeconomics*. 2012;30(8):661–79.

33. Gotham K, Pickles A, Lord C. Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *J Autism Dev Disord.* 2009;39(5):693–705.
34. Sparrow SS, Cicchetti DV, Balla DA. *Vineland-II adaptive behavior scales.* 2nd ed. Circle Pines: AGS Publishing; 2005.
35. Roid GH. *Stanford–Binet intelligence scales.* 5th ed. Rolling Meadows: Riverside Publishing; 2003.
36. Mullen E. *Mullen scales of early learning.* Los Angeles: Western Psychological Services; 1997.
37. Bayley N, Reuner G. *Bayley scales of infant and toddler development: Bayley-III.* San Antonio: Harcourt Assessment, Psychological Corporation; 2006.
38. Duarte CS, Bordin IA, de Oliveira A, Bird H. The CBCL and the identification of children with autism and related conditions in Brazil: pilot findings. *J Autism Dev Disord.* 2003;33(6):703–7.
39. Owens JA, Spirito A, McGuinn M. The children's sleep habits questionnaire (CSHQ): psychometric properties of a survey instrument for school-aged children. *Sleep.* 2000;23(8):1043–51.
40. Varni JW, Seid M, Kurtin PS. PedsQL™ 4.0: reliability and validity of the Pediatric Quality of Life Inventory™ version 4.0 generic core scales in healthy and patient populations. *Med Care.* 2001;39(8):800–12.
41. Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, DePauw S, et al. Multiattribute and single-attribute utility functions for the Health Utilities Index Mark 3 system. *Med Care.* 2002;40(2):113–28.
42. Esbensen AJ, Seltzer MM, Lam KS, Bodfish JW. Age-related differences in restricted repetitive behaviors in autism spectrum disorders. *J Autism Dev Disord.* 2009;39(1):57–66.
43. Shattuck PT, Seltzer MM, Greenberg JS, Orsmond GI, Bolt D, Kring S, et al. Change in autism symptoms and maladaptive behaviors in adolescents and adults with an autism spectrum disorder. *J Autism Dev Disord.* 2007;37(9):1735–47.
44. Tilford JM, Payakachat N, Kuhlthau KA, Pyne JM, Kovacs E, Bellando J, et al. Treatment for sleep problems in children with autism and caregiver spillover effects. *J Autism Dev Disord.* 2015;45(11):3613–23.
45. Khanna R, Madhavan SS, Smith MJ, Patrick JH, Tworek C, Becker-Cottrill B. Assessment of health-related quality of life among primary caregivers of children with autism spectrum disorders. *J Autism Dev Disord.* 2011;41(9):1214–27.
46. Radloff LS. The CES-D scale a self-report depression scale for research in the general population. *Appl Psych Meas.* 1977;1(3):385–401.
47. Brouwer W, Van Exel N, Van Gorp B, Redekop W. The CarerQol instrument: a new instrument to measure care-related quality of life of informal caregivers for use in economic evaluations. *Qual Life Res.* 2006;15(6):1005–21.
48. Hoefman RJ, van Exel J, Brouwer WBF. Measuring care-related quality of life of caregivers for use in economic evaluations: carerqol tariffs for Australia, Germany, Sweden, UK, and US. *Pharmacoeconomics.* 2017;35(4):469–78.
49. Hoffman L, Marquis J, Poston D, Summers JA, Turnbull A. Assessing family outcomes: psychometric evaluation of the beach center Family Quality Of Life Scale. *J Marriage Fam.* 2006;68(4):1069–83.
50. Dolan P. The measurement of health-related quality of life for use in resource allocation decisions in health care. In: Culyer A, Newhouse J, editors. *Handbook of health economics, vol. 1B.* North Holland: University of York and Harvard University of Medical School; 2000. p. 1724–48.
51. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Med Care.* 2005;43(3):203–20.
52. Craig BM, Pickard AS, Stolk E, Brazier JE. US valuation of the SF-6D. *Med Decis Mak.* 2013;33(6):793–803.
53. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull.* 1959;56(2):81–105.
54. Cohen J. A power primer. *Psychol Bull.* 1992;112(1):155–9.
55. Heintz E, Wiréhn A, Peebo BB, Rosenqvist U, Levin L. 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):475–84.
56. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *Statistician.* 1983;32:307–17.
57. Petrou S, Hockley C. An investigation into the empirical validity of the EQ-5D and SF-6D based on hypothetical preferences in a general population. *Health Econ.* 2005;14(11):1169–89.
58. Hoefman RJ, van Exel J, Brouwer WB. Measuring the impact of caregiving on informal carers: a construct validation study of the CarerQol instrument. *Health Qual Life Out.* 2013;11(173):1–13.
59. Sprangers MA, Groenvold M, Arraras JI, Franklin J, te Velde A, Muller M, et al. The European Organization for Research and Treatment of Cancer breast cancer-specific quality-of-life questionnaire module: first results from a three-country field study. *J Clin Oncol.* 1996;14(10):2756–68.
60. Luo N, Wang P, Fu AZ, Johnson JA, Coons SJ. Preference-based SF-6D scores derived from the SF-36 and SF-12 have different discriminative power in a population health survey. *Med Care.* 2012;50(7):627–32.
61. 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.
62. Basu A, Dale W, Elstein A, Meltzer D. A time tradeoff method for eliciting partner's quality of life due to patient's health states in prostate cancer. *Med Decis Mak.* 2010;30(3):355–65.
63. Wittenberg E, Prosser LA. Disutility of illness for caregivers and families: a systematic review of the literature. *Pharmacoeconomics.* 2013;31(6):489–500.
64. Prosser LA, Lamarand K, Gebremariam A, Wittenberg E. Measuring family HRQoL spillover effects using direct health utility assessment. *Med Decis Mak.* 2015;35(1):81–93.
65. Weatherly H, Faria R, Van Den Berg B. Valuing informal care for economic evaluation. In: Culyer AJ, editor. *Encyclopedia of health economics, vol. 1.* San Diego: Elsevier; 2014. p. 459–67.
66. Janssen M, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. *Qual Life Res.* 2013;22(7):1717–27.