Published in final edited form as:

Autism Res. 2014 December ; 7(6): 649–663. doi:10.1002/aur.1409.

Predicting Health Utilities for Children with Autism Spectrum Disorders

Nalin Payakachat¹, J. Mick Tilford^{1,2,3}, Karen A. Kuhlthau⁴, N. Job van Exel⁵, Erica Kovacs⁶, Jayne Bellando³, Jeffrey M. Pyne^{7,8}, and Werner BF Brouwer⁵

¹Division of Pharmaceutical Evaluation and Policy, University of Arkansas for Medical Sciences, Little Rock, AR ²Department of Health Policy and Management, University of Arkansas for Medical Sciences, Little Rock, AR ³Department of Pediatrics, University of Arkansas for Medical Sciences, Little Rock, AR ⁴Department of Pediatrics, Harvard Medical School, Boston, MA and Center for Adolescent Health Policy, Massachusetts General Hospital, Boston, MA ⁵Department of Health Policy and Management, Erasmus University, Rotterdam, The Netherlands

⁶Department of Psychiatry, Columbia University Medical Center, New York, NY ⁷Center for Mental Healthcare and Outcomes Research, Central Arkansas Veterans Healthcare System, Little Rock, AR ⁸Psychiatric Research Institute, University of Arkansas for Medical Sciences, Little Rock, AR

Scientific Abstract

Comparative effectiveness of interventions for children with ASDs that incorporates costs is lacking due to the scarcity of information on health utility scores or preference-weighted outcomes typically used for calculating quality-adjusted life years (QALYs). This study created algorithms for mapping clinical and behavioral measures for children with ASDs to health utility scores. The algorithms could be useful for estimating the value of different interventions and treatments used in the care of children with ASDs. Participants were recruited from two Autism Treatment Network sites. Health utility data based on the HUI3 for the child was obtained from the primary caregiver (proxy-reported) through a survey (N=224). During the initial clinic visit, proxy-reported measures of the Child Behavior Checklist, Vineland II Adaptive Behavior Scales, and the PedsQL 4.0 (start measures) were obtained and then merged with the survey data. Nine mapping algorithms were developed using the HUI3 scores as dependent variables in ordinary least squares regressions along with the start measures, the Autism Diagnostic Observation Schedule to measure severity, child age, and cognitive ability as independent predictors. In-sample cross-validation was conducted to evaluate predictive accuracy. Multiple imputation techniques were used for missing data. The average age for children with ASDs in this study was 8.4 (SD=3.5) years. Almost half of the children (47%) had cognitive impairment (IQ≤70). Total scores for all of the outcome measures were significantly associated with the HUI3 score. The algorithms can be applied to clinical studies containing start measures of children with ASDs to predict QALYs gained from interventions.

Keywords

mapping; predictive algorithms; equating measure; autism; health utilities; clinical measure; behavioral measure; quality of life measure

Introduction

The prevalence of children with autism spectrum disorders (ASDs) has increased significantly with recent estimates suggesting 1 in 88 to 1 in 50 children are now affected (Centers for Disease Control and Prevention 2012; Blumberg et al. 2013). ASDs are neurodevelopmental disorders beginning in childhood and affecting outcomes into adulthood (Howlin et al. 2004). ASDs are characterized, in varying degrees, by social interaction difficulties, communication challenges, and a tendency to engage in repetitive behaviors (Lord & McGee J 2001; Lord & Bishop 2010). With the publication of the DSM-5, the three categories of ASD symptoms (social difficulties, communication impairments, and repetitive/restricted behaviors) have been merged into two categories (social-communication impairment and repetitive/restricted behaviors) and the former diagnostic subtypes have been merged into one umbrella diagnosis of ASD (American Psychiatric Association. 2013).

While ASD is usually a lifelong condition, both children and adults benefit from interventions that reduce ASD symptoms and improve skills and abilities (such as language, cognitive and adaptive skills) (Howlin et al. 2009; Warren et al. 2011; Dawson et al. 2010). Because of the increased prevalence of children with ASDs, resources for understanding the comparative effectiveness of alternative interventions has increased commensurate with the burden of the condition on the population. Understanding the value of interventions for children with ASDs, however, has the potential to translate clinical comparative effectiveness research findings into sustained practice (Glasgow et al. 2012; Glasgow et al. 2013; Glasgow & Steiner 2012).

For example, there is clear evidence that children with ASDs benefit from intensive behavioral services (Rothenberg & Samson 2009; Dawson et al. 2010; Warren et al. 2011). The National Research Council guidelines for treating children with autism recommend providing 25 hours a week of intensive therapy (Lord & McGee J 2001; Myers et al. 2007; Scottish Intercollegiate Guidelines Network (SIGN) 2007; The National Autism Center 2009; Lord & Bishop 2010). Yet, payers raise concerns about the costs of intensive behavioral interventions and who should pay for them (Rothenberg & Samson 2009). Information on the value of intensive behavioral services based on formal cost-effectiveness analyses could inform guidelines for providing services to children with ASDs, but are lacking. Indeed, implementation scientists recognize that one of the “greatest opportunities” to translate clinical comparative effectiveness evidence into practice or policy is an understanding of costs in relation to outcomes because it provides a clear rationale for decision-makers to act (Glasgow et al. 2013).

The lack of information on cost-effectiveness seems partly due to the scarcity of information in this population on health utility scores (alternatively referred to as preference-weighted

outcomes), which are necessary for calculating quality-adjusted life years (QALYs). The QALY is a measure that combines both health-related quality and quantity of life into a single index (Neumann et al. 2000). Economic evaluations that use the cost per QALY metric theoretically permit comparison of interventions for ASD with other physical and mental health interventions in both child and adult populations (Weinstein MC 1996; Neumann et al. 2000; Neumann & Greenberg 2009; Greenberg & Neumann 2011).

To generate QALY estimates for economic evaluation, researchers need evidence on health utilities. Typically, utilities are expressed on a scale ranging from 0 (the value attached to the state 'dead') to 1 (representing the value of the state of full health). Negative health utility scores are possible since there may be health states people consider to be worse than dead (e.g., vegetative states). However, health utility scores cannot exceed 1 as there is no health state better than perfect health (Gold et al. 1996). Health utility scores can be estimated directly using methods to express the value of a health state like a visual analog scale, standard gamble, or a time trade-off method (Brazier et al. 2007). Often, however, they are estimated indirectly using existing multi-attribute health classification systems (Brazier et al. 2007) such as the HUI3 (Feeny et al. 2002), QWB-SA (Seiber et al. 2008), EQ-5D (The EuroQol Group 1990), or SF-6D (Brazier & Roberts 2004). These systems are based on instruments to classify health states described by a number of health domains with predefined health utility scores that enable translating all possible health states measured within the instrument into a health utility score. The basic idea behind these instruments is that they can be applied generally across conditions and populations to facilitate standardized comparisons, although much has been written about the comparability of scores generated by different instruments (Lipscomb et al. 2009). To date, only a few studies have estimated health utilities for children with ASDs (Petrou & Kupek 2009; Petrou et al. 2010; Tilford et al. 2012) and no study has estimated health utility gains associated with specific treatments such as intensive behavioral services. Thus, formal cost-effectiveness of services for children with ASDs according to the cost per QALY metric as recommended by guidelines for conducting such evaluations in both the US and the UK (Weinstein MC 1996; National Institute for Health and Clinical Excellence (NICE) 2008; The National Institute for Health and Care Excellence 2013) are lacking despite their potential to inform discussions of value. Following Bailey (2009), we believe more discussion of the value of services for children with ASDs appear warranted (Bailey 2009).

Because estimates of health utility gains associated with interventions for children with ASDs are central to describing cost-effectiveness, and data are lacking, there is a need to generate estimates of health utility gains with new methods. In the absence of health utility scores in clinical studies, determining value using cost-effectiveness analysis can be accomplished with an indirect approach, which is called "mapping" (Longworth & Rowen 2011). Mapping can be used to predict health utilities based on clinical or behavioral data to estimate QALY gains (Longworth & Rowen 2011). Algorithms for mapping clinical outcomes in terms of health utilities have been reported for a number of conditions (Payakachat et al. 2009; Versteegh et al. 2012; Longworth et al. 2005; Crott & Briggs 2010; Goldsmith et al. 2010; Dakin et al. 2010) but no such research has been developed specifically to predict health utilities for children with ASDs.

Mapping is defined as the development and use of an algorithm to predict health utility scores using data on other indicators or measures of health (Longworth & Rowen 2011). In the absence of direct elicitation of health utility scores in clinical trials, a mathematical mapping approach that explains relationships between health utility scores and clinical and behavioral measures or other quality of life measures is a useful (although necessarily second-best) alternative (National Institute for Health and Clinical Excellence (NICE) 2008; The National Institute for Health and Care Excellence 2013). Predictive algorithms from the mapping approach can also help identify the aspects of behavioral functioning among children with ASDs that particularly impact health utility scores and the magnitude of this impact. The models can then be applied to data from existing clinical trials or other studies containing the predictive measures necessary to predict health utility scores. The predicted health utility scores can then be linked back to data collected within the original study and can be used to calculate QALYs for use in economic evaluations of treatments or interventions.

This study provides algorithms to map clinical, behavioral, and health-related quality of life (HRQL) measures including the Child Behavior Checklist (CBCL), Vineland-II Adaptive Behavior Scales (Vineland-II), and Pediatric Quality of Life Inventory (PedsQL) version 4.0 that are typically used in clinical studies of ASD treatments and interventions onto a health utility measure (the Health Utilities Index Mark 3, HUI3) using statistical association. The predictive algorithms provided can be applied to clinical data generated in comparative effectiveness evaluations of ASD treatments to generate health utility scores. The findings thus could be helpful in estimating value in terms of cost-effectiveness of different ASD treatments, and to compare estimates with treatments for other populations.

Methods

Sample and Data Collection—The sampling frame for this study was children aged 4 – 17 years old diagnosed with ASDs and their parents who participated in two Autism Treatment Network sites (a developmental center in Little Rock, Arkansas and an outpatient psychiatric clinic at Columbia University Medical Center in New York). Families of diagnosed children who had been seen at the clinics and previously agreed to be contacted about future studies as part of their participation in the Autism Treatment Network registry were identified as potential study subjects. A survey package was sent by mail to eligible families with a self-addressed envelope for return. Families that returned the survey, signed Health Insurance Portability and Accountability Act forms, and consent/assent forms were provided a \$25 gift card for participating in the research. Approximately 10% of the families in the Little Rock clinic and 5% of the families in the Columbia clinic did not agree to be contacted about future research studies. The study protocol was approved by the institutional review boards at Columbia University/New York State Psychiatric Institute (NYSPI) and the University of Arkansas for Medical Sciences.

The survey package contained instruments to measure health utility scores for the child with ASD and the primary caregiver. An information sheet specified that the primary caregiver of the child should complete the survey marked “to be completed by the caregiver about the child with the ASD.” Data from the survey was then linked to the clinical information of

children with ASDs obtained from the Autism Treatment Network to create the final data file.

Outcome Measures—There are several generic, preference-based HQRL instruments that can be used to produce health utility scores, however, not all of them have been used nor are designed specifically for children (Payakachat et al. 2012). The HUI3 (Feeny et al. 2002) was selected to serve as our primary outcome for measuring health utilities. We decided to use the primary caregiver as a proxy to report health for their child with ASD because the comprehension level for the HUI3 was fairly high for children with ASDs, more importantly, a significant proportion of children with ASDs has cognitive disability, in addition to developmental disability. The HUI3 has been used previously in studies involving children in the same age ranges with our current study (Petrou & Kupek 2009; Buysse CP 2008; Prince et al. 2010). Previous research showed that HUI3 was able to differentiate health status between children with ASDs and unaffected children. At the time this survey research was conducted, the EQ-5Y (EQ-5D Youth version) (Wille et al. 2010) and CHU9D (Child Health Utility 9D) (Stevens 2010) were not available for public use.

The HUI3 measures responses on eight domains of health including vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain. Not all of the health domains may be affected by ASDs, since the instrument is intended to be generic, i.e. to be able to capture health utility associated with different conditions, in order to be able to produce comparable health utility scores in a variety of diseases. The HUI3 includes a multi-attribute scoring algorithm associated with the domains to generate a utility score based on subject responses. The resulting health utility scores range from -0.36 (worst possible health) to 1.0 (best possible health). Subject responses on the eight domains of HUI3 were used to calculate HUI3 utility scores using a multiplicative, multi-attribute utility function (Feeny et al. 2002). Differences or changes on HUI3 utility scores of 0.03 are considered to be clinically important (Luo et al. 2010).

Detailed information on patient and family demographics, patient diagnosis, medical history, physical and neurological exam, clinical and behavioral measures, quality of life measures, and experience with care were obtained from the Autism Treatment Network registry. From the available measures, we selected relevant outcomes that could be mapped onto the HUI3 health utility scores. The clinical and behavioral outcome measures included the CBCL, Vineland-II, and PedsQL version 4.0, all of which were administered at the Autism Treatment Network clinic baseline visit and continuing through follow-up. As many children had both baseline and follow-up data, all outcomes used for analysis were chosen based on the closest time to survey completion.

Child Behavior Checklist

The Child Behavior Checklist (CBCL) is a standardized behavioral inventory that covers a wide range of psychiatric symptoms and undesirable behaviors and is based on parent report (Achenbach & Ruffle 2000). The CBCL assesses both internalizing (i.e., anxious, depressive, and overcontrolled) and externalizing (i.e., aggressive, hyperactive, noncompliant, and undercontrolled) behavior symptoms. The CBCL total problems score is

a combined measure of externalizing and internalizing problems. CBCL scores are expressed as T-scores (mean of 50, SD=10. T-scores of 60 to 69 are considered to be in a borderline clinical range while T-scores of 70 or above reflect clinically significant behavioral problems in comparison to same age/gender peers. The CBCL has been shown to be valid and reliable in children with ASDs (Dekker et al. 2002; Wallander et al. 2006).

Vineland-II Adaptive Behavior Scales

The Vineland-II, is a semi-structured, clinician-administered interview conducted with a caregiver that is used to assess the performance of everyday activities required for personal and social sufficiency across a broad range of conditions (Sparrow et al. 2005). The Vineland-II assesses overall adaptive functioning as well as adaptive behavior in specific domains of communication, socialization, daily living skills, and motor skills (the latter for fine and gross movements in children aged < 7 years). The adaptive composite and domain scores are expressed as age-based standard score (mean of 100, SD=15), with higher scores reflecting better adaptive functioning. The cut-off point of 70 is considered to reflect functioning within a low adaptive level.

Pediatric Quality of Life Inventory

The PedsQL version 4.0 Generic Module is designed for use in children aged 2-18 years and has 23 items grouped into four domains: physical, emotional, social, and school functioning (Varni et al. 2001). Each item is converted into a 0 – 100 scale, with higher scores indicating better quality of life. The PedsQL total score is the unweighted average of all item scores. Domain scores are calculated by averaging all items in that domain. This instrument is able to differentiate quality of life outcomes between children with ASDs and unaffected children (Kuhlthau et al. 2010). A cut-off point for at-risk status for impaired HRQL (parent proxy-report) is 65.4 (Varni et al. 2003).

Statistical Analyses—The HUI3 health utility score was used as the “target” measure. The CBCL, Vineland-II, and PedsQL were used as the “start” measures. We also included child's age, cognitive ability (IQ), and ADOS calibrated severity scores as other predictor variables in each model. Spearman's correlations were performed to examine the strength of the associations between the target and each of the start measures.

The aim of the statistical analysis is to establish mapping algorithms to predict HUI3 health utility scores from different start variables and other control variables. Because health utility scores in general have non-normal distributions with negative skew, as well as ceiling effects (a large spike at the upper bound) that violate the assumptions of ordinary least squares (OLS) regression, various statistical modeling approaches have been considered in previous studies (Brazier et al. 2010; Longworth & Rowen 2011). A number of models including the censored least absolute deviation (CLAD) model, the Tobit model, generalized linear models (GLM), latent class models, two-part or two-step models (TPM), and a random effects censored mixture model have been evaluated. A recent simulation study found that OLS is still superior to many of the alternative approaches including CLAD, Tobit, TPMs, and LCMs (Pullenayegum et al. 2010). The beta regression approach was also used in comparison to OLS to address the challenges of the presence of a large spike at 1 in

the health utility score distribution (Basu & Manca 2012). The results indicated that the beta regression is useful when covariate effects are large and there are large spikes at the upper bound of the distribution. OLS, however, still provides unbiased estimates when covariate effects are less than 0.03 following a standard deviation change in the independent variable and there are small spikes at the upper end of the distribution.

The distribution of HUI3 scores in our sample is negatively skewed (skewness of -0.9) but only a small percentage of respondent HUI3 scores reach the ceiling of 1 (4.1%, Figure 1). Thus, the distribution of HUI3 scores for this study does not correspond to the distribution patterns that benefit from beta regression (Basu & Manca 2012). For this reason, we decided to use OLS for all mapping algorithms in this study.

Model Specifications

Seven models were specified in this study based on either domain scores or a total score of each instrument as a start measure (Table 1). The rationale for model specification was based on finding start measures that are normally reported as domain or total scores in ASD research. Mapping algorithms could then be created to produce health utility scores based on the availability of start measures (CBCL, Vineland-II, PedsQL) in clinical studies. Two models and resulting algorithms were created for the CBCL and PedsQL while three models were generated for the Vineland-II. The first CBCL model used the internalizing and externalizing problem T scores, which are the two subdomains of the CBCL. The second CBCL model was estimated only with the total problem T scores, which is the total composite score of the CBCL. The first PedsQL model consisted of all four domains, while the second PedsQL model included only the total score. The first Vineland-II model comprised all four Vineland-II domain scores (Communication, Daily Living Skills, Socialization, and Motor Skills) while the second Vineland-II model had only three domain scores (Communication, Daily Living Skills, and Socialization) because the Motor Skills domain is used only for children < 7 years old. The third Vineland-II model was estimated with only the adaptive composite score. Two additional models were proposed (Combined Models 1 and 2, Table 1) that included all three start measures (the total scores from CBCL, Vineland-II, and PedsQL).

Clinical information on children with ASDs for model building in the models included the Autism Diagnostic Observation Schedule (ADOS) severity score and cognitive ability scores. The ADOS, a clinician administered interview, considered the “gold standard” observational assessment for diagnosing ASD. The ADOS contains four assessment modules designed to elicit behaviors directly relevant to the diagnosis of ASD at different developmental levels and chronological ages (Lord et al. 2000). The ADOS Severity Score is a calibrated metric used to facilitate comparison of scores across the four different ADOS modules and offers a method to quantify relative severity of ASD symptoms that is independent of age and language level (Gotham et al. 2009; Shumway et al. 2012). ADOS severity scores range from 1 to 10, with scores of 1 – 3 representing nonspectrum classification, 4 – 5 corresponding to an “autism spectrum” classification, and 6 – 10 indicating an “autism” classification on the ADOS. Cognitive ability tests used in the Autism Treatment Network registry included the Stanford-Binet Intelligence Scales, 5th

Edition, Abbreviated Battery (SB-5), an individually administered, standardized cognitive assessment that can be used with individuals age 2 years and older. However, the Mullen Scales of Early Learning, American Guidance Service Edition (Mullen 1997) or the Bayley Scales of Infant Development, 3rd Edition (Bayley 2005) were also used in the Autism Treatment Network registry to assess cognitive functioning when the child could not obtain a basal score on the SB-5 during the initial baseline evaluation. In our sample, the majority of children were assessed using the SB-5 (84.3%). The SB5 Nonverbal and Verbal IQ and the other two measures of cognitive functioning are comparable for the purposes of the Autism Treatment Network registry as well as within our study, with correlations ranging from 0.78 to 0.84 (Roid 2003). All three cognitive measures are in comparison to age based norms and provide a standard score with a mean of 100 and standard deviation of 15 to describe a child's cognitive ability. A standard score less than 70 indicates impaired cognitive functioning.

The ADOS severity score and cognitive ability tests were administered only at the baseline visit. ADOS severity scores were added in each model to control for the severity of the condition across the different modules of the ADOS instrument. Child age (both linear and squared terms) and logarithm of IQ were included in the model. No interaction terms were included in any models in order to keep the models as simple as possible.

We recognize that different combinations of models could be specified to reach the most predictive model. However, the ability to use such a model in real world settings may be limited if the clinical research project did not have all of the various start measures. Thus, we report the results of parsimonious models that can be used by researchers with only limited start measures. Models that are more complicated, and thus more predictive, may be considered in future research studies or on an ad hoc basis.

Model specification was tested using the Ramsey RESET test (Ramsey 1969) and Link test (Pregibon 1980). The OLS estimates were generated with robust standard errors to address heteroscedasticity (Long & Ervin 2000). All statistical significance levels were set at a *p*-value of 0.05.

Model Comparisons

We conducted an in-sample, cross-validation analysis to assess predictive accuracy of the mapping algorithms using a *k*-fold technique where the dataset was randomly partitioned into *k* subsamples (*k*=5) with 1,000 replications (Kohavi 1995). One subsample was retained as the validation data for testing the predictive accuracy of the model and the remaining four subsamples were used as training data. The process was then repeated five times with each of the five subsamples used exactly once as the validation data. If any health utility scores were predicted to be greater than 1.0, they were then truncated at 1.0 to remain consistent with the health utility scale bounds.

The model predictive accuracy was determined using the individual mean absolute prediction error (MAPE) and root mean squared error (RMSE), which provided the deviation between the predicted and observed health utility scores (Longworth & Rowen 2011). The average MAPE and RMSE from the 5-fold, in-sample cross-validation analysis

were then calculated. The mean error for an OLS model is usually near zero due to the estimation technique. Therefore, we did not include the mean errors in the results. Error patterns across the scale of the health utility scores were reported to provide detailed information on how each mapping algorithm performed across the range of scores. Finally, the intraclass correlation coefficients (ICC), a measure of agreement between predicted and observed HUI3 scores, were calculated and reported for each algorithm using a two-way mixed model analysis of variance.

Missing Data

A multiple imputation approach was used to handle missing data. Missing observations were replaced with a set of plausible values that account for the uncertainty about the right value to impute using PROC MI. The imputed data sets are then analyzed for each OLS model. The results from each analysis were then combined using PROC MIANALYZE. Both procedures were conducted in SAS software (SAS 9.3, SAS Institute Inc.).

Results

224 surveys were returned from 2 Autism Treatment Network sites (a response rate of 54.6%). Demographic characteristics of the sample are presented in Table 2. The average age for children with ASDs in this study was 8.4 (SD=3.5) years. The sample was predominantly Caucasian children (75.2%) relative to African American (9.5%) and Hispanic (9%) children. Most children were in preschool or kindergarten (45%) or elementary school (33%). 98.6% of children lived at home with the primary caregivers being biological parents. Almost half of the children (47%) had cognitive impairment, scoring within the 2nd percentile rank or below (cognitive ability score (IQ) $\leq$ 70). The average IQ was 75.7 (SD=24.3), which was positively correlated with the HUI3 scores ($\rho=0.34$, $p<0.001$). The ADOS severity score ranged from 2 to 10 with a mean of 7.2 (SD=1.8). The severity score had mild, but significant correlation with HUI3 scores ($\rho=-0.14$, $p=0.04$).

Average health utility scores as derived from the HUI3 for the full sample of children with ASDs was 0.66 (SD=0.23). Children with autistic disorder had the lowest health utility scores on the HUI3 (mean of 0.63, SD=0.24). The average HUI3 scores were the highest in the children with Asperger's Disorder (mean of 0.77, SD=0.15). The common problem of a spike in scores at 1 (perfect health) was not apparent with the HUI3 scores as only 9 observations (4.1%) reached the upper bound of the distribution (Figure 1).

Outcome measures from the CBCL, Vineland-II, and PedsQL instruments are presented in Table 3. The average of the CBCL total problem T scores was 63.7 (SD=9.4), which was higher than norm populations (mean of 50). 31% of affected children had the CBCL total problem T score $\geq$ 70, which is a cut-off point for clinically significant behavioral problems. The CBCL total problem T scores were negatively associated with the HUI3 scores as expected ($\rho=-0.22$, $p=0.002$); the more behavior problems, the lower the health utility scores. Examination of the adaptive behavior domains indicated the lowest scores were achieved on the socialization domain (average score=68.1, SD=11.3). The Vineland-II adaptive composite scores averaged 68.4 (SD=11.1) and 58% had scores $\leq$ 70 (low adaptive behavior skills). The Vineland-II composite scores were positively correlated with the HUI3

health utility score ($\rho=0.45$, $p<0.001$). The PedsQL 4.0 also had the lowest scores on the social functioning (average score=49.5, SD=24.0) domain, followed by school functioning (average score=63.1, SD=20.3). The total PedsQL score averaged 63.0 (SD=15.7) and 51% had impaired HRQL (score = 64.5). All PedsQL domain and total scores were positively associated with the HUI3 scores ($p<0.05$). Overall, the utility measures had weak to moderate correlations with the outcome or “start” measures.

The Ramsey RESET and Link tests did not indicate a problem with model specification as p -values for both tests were greater than 0.05 in all models. Missing data patterns appeared to be arbitrary. Percentage of missing data ranges from 2.7 (HUI3 scores) to 16.5% (the CBCL and PedsQL Social domain), Table 3. We specified five imputation sets for the multiple imputation process since the amount of missing data was less than 20%, to achieve at least 95% relative efficiencies (SAS Institute Inc 2011). A Markov chain Monte Carlo method with the expectation-maximization (EM) algorithm was used to impute missing values before entering into the 5-fold in-sample, cross-validation procedure for each model. The EM algorithm (posterior mode) converged and relative efficiency ranged from 95.7 (the Vineland-II Motor Skills) to 99.7% (HUI3 and IQ). Model prediction errors (MAPE and RMSE) were reported in Appendix I. Model MAPEs ranged from 0.1591 (the PedsQL Model 1) to 0.1819 (the CBCL Model 1). Model RMSEs ranged from 0.2001 (the PedsQL Model 1) to 0.2308 (the CBCL Model 1). Agreement between the predicted and observed HUI3 scores was the lowest for the CBCL Model 1 (ICC=0.455) and highest for the Combined Model 1 (ICC=0.681), Appendix I. Absolute prediction errors (APEs) were also reported by size of error for each model (Appendix II). The Combined Model 2 and PedsQL Models 1 have the lowest percentage of APEs > 0.20 at 31.7% and 31.9%, respectively.

The in-sample cross-validation prediction errors (MAPEs) for the observed HUI3 score range were reported to examine degree of prediction errors in different health states (Appendix III). In our sample, only 7.8% ($n=17$) reported HUI3 scores lower than 0.25 and 10.1% ($n=22$) reported HUI3 score between 0.9 – 1.0 (Appendix III). All predicted models performed poorly at the lower end of the HUI3 scores ($HUI3 < 0.25$) and performed best at the HUI3 range of 0.5 – 0.9. Analysis of scatter plots between predicted and observed HUI3 scores indicated that there was a tendency for the OLS models to over-predict at the lower end and under-predict at the upper end of the HUI3 (results not shown).

Algorithms presented as mathematical equations to predict HUI3 scores for each start measure, as well as robust standard errors, are provided in Table 4. The CBCL models, the internalizing problems T score (Model 1) and the total problem T score (Model 2) were significantly associated with the HUI3 scores ($p=0.003$ and $p<0.001$, respectively). The communication and motor skill domains of the Vineland-II were significantly related to the HUI3 scores in the Vineland-II model 1 ($p=0.008$ and $p=0.049$, respectively). For the Vineland-II Model 2 which excludes the motor skill domain, the communication and daily living skill domains are significantly related with the HUI3 scores ($p=0.041$ and $p=0.016$, respectively). The Vineland-II composite score is also significantly associated with the HUI3 scores ($p<0.001$). The PedsQL model 1 found significant relationships between the physical and social functioning domains and the HUI3 ($p=0.001$ and $p=0.003$, respectively). The PedsQL total score was statistically significant in model 2 ($p<0.001$). The Vineland-II

composite score and PedsQL total score were statistically significantly associated with HUI3 in both Combined Models.

Discussion

The increased prevalence of autism has created renewed interest in ensuring that children with ASDs receive services to achieve optimal outcomes (Interagency Autism Coordinating Committee 2011). Evaluations of the effectiveness of services for children with ASDs typically include clinical, behavioral, and HRQL outcome measures that cannot be translated into the cost per QALY for informing health policy decision makers despite the potential for information on cost-effectiveness to translate effective treatments into sustained practice (Glasgow & Steiner 2012). Cost-effectiveness evaluations rely on QALYs as the recommended outcome measure because it can facilitate comparisons across different type of treatments or interventions, health conditions and patient groups (Weinstein MC 1996). The lack of trials and other clinical studies that prospectively measure QALYs has led to limited information on the cost-effectiveness of services for children with ASDs. Indeed, we are unaware of any studies that use the cost per QALY metric to evaluate services for children with ASDs.

This study seeks to increase the amount of information available on QALY gains associated with services for children with ASDs by creating mapping algorithms that can be used with clinical outcome measures. In this study, we combined information on outcome measures typically used in clinical trials of services for children with ASDs with health utility data as measured by the HUI3. With these data, the investigators developed several mapping algorithms that can be used to predict health utility scores based on clinical, behavioral, and HRQL outcome measures. Investigators can use the algorithms along with data from successful trials of alternative treatments for children with autism to predict the QALYs gained from the interventions studied. If investigators have only one of the “start” measures, they can choose an algorithm with the same subdomain scores used in their original studies to predict HUI3 scores based on the lowest MAPE (Appendix I). Similarly, when investigators have more than one of the “start” measures, they may choose an algorithm that has the lowest MAPE. The suggestions are based solely on authors' opinions. The QALY information could then be combined with cost data to calculate incremental cost effectiveness ratios for use in resource allocation decisions (Weinstein MC 1996). In short, investigators will have a greater opportunity to ask whether an effective intervention is “worth it” relative to the cost. Asking such a question has the potential to translate interventions into sustained practice if decision-makers can see that the cost-effectiveness profiles for autism services are similar to other medical or public health interventions or some services achieve similar outcomes at lower cost.

To illustrate, the investigators used information from the Vineland-II model 1 mapping algorithm derived in this paper (Table 4.2) to evaluate the QALYs gained from early intervention services: $\text{HUI3 utility scores} = -0.2367 + 0.0033 \times \text{communication} + 0.0032 \times \text{daily living skills} + 0.0002 \times \text{socialization} + 0.0033 \times \text{motor skills} + (-0.0050) \times \text{ADOS severity scores} + 0.0252 \times \log(\text{IQ})$. Findings from the Early Start Denver model (ESDM) showed improvements or less regression on the Vineland-II domains and clinical outcomes

(ADOS severity score and IQ) when compared to the standard intervention (assess-and-monitor, A/M) (Dawson et al. 2010). The changes in Vineland-II domains from the ESDM and A/M were the communication (+13.7 vs. -0.7), daily living skills (-6.1 vs. -14.5), socialization (-4.6 vs. -8.9), and motor skills (-9.9 vs. -23.1). The ADOS severity among the cohort who received the ESDM was improved (-0.2) when compared to the A/M group (+0.3). The IQ among children who received the ESDM increased 17.6 points when compared to 7.0 points in the A/M group. While the improved outcomes clearly suggest benefits associated with the intervention, it is unclear whether the ESDM is cost-effective relative to other medical and public health services. Application of the mapping algorithm suggests average QALY gains of 0.13 at 2 years post baseline. This summary measure of health gains points to large increases in HRQL that has a high probability of being a cost-effective investment of societal resources. While our findings are intended for illustrative purposes as we lack access to the data, the evidence clearly suggests a full-scale investigation of the cost-effectiveness of the ESDM for children with ASDs is warranted to provide evidence for health system decision-making.

Limitations

There are a number of limitations that need to be addressed in this study. The mapping algorithms provided in this study were modeled based on children from two locations with ADOS severity scores ranging from mild to severe. The average ADOS severity scores were not statistically significant different between the two Autism Treatment Network locations. Additionally, each Autism Treatment Network site had only 9% of the affected children who had ADOS severity scores ≥ 5 . Two thirds of the sample from both Autism Treatment Network sites had ADOS severity scores in the range of 6 and 8. Generalizability of these algorithms may be limited and additional research on applying them to data from clinical trials is needed. Researchers who desire to use these mapping algorithms in different populations should recognize these limitations. The sample size for this mapping exercise was adequate but nonetheless limited. Naturally, a larger sample size is always desirable, particularly if one would like to include additional variables or interaction terms in the model.

Although the mapping approach is gaining popularity, as it permits researchers to predict health utility scores when original studies did not include such a measure, mapping should be considered a second-best solution to directly collected health utility values (Longworth & Rowen 2011). Hence, the use of mapping algorithms will lead to increased uncertainty and error around the estimated health utility values. In this regard, using the mapping algorithms to predict HUI3 scores for cost-effectiveness analyses should incorporate sensitivity analyses to test the robustness of the QALY estimates (Chapman et al. 2004). Mapping algorithms also greatly expand the potential information set from prior comparative effectiveness research studies. It will take much longer to accumulate similar information based on primary data collection of health utility scores in prospective clinical studies.

The generic instruments for measuring health utilities are applicable for a wide range of health conditions and treatments, although there are special issues in applications involving children (Ungar 2010; Prosser et al. 2007; Grosse et al. 2010; Prosser et al. 2012;

Payakachat et al. 2012; Ungar 2007; Ungar 2011). In particular, evaluation of health utilities in young children (<5) is especially difficult as instruments are not designed for this age group and typically requires out-of-sample prediction. Most of the children in this sample were above 5 years old (87.9%). Importantly, we only map outcomes of the child onto health utilities. It is now recognized that interventions affecting children have a high probability of affecting the health of caregivers and other family members and these effects should be incorporated in economic evaluations (Meltzer & Smith 2012). However, mapping family ‘spillover’ effects for economic evaluation is beyond the scope of this paper.

The recognition that children represent a special population for economic evaluation has led to the creation of new utility measures (such as the Child Health Utility 9D (Stevens 2010) and the EQ-5D Youth (Wille et al. 2010)) that may improve prediction. We could not include these measures at the time this study began, as they were not available for public use. Initial evidence suggests sufficient construct validity with the HUI3 to warrant its use in mapping exercises predicting QALY gains (Horsman et al. 2003). However, it should also be pointed out that QALYs pertain to health-related utility only. One may consider conditions such as ASDs to have an impact on broader well-being also. Hence, using restrictive outcome measures could result in ignoring relevant outcomes. This needs to be investigated further, since cost-effectiveness, defined broadly, allows the use of broader outcome measures than QALYs (Brouwer et al. 2008). While this debate is directly relevant for how resources should be allocated to the treatment and prevention of autism, it remains beyond the scope of this study.

The predictive accuracy of the mapping algorithms was examined using the difference between predicted and observed values (i.e., MAPE and RMSE) from in-sample cross-validation to provide an indication of the size of the prediction errors. Unfortunately, we cannot report out-of-sample predictive accuracy because there is currently no other study in children with ASDs that contains HUI3 scores and the “start” measure. The predictive errors reported from our mapping exercises are comparable to those reported in the literature (Brazier et al. 2010; Longworth & Rowen 2011). However, the prediction errors are often larger for models mapping a condition-specific measure onto a generic utility score than mapping generic health outcome measures onto generic utility measures (Longworth & Rowen 2011). One reason may be the limited conceptual overlap between the start measures and the HUI3 utility measure. We are aware of only two studies that mapped generic preference-based scores from other measures used in child health conditions. Dakin (2010) mapped OM8-30 measure for children with otitis media to HUI2 and HUI3 (Dakin et al. 2010). Our algorithms performed relatively similar to Dakin's algorithms for the HUI3, with prediction errors in the same ranges. Khan (2014) mapped the PedsQL to EQ-5D in a relatively healthy population of school-age children (Khan et al. 2014). Khan's algorithms have similar prediction errors in terms of RMSEs when compared to ours, but their mean prediction errors are relatively smaller. However, Khan's algorithms are robust only to healthy children aged 11-15 years and EQ-5D scores ≥ 0.6 . We did not report explanatory power using R-squared since it is less informative for evaluating mapping performance relative to prediction errors (Brazier et al. 2010; Longworth & Rowen 2011).

When predicting the HUI3 scores using the mapping algorithms, it is important to be aware that predicted scores greater than 1.0 are possible. Since any predicted HUI3 scores above 1.0 is an artifact of the statistical model, researchers should truncate any predicted score of greater than 1.0 to 1.0 to be consistent with the original health utility scale. In addition, sensitivity analyses of point estimates mapped from any start measures should be incorporated into cost-effectiveness analyses (Briggs 2000). In probabilistic sensitivity analyses, robust standard errors of each start measure should be used. The pattern of errors in this study is similar to other mapping studies that used condition-specific start measures (Goldsmith et al. 2010; Payakachat et al. 2009). The algorithms do not perform well in the lower end of the health utility scores (i.e., $HUI3 < 0.25$). The problem of over-predicted health utility scores for patients in poor health was also reported in the literature (Brazier et al. 2010; Versteegh et al. 2012). A separate mapping algorithm to predict health utility for children with ASDs who are in poor health may reduce overprediction problem (Versteegh et al. 2010), but we feel that on average, the algorithms provided in this study will permit new estimates of the cost per QALY gained for services provided to children with ASDs. Our current sample contained only 17 children (7.8%) who had HUI3 scores of < 0.25 and only 2 (0.9%) out of the entire sample had HUI3 scores lower than 0. Future research is needed to test these mapping algorithms in different data sets containing health outcome measures for children with ASDs. Lastly, the mapping algorithms are generated from imputed data, which may influence the parameters of the estimated algorithms as well as underestimate underlying relationships between the HUI3 scores and the “start” measures.

Conclusions

The mapping algorithms provided in this study can be applied to clinical trials and other studies containing one of the start measures (CBCL, Vineland-II, PedsQL) for children with ASDs in order to predict health utility scores when direct elicitation of health utility scores were not possible. The predicted health utilities can be incorporated into economic evaluations in terms of QALYs. Evidence from economic evaluations can provide useful information to decision-makers, including healthcare providers, policy makers, and families as cost-effectiveness information has the potential to justify the provision of services that are under-provided.

Acknowledgments

The project was supported by a grant no. R01MH089466 from the National Institute of Mental Health with JMT and KAK serving as principal investigators and a grant no. R03MH102495 with NP as the principal investigator. The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Mental Health or the National Institutes of Health. The authors acknowledge the members of the Autism Treatment Network for use of the data. The data for the study were collected as part of the Autism Treatment Network, a program of Autism Speaks. Further support came from a cooperative agreement (UA3MC11054) from the U.S. Department of Health and Human Services, Health Resources and Services Administration, Maternal and Child Health Research Program, to the Massachusetts General Hospital. The work described in this article represents the independent efforts of the authors with no restrictions from the funding source or the Autism Treatment Network. None of the authors of this study reported a conflict of interest associated with the preparation of the manuscript. Maria Melguizo, Nupur Chowdhury, Rebecca Rieger and Latunja Sockwell provided excellent research assistance

Grant sponsor: The National Institute of Mental Health; Grant numbers: R01MH089466, R03MH102495

References

- Achenbach TM, Ruffle TM. The Child Behavior Checklist and related forms for assessing behavioral/emotional problems and competencies. *Pediatrics in Review*. 2000; 21:265–271. [PubMed: 10922023]
- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th2013. (DSM-V). Retrieved November, 12, 2013, from <http://www.dsm5.org/Pages/Default.aspx>
- Bailey AJ. Where are the autism economists? *Autism Research*. 2009; 2:245. [PubMed: 19844947]
- Basu A, Manca A. Regression Estimators for Generic Health-Related Quality of Life and Quality-Adjusted Life Years. *Medical Decision Making*. 2012; 32:56–69. [PubMed: 22009667]
- Bayley, N. Bayley scales of infant and toddler development. 3rd. San Antonio, TX: Pearson Education, Inc; 2005.
- Blumberg, SJ.; Bramlett, MD.; Kogan, MD.; Schieve, LA.; Jones, JR.; Lu, MC. Changes in prevalence of parent-reported autism spectrum disorder in school-aged U S children: 2007 to 2011-2012 (Rep No 65). Hyattsville, MD: National Center for Health Statistics, MD; 2013.
- Brazier, JE.; Ratcliffe, J.; Salomon, J.; Tsuchiya, A. Valuing health. In: Brazier, JE.; Ratcliffe, J.; Salomon, J.; Tsuchiya, A., editors. *Measuring and valuing health benefits for economic evaluation*. New York, NY: Oxford University Press Inc; 2007. p. 83-138.
- Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Medical Care*. 2004; 42:851–859. [PubMed: 15319610]
- Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *The European Journal of Health Economics*. 2010; 11:215–225. [PubMed: 19585162]
- Briggs AH. Handling Uncertainty in Cost-Effectiveness Models. *Pharmacoeconomics*. 2000; 17:479–500. [PubMed: 10977389]
- Brouwer WBF, Culyer AJ, van Exel NJ, Rutten FFH. Welfarism vs. extra-welfarism. *Journal of Health Economics*. 2008; 27:325–338. [PubMed: 18179835]
- Buyse CP, R H. Long-term health status in childhood survivors of meningococcal septic shock. *Archives of Pediatrics & Adolescent Medicine*. 2008; 162:1036–1041. [PubMed: 18981351]
- Centers for Disease Control and Prevention. Prevalence of autism spectrum disorders -Autism and Developmental Disabilities Monitoring Network, 14 Sites, United States, 2008. *MMWR Surveillance Summ*. 2012; 61:1–19.
- Chapman RH, Berger M, Weinstein MC, Weeks JC, Goldie S, Neumann PJ. When does quality-adjusting life-years matter in cost-effectiveness analysis? *Health Economics*. 2004; 13:429–436. [PubMed: 15127423]
- Crott R, Briggs A. Mapping the QLQ-C30 quality of life cancer questionnaire to EQ-5D patient preferences. *European Journal of Health Economics*. 2010; 11:427–434. [PubMed: 20473703]
- Dakin H, Petrou S, Haggard M, Benge S, Williamson I. Mapping analyses to estimate health utilities based on responses to the OM8-30 otitis media questionnaire. *Quality of Life Research*. 2010; 19:65–80.
- Dawson G, Rogers S, Munson J, Smith M, Winter J, Greenson J, et al. Randomized, controlled trial of an intervention for toddlers with autism: the Early Start Denver Model. *Pediatrics*. 2010; 125:e17–e23. [PubMed: 19948568]
- Dekker MC, Nunn R, Koot HM. Psychometric properties of the revised Developmental Behaviour Checklist scales in Dutch children with intellectual disability. *Journal of Intellectual Disability Research*. 2002; 46:61–75. [PubMed: 11851857]
- 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. *Medical Care*. 2002; 40:113–128. [PubMed: 11802084]
- Glasgow RE, Doria-Rose VP, Khoury MJ, Elzarrad M, Brown ML, Stange KC. Comparative Effectiveness Research in Cancer: What Has Been Funded and What Knowledge Gaps Remain? *Journal of the National Cancer Institute*. 2013; 105:766–773. [PubMed: 23578853]

- Glasgow, RE.; Steiner, JF. Comparative effectiveness research to accelerate translation: Recommendations for an emerging field of science. In: Brownson, RC.; Colditz, GA.; Proctor, EK., editors. Dissemination and implementation research in health. New York, NY: Oxford University Press; 2012. p. 72-92.
- Glasgow RE, Vinson C, Chambers D, Khoury MJ, Kaplan RM, Hunter C. National Institutes of Health Approaches to Dissemination and Implementation Science: Current and Future Directions. *American Journal of Public Health*. 2012; 102:1274–1281. [PubMed: 22594758]
- Gold, M.; Siegel, J.; Russell, L.; Weinstein, M. Cost effectiveness in health and medicine. New York, NY: Oxford University Press; 1996.
- Goldsmith K, Dyer M, Buxton M, Sharples L. Mapping of the EQ-5D index from clinical outcome measures and demographic variables in patients with coronary heart disease. *Health and Quality of Life Outcomes*. 2010; 8:54. [PubMed: 20525323]
- Gotham K, Pickles A, Lord C. Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorder*. 2009; 39:693–705.
- Greenberg D, Neumann PJ. Does adjusting for health-related quality of life matter in economic evaluations of cancer-related interventions? *Expert Review of Pharmacoeconomics & Outcomes Research*. 2011; 11:113–119. [PubMed: 21351863]
- Grosse SD, Prosser LA, Asakawa K, Feeny D. QALY weights for neurosensory impairments in pediatric economic evaluations: case studies and a critique. *Expert Review of Pharmacoeconomics & Outcomes Research*. 2010; 10:293–308. [PubMed: 20545594]
- Horsman J, Fulong W, Feeny D, Torrance G. The Health Utilities Index (HUI®): Concepts, Measurement Properties and Applications. *Health and Quality of Life Outcomes*. 2003; 1(54)
- Howlin P, Goode S, Hutton J, Rutter M. Adult outcome for children with autism. *Journal of Child Psychology and Psychiatry*. 2004; 45:212–229. [PubMed: 14982237]
- Howlin P, Magiati I, Charman T. Systematic review of early intensive behavioral interventions for children with autism. *American Journal on Intellectual and Developmental Disabilities*. 2009; 114:23–41. [PubMed: 19143460]
- Interagency Autism Coordinating Committee. 2011 IACC Strategic Plan for Autism Spectrum Disorder Research. 2011. Retrieved October 3, 2012, from http://iacc.hhs.gov/strategic-plan/2011/IACC_2011_Strategic_Plan.pdf
- Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence*. 1995; 2:1137–1143.
- Khan KA, Petrou S, Rivero-Arias O, Walters SJ, Boyle SE. Mapping EQ-5D utility scores from the PedsQL generic core scales. *Pharmacoeconomics*. 2014 Apr 9.10.1007/s40273-014-0153-y
- Kuhlthau K, Orlich F, Hall T, Sikora D, Kovacs E, Delahaye J, et al. Health-Related Quality of Life in Children with Autism Spectrum Disorders: Results from the Autism Treatment Network. *Journal of Autism and Developmental Disorders*. 2010; 40:721–729. [PubMed: 20033762]
- Lipscomb J, Drummond M, Fryback D, Gold M, Revicki D. Retaining, and enhancing, the QALY. *Value in Health*. 2009; 12(Suppl 1):S18–26. [PubMed: 19250127]
- Long S, Ervin L. Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician*. 2000; 54:217–224.
- Longworth L, Buxton M, Sculpher M, Smith D. Estimating utility data from clinical indicators for patients with stable angina. *European Journal of Health Economics*. 2005; 6:347–353. [PubMed: 16193322]
- Longworth, LJ.; Rowen, D. NICE DSU Technical Support Document 10: The use of mapping methods to estimate health state utility values. Sheffield, UK: Decision Support Unit, ScHARR, University of Sheffield; 2011.
- Lord, C.; McGee, J. Educating Children with Autism. Washington, DC: National Academies Press; 2001.
- Lord C, Risi S, Lambrecht L, Cook EH Jr, Leventhal BL, Dilavore PC, et al. The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*. 2000; 30:205–223. [PubMed: 11055457]

- Lord C, Bishop SL. Autism spectrum disorders: Diagnosis, prevalence, and services for children and families. Society for Research in Child Development. Sharing Child and Youth Development Knowledge. Social Policy Report. 2010; 24(2):1–25.
- Luo N, Johnson JA, Coons SJ. Using Instrument-Defined Health State Transitions to Estimate Minimally Important Differences for Four Preference-Based Health-Related Quality of Life Instruments. *Medical Care*. 2010; 48(4):365–371. [PubMed: 20355266]
- Meltzer, DO.; Smith, PC. Theoretical Issues Relevant to the Economic Evaluation of Health Technologies. In: Mark, VP.; McGuire, TG.; Barros, PP., editors. *Handbook of Health Economics*. Vol. 2. Elsevier Science Ltd: 2012. p. 433-469.
- Mullen, E. Mullen scales of early learning. Los Angeles, CA: Western Psychological Services; 1997.
- Myers SM, Johnson CP. the Council on Children With Disabilities. Management of Children With Autism Spectrum Disorders. *Pediatrics*. 2007; 120:1162–1182. [PubMed: 17967921]
- National Institute for Health and Clinical Excellence (NICE). Guide to the methods of technology appraisal. London: NICE; 2008.
- Neumann PJ, Goldie SJ, Weinstein MC. Preference-based measures in economic evaluation in health care. *Annual Review of Public Health*. 2000; 21:587–611.
- Neumann PJ, Greenberg D. Is the United States ready for QALYs? *Health Aff (Millwood)*. 2009; 28:1366–1371. [PubMed: 19738253]
- Payakachat N, Summers KH, Pleil AM, Murawski MM, Thomas J III, Jennings K, et al. Predicting EQ-5D utility scores from the 25-item National Eye Institute Vision Function Questionnaire (NEI-VFQ 25) in patients with age-related macular degeneration. *Quality of Life Research*. 2009; 18:801–813. [PubMed: 19543808]
- Payakachat N, Tilford JM, Kovacs E, Kuhlthau K. Autism spectrum disorders: a review of measures for clinical, health services and cost-effectiveness applications. *Expert Review of Pharmacoeconomics & Outcomes Research*. 2012; 12:485–503. [PubMed: 22971035]
- Petrou S, Johnson S, Wolke D, Hollis C, Kochhar P, Marlow N. Economic costs and preference-based health-related quality of life outcomes associated with childhood psychiatric disorders. *The British Journal of Psychiatry*. 2010; 197:395–404. [PubMed: 21037217]
- Petrou S, Kupek E. Estimating Preference-Based Health Utilities Index Mark 3 Utility Scores for Childhood Conditions in England and Scotland. *Medical Decision Making*. 2009; 29(3):291–303. [PubMed: 19264726]
- Pregibon D. Goodness of link tests for generalized linear models. *Applied Statistics*. 1980; 29:15–24.
- Prince FHM, Geerdink LM, Borsboom GJJM, Twilt M, van Rossum MAJ, Hoppenreijns EPAH, et al. Major improvements in health-related quality of life during the use of etanercept in patients with previously refractory juvenile idiopathic arthritis. *Annals of the Rheumatic Diseases*. 2010; 69:138–142. [PubMed: 19581280]
- Prosser LA, Grosse SD, Kemper AR, Tarini BA, Perrin JM. Decision analysis, economic evaluation, and newborn screening: challenges and opportunities. *Genetics in Medicine*. 2012; 14:703–712.
- Prosser LA, Hammitt JK, Keren R. Measuring health preferences for use in cost-utility and cost-benefit analyses of interventions in children: theoretical and methodological considerations. *Pharmacoeconomics*. 2007; 25:713–726. [PubMed: 17803331]
- Pullenayegum EM, Tarride JE, Xie F, Goeree R, Gerstein HC, O'Reilly D. Analysis of health utility data when some subjects attain the upper bound of 1: are Tobit and CLAD models appropriate? *Value in Health*. 2010; 13:487–494. [PubMed: 20230549]
- Ramsey J. Tests for specification error in classical linear least squares regression analysis. *Journal of the Royal Statistical Society Series B*. 1969; 31:350–371.
- Roid, G. Stanford binet intelligence scales. 5th. Itasca, IL: Riverside Publishing; 2003.
- Rothenberg B, Samson D. Early Intensive Behavioral Intervention based on Applied Behavior Analysis among Children with Autism Spectrum Disorders. Technology Evaluation Center Assessment Program. Executive Summary. 2009; 23:1–61.
- SAS Institute Inc. SAS/STAT 9.3 User's Guide. Cary, NC: SAS Institute Inc; 2011. The MI procedure; p. 4610
- Scottish Intercollegiate Guidelines Network (SIGN). A national clinical guideline (Rep No SIGN publication No 98). Edinburgh, Scotland: Scottish Intercollegiate Guidelines Network (SIN);

2007. Assessment, diagnosis and clinical interventions for children and young people with autism spectrum disorders.
- Seiber, WJ.; Groessl, EJ.; Ganiats, TG.; Kaplan, RM. Quality of Well Being Self-Administered (QWB-SA) Scale: User's Manual. San Diego, CA: Health Services Research Center, University of California, San Diego; 2008.
- Shumway S, Farmer C, Thurm A, Joseph L, Black D, Golden C. The ADOS Calibrated Severity Score: Relationship to Phenotypic Variables and Stability over Time. *Autism Research*. 2012; 5:267–276. [PubMed: 22628087]
- Sparrow, SS.; Cicchetti, DV.; Balla, DA. Vineland Adaptive Behavior Scales. Second. American Guidance Services, Inc; 2005.
- Stevens, K. Valuation of the Child Health Utility Index 9D (CHU9D). University of Sheffield; 2010.
- The EuroQol Group. EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy*. 1990; 16:199–208. [PubMed: 10109801]
- The National Autism Center. National Standards Report. Randolph, MA: National Autism Center; 2009.
- The National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. London, England: The National Institute for Health and Care Excellence; 2013.
- Tilford JM, Payakachat N, Kovacs E, Pyne JM, Brouwer W, Nick T, et al. Preference-based health-related quality of life outcomes in children with autism spectrum disorders: A comparison of generic instruments. *Pharmacoeconomics*. 2012; 30:1–19. [PubMed: 22201520]
- Ungar WJ. Paediatric health economic evaluations: a world view. *Healthcare Quarterly*. 2007; 10:134–135. [PubMed: 17326379]
- Ungar, WJ. Economic evaluation in child health. New York, NY: Oxford Publisher Press; 2010.
- Ungar WJ. Challenges in Health State Valuation in Paediatric Economic Evaluation: Are QALYs Contraindicated? *Pharmacoeconomics*. 2011; 29:641–652. [PubMed: 21604821]
- Varni JW, Burwinkle TM, Seid M, Skarr D. The PedsQLTM 4.0 as a pediatric population health measure: Feasibility, reliability, and validity. *Ambulatory Pediatrics*. 2003; 3:329–341. [PubMed: 14616041]
- Varni JW, Seid M, Kurtin PS. The PedsQL 4.0: Reliability and validity of the Pediatric Quality of Life InventoryTM Version 4.0 Generic Core Scales in healthy and patient populations. *Medical Care*. 2001; 39:800–812. [PubMed: 11468499]
- Versteegh MM, Leunis A, Luime JJ, Boggild M, Uyl-de Groot CA, Stolk EA. Mapping QLQ-C30, HAQ, and MSIS-29 on EQ-5D. *Medical Decision Making*. 2012; 32:554–568. [PubMed: 22114301]
- Versteegh MM, Rowen D, Brazier JE, Stolk EA. Mapping onto EQ-5D for patients in poor health. *Health and Quality of Life Outcomes*. 2010; 8:141–148. [PubMed: 21110838]
- Wallander JL, Dekker MC, Koot HM. Risk factors for psychopathology in children with intellectual disability: a prospective longitudinal population-based study. *Journal of Intellectual and Disability Research*. 2006; 50:259–268.
- Warren Z, McPheeters ML, Sathe N, Foss-Feig JH, Glasser A, Veenstra-VanderWeele J. A Systematic Review of Early Intensive Intervention for Autism Spectrum Disorders. *Pediatrics*. 2011; 127:e1303–e1311. [PubMed: 21464190]
- Weinstein MC, S J, G M, K M, R L. Recommendations of the panel on cost-effectiveness in health and medicine. *JAMA*. 1996; 276:1253–1258. [PubMed: 8849754]
- Wille N, Badia X, Bonsel G, Burström K, Cavrini G, Devlin N, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. *Quality of Life Research*. 2010; 19:875–886. [PubMed: 20405245]

Abbreviations

APE	Absolute prediction error
ADOS	Autism Diagnostic Observation Schedule

ASD	Autism spectrum disorder
CBCL	Child Behavior Checklist
HUI	Health Utilities Index
HRQL	Health-related quality of life
MAPE	Mean absolute prediction error
PedsQL	Pediatric Quality of Life Inventory
QALY	Quality Adjusted Life Year
RMSE	Root mean squared error
Vineland-II	Vineland-II Adaptive Behavior Scales

Figure 1.

Table 1
Model specifications

Models	Dependent variables	Independent variables	Covariates*
Child Behavior Checklist (CBCL) Model 1	HUI3	Internalizing problems T score	Yes
		Externalizing problems T score	
Child Behavior Checklist (CBCL) Model 2	HUI3	CBCL Total problems T score	Yes
Vineland-II Adaptive Behavior Scales Model 1	HUI3	Communication	Yes
		Daily living skills	
		Socialization	
		Motor skills	
Vineland-II Adaptive Behavior Scales Model 2	HUI3	Communication	Yes
		Daily living skills	
		Socialization	
Vineland-II Adaptive Behavior Scales Model 3	HUI3	Vineland-II Composite score	Yes
Pediatrics Quality of Life Inventory (PedsQL) Model 1	HUI3	Physical functioning	Yes
		Emotional functioning	
		Social functioning	
		School functioning	
Pediatrics Quality of Life Inventory (PedsQL) Model 2	HUI3	PedsQL Total score	Yes
Combined Model 1	HUI3	CBCL Total problems T score	Yes
		Vineland-II Composite score	
		PedsQL Total score	
Combined Model 2	HUI3	CBCL Total problems T score	No
		Vineland-II Composite score	
		PedsQL Total score	

* Covariates are child age, squared child age, log cognitive ability (IQ), Autism Diagnostic Observation Schedule (ADOS) severity score.

Table 2

Demographic characteristics of children with ASDs from two sites of the Autism Treatment Network.

	N = 224
Age, mean $\pm$ SD (range)	8.4 $\pm$ 3.5 (4.0 – 17.9)
Gender	
Male	86.6%
Female	13.4%
Race/Ethnicity	
Caucasian	75.2%
African American	9.5%
Hispanic	9.0%
Asian	2.3%
Other	4.1%
Education level	
Preschool/kindergarten	45.0%
1 st – 2 nd grade	18.8%
3 rd – 5 th grade	14.2%
6 th – 8 th grade	7.8%
9 th – 12 th grade	5.0%
Other (e.g., home-based, special education)	8.3%
Birth order	
1 st child	50.7%
2 nd child	29.2%
3 rd child	13.0%
Others	7.2%
Child living arrangement	
At home	98.6%
In an institution or developmental center	1.4%
Type of school the child attend	
Non-specialized public school	44.6%
Specialized public school	8.2%
Vocational public school	0.5%
Private school	7.3%
Home school	2.3%
Special education school	26.8%
Others (e.g., special education program, self-contained class)	10.5%
Primary caregiver	
Biological parents	94.6%
adoptive parents	3.6%

Guardian	0.5%
others	1.3%
Diagnosis	
Autistic disorder	73.4%
Asperger's Disorder	8.7%
Pervasive developmental disorder-not otherwise specified	17.9%
Cognitive ability^a(IQ) 70	47%

^a Cognition scores are based on the Stanford-Binet Intelligence Scales, Fifth Edition, Abbreviated Battery or either the Mullen Scales or the Bayley Scales.

Table 3

Outcome measures and Spearman correlations with HUI3 scores (N=224).

Outcome Measures	N	Missing data (%)	Mean (SD), range	Spearman Correlations
				HUI3
HUI3 Scores	218	2.7	0.66 (0.23), -0.098 – 1.0	1.000
ADOS Severity Scores	205	8.4	7.2 (1.8), 2 – 10	-0.143*
Cognitive ability	197	12.1	75.7 (24.3), 41 – 148	0.310*
Child Behavior Checklist				
Externalizing problems T score	187	16.5	58.3 (11.0), 33 – 85	-0.110
Internalizing problems T score	187	16.5	60.9 (10.3), 33 – 83	-0.203**
Total problems T score	187	16.5	63.7 (9.4), 39 – 92	-0.224**
Vineland-II Adaptive Behavior Scales				
Communication	197	12.1	71.7 (15.2), 33 – 135	0.385***
Daily living skills	197	12.1	71.2 (13.3), 33 – 107	0.392***
Socialization	196	12.5	68.1 (11.3), 40 – 103	0.334***
Motor skills ^a	122	0	76.3 (12.4), 40 – 114	0.483***
Composite score	194	13.4	68.4 (11.1), 36 – 105	0.445***
Pediatrics Quality of Life Inventory 4.0				
Physical functioning	194	13.4	71.0 (20.0), 3.1 – 100	0.333***
Emotional functioning	190	15.2	64.8 (20.8), 5.0 – 100	0.152*
Social functioning	192	14.3	49.5 (24.0), 0 – 100	0.343***
School functioning	187	16.5	63.1 (20.3), 20 – 100	0.189*
Total score	194	13.4	63.0 (15.7), 11.8 – 100	0.375***

^aThe Vineland-II Motor Skills domain was used only in children age < 7 years.*
 $p < 0.05$,**
 $p < 0.01$,***
 $p < 0.001$

ADOS: Autism Diagnostic Observation Schedule

Table 4

Algorithms for estimating HUI3 scores by model.

4.1 HUI3 and Child Behavior Checklist models						
Parameter	Model 1			Model 2		
	Estimate	Robust STDERR	p	Estimate	Robust STDERR	p
Externalizing problems T score	-0.0015	0.0014	0.435			
Internalizing problems T score	-0.0043	0.0016	0.003			
Total problems T score				-0.0068	0.0015	<0.001
ADOS Severity	-0.0124	0.0085	0.165	-0.0136	0.0082	0.107
Child age	0.0078	0.0192	0.709	0.0116	0.0189	0.569
Squared child age	-0.0001	0.0011	0.924	-0.0002	0.0011	0.832
Log Cognitive ability (IQ)	0.2282	0.0511	0.033	0.2217	0.0503	0.025
Intercept	0.0744	0.2567	0.843	0.1726	0.2505	0.614
STDERR: Standard Error						
4.2 HUI3 and Vineland-II models						
Parameter	Model 1			Model 2		
	Estimate	Robust STDERR	p	Estimate	Robust STDERR	p
Communication	0.0033	0.0015	0.008	0.0037	0.0015	0.041
Daily living skills	0.0032	0.0018	0.152	0.0046	0.0017	0.016
Socialization	0.0002	0.0019	0.923	0.0010	0.0019	0.651
Motor skills	0.0033	0.0017	0.049			
Composite score						
ADOS Severity	-0.0050	0.0079	0.553	-0.0051	0.0079	0.539
Child age	0.0097	0.0193	0.644	0.0105	0.0192	0.619
Squared child age	0.0004	0.0010	0.746	0.0003	0.0010	0.788
Log Cognitive ability (IQ)	0.0252	0.0485	0.761	0.0238	0.0494	0.771
Intercept	-0.2367	0.2032	0.393	-0.1630	0.2008	0.525
STDERR: Standard Error						
				0.0103	0.0016	<0.001
				-0.0063	0.0078	0.437
				0.0119	0.0186	0.553
				0.0003	0.0010	0.805
				0.0304	0.0478	0.705
				-0.2438	0.2015	0.342

4.3 HUI3 and Pediatrics Quality of Life Inventory models					
Parameter	Model 1			Model 2	
	Estimate	Robust STDERR	p	Estimate	p
Physical functioning	0.0031	0.0008	0.001		
Emotional functioning	0.0004	0.0008	0.656		
Social functioning	0.0021	0.0007	0.003		
School functioning	-0.0001	0.0008	0.957		
Total score				0.0059	0.0008
ADOS Severity	-0.0104	0.0075	0.196	-0.0100	0.0077
Child age	0.0016	0.0183	0.465	0.0177	0.0184
Squared child age	-0.0003	0.0010	0.785	-0.0039	0.0010
Log Cognitive ability (IQ)	0.1875	0.0445	0.019	0.1971	0.0454
Intercept	-0.5172	0.2262	0.160	-0.5374	0.2298
STDERR: Standard Error					

4.4 HUI3 and combined measures models					
Parameter	Model 1			Model 2	
	Estimate	Robust STDERR	p	Estimate	p
CBCL total problems T score	-0.0028	0.0017	0.176	-0.0023	0.0019
Vineland II composite score	0.0067	0.0015	<0.001	0.0076	0.0013
PedsQL total score	0.0041	0.0009	0.001	0.0034	0.0011
ADOS Severity	-0.0120	0.0072	0.122		
Child age	0.0273	0.0197	0.192		
Squared child age	-0.0006	0.0010	0.558		
Log Cognitive ability (IQ)	0.1204	0.0511	0.100		
Intercept	-0.4850	0.2345	0.079	0.0667	0.1886
STDERR: Standard Error; CBCL: Chile Behavior Checklist; PedsQL: Pediatrics Quality of Life Inventory					