Mapping scores from the Strengths and Difficulties Questionnaire (SDQ) to preference-based utility values

Article *in* Quality of Life Research · August 2013

DOI: 10.1007/s11136-013-0494-6 · Source: PubMed

CITATIONS
8

READS
189

4 authors:

Gareth Furber
Flinders University
34 PUBLICATIONS **169** CITATIONS
SEE PROFILE

Leonie Segal
University of South Australia
153 PUBLICATIONS **1,857** CITATIONS
SEE PROFILE

Matthew Leach
University of South Australia
91 PUBLICATIONS **877** CITATIONS
SEE PROFILE

Jane Cocks
University of South Australia
1 PUBLICATION **8** CITATIONS
SEE PROFILE

Qual Life Res
DOI 10.1007/s11136-013-0494-6

Mapping scores from the Strengths and Difficulties Questionnaire (SDQ) to preference-based utility values

Gareth Furber · Leonie Segal · Matthew Leach · Jane Cocks

Accepted: 30 July 2013

Abstract
Purpose Quality of life mapping methods such as "Transfer to Utility" can be used to translate scores on disease-specific measures to utility values, when traditional utility measurement methods (e.g. standard gamble, time trade-off, preference-based multi-attribute instruments) have not been used. The aim of this study was to generate preliminary ordinary least squares (OLS) regression-based algorithms to transform scores from the Strengths and Difficulties Questionnaires (SDQ), a widely used measure of mental health in children and adolescents, to utility values obtained using the preference-based Child Health Utility (CHU9D) instrument.
Methods Two hundred caregivers of children receiving community mental health services completed the SDQ and CHU9D during a telephone interview. Two OLS regressions were run with the CHU9D utility value as the dependent variable and SDQ subscales as predictors. Resulting algorithms were validated by comparing predicted and observed group mean utility values in randomly selected subsamples.
Results Preliminary validation was obtained for two algorithms, utilising five and three subscales of the SDQ, respectively. Root mean square error values (.124) for both models suggested poor fit at an individual level, but both algorithms performed well in predicting mean group observed utility values.
Conclusion This research generated algorithms for translating SDQ scores to utility values and providing researchers with an additional tool for conducting health economic evaluations with child and adolescent mental health data.

Keywords Utility · Mapping · Mental health · Child and adolescent

Abbreviations

OLS	Ordinary least squares
SDQ	Strengths and Difficulties Questionnaire
CHU9D	Child Health Utility—9 Dimension
QALY	Quality adjusted life year
CUA	Cost-utility analysis
CAMH	Child and adolescent mental health
TTU	Transfer to Utility

Background

Governments and health agencies are increasingly concerned with allocating scarce health care resources in a way that is efficient, so as to maximise health and well-being of the population. This requires techniques that can compare performance of services across modalities and disease areas and where quality of life and mortality impacts may be expected. Health economists have developed the evaluation method of cost-utility analysis (CUA) for this

G. Furber (✉) · L. Segal
Health Economics and Social Policy Group, University of South Australia, Playford Building, City East Campus, North Terrace, Adelaide, SA 5000, Australia
e-mail: Gareth.furber@unisa.edu.au

M. Leach
School of Nursing and Midwifery, University of South Australia, Centenary Building, City East Campus, North Terrace, Adelaide, SA 5000, Australia

J. Cocks
Social Epidemiology and Evaluation Research Group, University of South Australia, Playford Building, City East Campus, North Terrace, Adelaide, SA 5000, Australia

Published online: 13 August 2013

purpose, in which performance is expressed as cost per gain in quality-adjusted life year (QALY). The QALY is a generic measure of health outcome that incorporates life years and quality of life. The QALY is calculated as the product of a utility value attached to individual health states and time in each health state, where the utility value falls between 0, which is equivalent to death (negative for states worse than death), and 1, which is equivalent to full health. The utility scale has equal interval properties and a precise and equivalent relationship with length of life (for example a quality of life utility increase from .6 to .8 over 5 years is equal to a gain of 1 quality-adjusted life year).

The strength of utilities and the associated QALYs is that a wide range of interventions can be meaningfully compared, using the metric of cost per QALY gain. It enables comparison across services that address different health problems, cover distinct disease groups and employ different modalities. Cost-utility analysis is the preferred option for economic evaluations in the Australian pharmaceutical approval system [1].

The measurement of utility values ideally proceeds through one of the following two approaches: preference-based scaling approaches, in which utility is elicited directly from patients using techniques such as the time trade-off (TTO) or standard gamble (SG) approach; or through completion by patients (or others with a knowledge of the health states) of a preference-based multi-attribute instrument (MAU) that assesses individual health across dimensions designed to cover all core inputs into quality of life. Utility values are then derived from published scoring algorithms developed from population preference studies. Preference-based multi-attribute instruments, such as the EQ-5D [2], are now quite widely used in adult health research.

Neither method of estimating quality of life utility values in child and adolescent mental health (CAMH) has been widely adopted. This reflects both the multiple challenges of health state valuation in children [3] and a broader gap in the economic evaluation of CAMH services [4–6]. Few published studies of child and adolescent mental health interventions have incorporated utility instruments [7, 8], and few (if any) services providing CAMH treatment are collecting utility values from their clients. This means we have a limited set of utility values for children and adolescents [9, 10], which is limiting the potential use of cost-utility analyses in determining treatment priorities for CAMH services.

One proposed solution to this problem is mapping [11–14], which is recognised by the National Institute for Health and Clinical Excellence [15] as a method for generating utility information in situations where utility measurement has not been undertaken. The most popular mapping method, possibly because of its simplicity and modest data requirements, is Transfer to Utility (TTU). This approach involves the concurrent administration of a descriptive measure and a preference-based multi-attribute instrument and the development of a regression-based algorithm; ordinary least squares (OLS) is the most popular [13], which transforms scores on the descriptive measure to the preference-based measure [16]. Once derived, the algorithm can be used to map available data from descriptive instruments (where preference-based utility instruments were not used) onto utility data for cost-utility analyses. Several reviews of studies that have utilised this approach have been published to date [11–13].

To our knowledge, there have been no mapping studies conducted in CAMH, limiting the tools available to health economic researchers who are attempting to conduct cost-utility analyses with outcomes from CAMH research studies and health services where utility measurement has not been conducted. Thus, the goal of this pilot study was to determine whether a suitably robust OLS-based transformation algorithm could be developed to translate scores on the widely used Strengths and Difficulties Questionnaire (SDQ) [17] to utility values. The SDQ was selected for the current study because it is mandated for use in Australia's specialised CAMH services as a consumer-oriented outcome assessment tool. Furthermore, national and international data coordination efforts (e.g. http://amhocn.org/, http://www.corc.uk.net/) have led to the creation of large SDQ data sets, which represent thousands of episodes of care in CAMH services across Australia and the United Kingdom. An algorithm to transform SDQ scores to utility values would therefore have immediate value in conducting cost-utility analyses of CAMH service data.

Methods

Design

The study employed a Transfer to Utility approach to map scores from the SDQ onto utility values derived from a preference-based multi-attribute instrument, the Child Health Utility—9 Dimension (CHU9D) [18]. The study was approved by both Health Service (#384.11) and University ethics committees (#25739).

Participants

Child health status and emotional and behavioural health were assessed by proxy. Proxy outcome measurement is a common practice both in CAMH services and quality of life studies where seeking self-report from children can be compromised by age and comprehension issues. Furthermore, limited resources and the use of a telephone interview format in this study precluded consenting and interviewing children and adolescents. Participants were

parents or other relatives of children aged 5–17 (inclusive) who were registered as "current clients" of a regional child and adolescent mental health service. Current client status was defined as having an open episode of care and a recorded contact within the last 6 weeks. Excluded were caregivers who had no recorded telephone number, had specific "no contact" instructions in the electronic clinical record, were foster carers or whose child was the subject of current guardianship or family court orders.

Procedure

Potential participants were identified from the electronic clinical record of the CAMH service and placed on a list. The order of participants on the list was randomised before being provided to telephone interviewers. All listed participants were sent out introductory letters at least 1 week prior to being contacted (by phone) by interviewers. Where a participant was identified as having more than one child receiving CAMH services, a coin toss method was used to identify which child the participant would be asked to rate. A minimum sample size of 138 was identified based on power estimates [19] for linear multiple regression assuming alpha at .05, power at 95 %, a medium effect size and five predictors (the five subscales of the SDQ). Interviews continued however until funding for interviewers was exhausted, which occurred when 200 participants had been recruited.

Measures

A telephone survey instrument was developed that consisted of both the SDQ and the CHU9D, in addition to questions about demographics, child's presenting issues and CAMH service satisfaction.

Strengths and Difficulties Questionnaire

The SDQ [17] was first developed as a shorter alternative to behavioural screening questionnaires such as the Rutter [20] and Child Behaviour Checklist [18, 19] but with an additional focus on consumer "strengths". The SDQ has repeatedly demonstrated equivalence to these longer measures in terms of factor structure, reliability, sensitivity to detecting psychiatric diagnoses and sensitivity to change [21–23]. The instrument is now a widely used mental health screening measure for use in children and adolescents aged 4–17.

The SDQ comprises 25 items, each describing a psychological or behavioural attribute (some positive, some negative) which the responder indicates as being "very true", "somewhat true" or "not true" of the child/adolescent in question over the last 6 months. The instrument generates both a total score and scores for five subscales, including emotional, conduct, hyperactivity–inattention, peer problems and prosocial behaviour. The total score ranges from 0 to 40 with higher values indicating greater behavioural and emotional pathology. Individual subscales scored from 0 to 10 with higher scores indicating poorer functioning for four of the subscales (emotional, conduct, hyperactivity–inattention and peer problems) and better functioning for one of the subscales (prosocial). The SDQ is available in three forms—adolescent self-report, caregiver administered and teacher administered. The form used for this study was the caregiver-administered form.

Child Health Utility—9D

The Child Health Utility—9D [18, 24, 25] is a 9-item preference-based multi-attribute utility instrument designed for use in children aged 7–11. The nine items, each with five response categories, assess the child's functioning "today" across domains of worry, sadness, pain, tiredness, annoyance, school, sleep, daily routine and activities. The instrument is available in both self-report (completed by the child) and proxy-report (caregiver completed) forms, and the wording of the instrument enables its use with children as young as 5. The decision to use the proxy-report form in the current study was driven.

The CHU9D was developed to address the paucity of paediatric preference-based measures for use in health care resource allocation decision-making [24]. The instrument focuses on the impact of health issues on quality of life, rather than measuring levels of impairment or disability. The domains were identified from qualitative interviews with children (aged 7–11) with a variety of health problems [18]. Two sets of preference weights are available. The first (original tariff) was generated from health state valuation interviews with 300 members of the UK adult general population [26]. The second set (alternative tariff) was developed by Ratcliffe and colleagues [27], based on best–worst scaling interviews with 590 Australian adolescents. Although relatively new, the CHU9D is increasingly being used in clinical outcome studies, and recent work in Australia is supporting its use in adolescents and pre-adolescents [28, 29].

Prior to analysis, a comparison of utility values generated using the two tariffs was conducted. Both tariffs generated average utility values lower than published general population utility norms in this age range (.90 to .92) [30], consistent with the sample being a clinical population. The mean CHU9D utility value calculated using the alternative tariff was significantly lower and more in line with values reported by Petrou and colleagues [9, 10] for children with psychiatric disorders, which ranged from .43 to .70. The alternative tariff also showed a more normal

distribution of values. This tariff was also generated from a survey of 590 Australian adolescents aged 11–17 and thus expected to be more pertinent to our population. Based on these considerations, the decision was made to utilise the alternative tariff for algorithm estimation. The decision is consistent with evidence from studies [29, 31] which showed that the original CHU9D tariff may overestimate average utility values when compared to the alternative tariff and utility values obtained from other instruments (e.g. HUI—Health Utilities Index, EQ5D).

Data analysis

Data analysis was conducted using SPSS version 19. Data were screened and cleaned. There was one missing CHU9D data point and 26 missing SDQ data points. Two SDQ items ("kind to younger children" and "steals") were the most frequently missed (nine data points in total), whilst other missing data points were scattered across the remaining 23 items. Missing data represented just .4 % of all data items. A review of the raw questionnaire data revealed caregivers had reported "don't know" on these items. For analysis purposes, a "no problem" approach was taken, where missing values were replaced with the equivalent value for "no problem".

Descriptive statistics for the full sample were calculated. The representativeness of the sample was assessed by comparing SDQ total scores with national SDQ data for Australian CAMHS services accessed through http://wdst.amhocn.org/. CHU9D scores were translated to utility values using both the original and alternative tariffs. Selection of the most appropriate tariff for further analysis was decided on the basis of utility value distribution, comparison with published utility values for similar populations and consideration of the sample frame for generating the utility scoring. A scatter plot of SDQ total scores and CHU9D utility values was visually inspected to determine whether a linear (OLS) regression model was appropriate.

Primary analysis involved running two OLS regressions, using enter and stepwise procedures. In both, the CHU9D utility value was estimated as a function of the five scales of the SDQ (emotion, conduct, peer, prosocial and hyperactivity) using OLS regression. For the enter procedure, all five scales were entered simultaneously, assuming equal importance of the scales for the final prediction. For the stepwise procedure, scales were retained based on probability of F (enter $P < .05$) determined by SPSS, in an attempt to isolate those scales with unique predictive value. No second-order, interaction or additional demographic terms were entered. In the case of demographic variables, age and gender were not significantly related to observed utility values. Alternative modelling strategies (e.g. item-based OLS, response mapping) were considered but rejected based on sample size and a consideration of the goal of the mapping—to derive a simple algorithm for generating mean utility values for groups based on SDQ scores, for which OLS is best suited.

Regression models were evaluated using a number of criteria:

1) R^2 and adjusted R^2 values were used as a measure of the explanatory power of the models.
2) Regression coefficients were used to determine which SDQ scales accounted for the greatest variation.
3) Root mean square error (RMSE) was used as a measure of the degree of individual predictor error.
4) Standardised residuals were used to identify outliers, using a criterion of no more than 5 % of the sample being $\geq$2 standard deviations from the mean. Cook's distance (>1 indicating concern) and leverage values (greater than 3 times average indicating concern) were used to identify influential cases.
5) Variance inflation factor (VIF) scores (>10) were used to identify multi-collinearity between predictor variables (i.e. SDQ subscales).
6) Durbin–Watson statistic was used to identify the presence of autocorrelation. A Durbin–Watson value of <1 or >3 was used as indicator of significant autocorrelation.
7) Residual plots were used to detect biases in prediction due to deviations in normality or heteroscedasticity.

Two transformation algorithms, based on the two regression models were calculated using the formula—"Utility = (Constant) + β_1 (scale 1) + β_2 (scale 2) + …". The predictive validity of these algorithms at the group level was tested by taking six randomly drawn subsamples of differing sizes (two each of approximately 25, 50 and 75 %) from the total sample and calculating the difference between predicted group utility values and observed group utility values. Acceptable performance was defined by group differences of less than .03—a "rule of thumb" based on Drummond [30].

Results

A total of 900 participants met the inclusion criteria during the data collection period and were randomised for contact. Interviewers attempted to contact caregivers by moving sequentially through the list of caregivers until two hundred interviews were completed. This resulted in 407 caregivers being approached, of whom 150 were not contactable, 37 declined to be interviewed, 14 were discovered not to meet criteria and 6 interviews were not complete. Descriptive statistics for the full sample (missing data imputed) are presented in Table 1.

Table 1 Descriptive characteristics of participating caregivers and their children

	Participants ($N = 200$)
Age of child, mean (SD)	11.71 (5.75)
Distribution by age band (%)	5–7—10.5
	8–10—25
	11–13—31
	14–17—33.5
Gender of Child (%)	Male = 52.5
	Female = 47.5
Caregiver (%)	Mother = 87
	Father = 8.5
	Other = 4.5
First time with CAMHS (%)	Yes = 74.5
	No = 25.5
Length of time with CAMHS for current episode (months), mean (SD)	11.95 (16.6)
SDQ total score, mean (SD)	19.52 (7.87)
SDQ emotion subscale, mean (SD)	5.40 (2.55)
SDQ conduct subscale, mean (SD)	4.01 (2.86)
SDQ hyperactivity subscale, mean (SD)	6.41 (2.81)
SDQ peer subscale, mean (SD)	3.71 (2.43)
SDQ prosocial subscale, mean (SD)	6.92 (2.23)
CHU-9D utility score, mean (SD)	Alternative tariff—.740 (.144)

Three quarters of participants were first-time CAMHS clients. Most (87 %) participants were mothers. The mean SDQ score for the children in the current study (19.52) was comparable to the national averages at admission and review in CAMH services (Table 2). Based on SDQ scoring guidelines for total problems, 132 of the children were in the clinical range, 24 were in the borderline range and 44 were in the normal range. The proportion of children with scores indicating clinically significant problems in specific domains were as follows: emotional problems (60 %), conduct problems (51 %), hyperactivity (51 %), peer problems (50 %) and prosocial (17 %). Two-thirds of children had difficulties in two or more areas, and almost 30 % of children had difficulties in 4 or more areas.

Results from the two OLS regression models are presented in Table 3. Both models were significantly better than the mean as an estimate of observed value, and SDQ subscales explained around 28 % of the variance in utility values. Model 1 had a potential utility prediction range of .51 to .93, whilst Model 2 had a potential utility prediction range of .528 to .918. This compares with a range of .33 to 1.0 for the CHU9D alternative scoring tariff.

Emotion, conduct and peer subscales accounted for the greatest amount of variance and were the only scales entered in the stepwise regression. Root mean square error (RMSE) values indicated significant error at the individual level (large differences between predicted and actual values) but did not favour one model over another. Standardised residuals identified 3 % (Model 1) and 2.5 % (Model 2) of sample as being potential outliers, which was within the 5 % target. Centred leverage values identified 1 case in Model 1 and 3 cases in Model 2 as influential cases, but removal of these cases did not improve the models. VIF scores did not identify significant multi-collinearity, and Durbin Watson values were within desired range.

Inspection of residual plots revealed mild biases common to both models. First, standardised residuals suggested a bias of both models to generate predicted values higher than observed. Second, there was greater residual error at lower values of utility, albeit still evenly scattered. Third, the peer and prosocial subscales demonstrated some heteroscedascity (bunching of scores at one end of the scale), raising some doubt as to the reliability of the peer and prosocial coefficients in the regression equations.

Overall, there was little to separate the two models in terms of statistical characteristics or prediction at the group level.

Table 4 summarises the differences between predicted and observed group means for six randomly drawn subsamples of differing sizes (two each of approximately 25, 50 and 75 % of the entire sample). The average difference between predicted and observed group means was .009 for Model 1 and .010 for Model 2. These are both well below the .03 "rule of thumb" value suggested by Drummond [32] as indicating a clinically significant difference. The largest difference was .02, still less than a clinically significant difference.

In terms of discriminative ability, mean predicted utility values for the clinical bandings of the SDQ are presented in Table 5. Both algorithms were able to discriminate between these groups with average mean utility values, demonstrating a clear relationship with SDQ severity.

Discussion

The current study demonstrates that it is possible to use simple OLS-based regression to develop algorithms that translate scores on descriptive measures of child and adolescent mental health to utility values for use in health economic evaluation. Using the SDQ, a widely used measure of emotional and behavioural functioning, and the CHU9D, a preference-based utility instrument designed specifically for use with children and adolescents, two preliminary algorithms for generating utility values from

Table 2 Comparison of SDQ total scores from study sample with national Australian averages

	N	Mean	SD	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
Current sample	200	19.5	7.9	8	14	19	25	30
National statistics—admission SDQ[a]	9292	20.2	6.9	11	15	21	25	29
National statistics—review SDQ[a]	4718	19.9	7.0	10	15	20	25	29

[a] Available from http://wdst.amhocn.org/

Table 3 OLS regressions predicting CHU9D utility score ("alternative tariff") from SDQ subscales

Model	Predictors	β	SE	*t*	Sig	R^2	Adjusted R^2	RMSE
1—Scale, enter	Constant	.880	.044	19.963	.000	.284	.265	.124
	Emotion	−.019	.004	−4.912	.000			
	Conduct	−.009	.004	−2.302	.022			
	Hyper	−.001	.004	−.278	.782			
	Peer	−.008	.004	−1.798	.074			
	Prosocial	.005	.004	1.171	.243			
2—Scale, stepwise	Constant	.918	.022	40.892	.000	.278	.267	.124
	Emotion	−.018	.004	−4.878	.000			
	Conduct	−.012	.003	−3.447	.001			
	Peer	−.009	.004	−2.189	.030			

Table 4 Predicted versus observed utility scores for random subsamples

Sample (*n*)	Observed	Model 1		Model 2	
	Mean utility value (SD)	Mean utility value (SD)	Difference	Mean utility value (SD)	Difference
Sample 1 (54)	.742 (.129)	.738 (.080)	.014	.740 (.082)	.016
Sample 2 (61)	.724 (.129)	.739 (.076)	.003	.738 (.077)	.004
Sample 3 (82)	.736 (.144)	.746 (.073)	.010	.747 (.073)	.011
Sample 4 (105)	.720 (.149)	.739 (.080)	.019	.740 (.079)	.020
Sample 5 (157)	.733 (.140)	.736 (.073)	.003	.737 (.073)	.005
Sample 6 (158)	.745 (.143)	.742 (.077)	.002	.741 (.076)	.004
Average difference			**.009**		**.01**

Table 5 Predicted utility scores for clinical bandings of the SDQ

	Model 1	Model 2
Normal ($N = 44$), mean (SD)	.84 (.040)	.84 (.040)
Borderline ($N = 24$), mean (SD)	.77 (.034)	.78 (.032)
Abnormal ($N = 132$), mean (SD)	.70 (.055)	.70 (.056)

SDQ subscales were generated, both with excellent predictive ability at the group level. The models are presented below (Table 6). To our knowledge, these are the first "Transfer to Utility" algorithms available for use in the child and adolescent mental health area. At this stage, there is little evidence to suggest one algorithm is superior to the other. If theoretical consistency is desired, we recommend the algorithm containing all five SDQ subscales.

The algorithms have significant potential value in the conduct of health economic evaluations where SDQ data, but not utility data, has been collected. In Australia and the United Kingdom, this could include analysis of national mental health data sets for children and adolescents. Internationally, this could include any study in which the SDQ has been used as an outcome, screening or descriptive instrument. This includes psychotherapy or drug trials, cross-sectional or longitudinal correlational studies and population health studies. The algorithms require only subscale scores and can thus be applied in situations where only subscale information is available. The development of

Table 6 Preliminary algorithms for generating utility scores from SDQ subscales

Algorithm using five SDQ subscales
Utility = .880 + (−.019 × emotion) + (−.009 × conduct) + (−.001 × hyper) + (−.008 × peer) + (.005 × prosocial)
Algorithm using three SDQ subscales
Utility = .918 + (−.018 × emotion) + (−.012 × conduct) + (−.009 × peer)

these algorithms is not intended to replace the use of validated preference-based instruments when evaluating outcomes in children and adolescents with mental health disorders, but rather to provide an alternative method for transforming SDQ data when preference-based instruments have not or cannot be used.

There are some cautions in using these algorithms. First, they are only recommended for predicting group means. They are not suitable for predicting individual values. The algorithms predict an estimated 28 % of the variance in observed CHU9D values. This is on the lower side of R^2 values for similar mapping studies reviewed by Brazier et al. [13]. This is not surprising given the two instruments relate to different time frames (i.e. "6 months" versus "today") have different foci (i.e. behaviours versus impact, population-specific versus generic) and utilise different response categories. Such differences are not uncommon when attempting to map scores generated from disease-specific measures to generic measures [11]. Analysis of the error variance revealed a couple of potential biases, namely that the algorithms may generate slightly higher average utility values, and that the accuracy of predicted utility values may be lower at lower utility values (i.e. more extreme SDQ scores). However, in the current sample, both algorithms predicted group utility means within .02 for all subsamples and within .01 for most. We recommend further testing of these algorithms in similar CAMHS samples.

Second, the range of the algorithms are limited to .51 to .93 (Model 1) and .528 to .918 (Model 2), a severe truncation of the utility scale. By these scales, children with severe SDQ profiles would score in the low .50s, whilst children with no behavioural or emotional issues would score in the high .80s/low .90s. Despite this limited range, there was evidence that the algorithm could discriminate between the different clinical bandings of the SDQ when the sample was divided into abnormal, borderline and normal categories (based on SDQ scores). We also note a utility value of .5 implies a severe reduction in quality of life—whereby an individual would be prepared to trade half of their remaining life expectancy in exchange for full health. Those attending a community-based child and adolescent mental health service (as against an inpatient service) are unlikely to have utility values below .5 or close to 1.0 (denoting full health). Whether the truncation of range has an impact on the usefulness of the algorithm is a topic for further investigation.

Third, the algorithms are currently recommended only for use in populations similar to that of the current study. Participants in the current study were selected randomly from a large regional CAMH services in Australia. Based on their SDQ total scores, the sample appears representative of children and families seeking services from CAMH services around Australia. We believe the algorithms have relevance for both Australian and international community-based child and adolescent mental health services; however, replication studies in other CAMH services would be highly valuable. The value of the algorithm in subsets of child and adolescent populations with severe difficulties and impairments (e.g. autism, bipolar disorder and early psychosis) is unknown and should be the subject of further study.

Finally, we made a number of methodological choices in this study that require addressing. The CHU9D instrument was chosen over other potential child or adolescent utility instruments (e.g. the HUI3) because of its brevity, relevant dimensions, availability of a scoring tariff developed with Australian adolescents and because pilot testing in the service had demonstrated strong face validity with children, adolescents and caregivers. The decision to go with proxy measurement was driven by practical constraints (use of telephone interviews) as well as attempting to reflect the common practice of completion of routine outcome measures by parents and caregivers in CAMH services. The decision to utilise SDQ scales rather than SDQ items was driven by both sample size and observations that SDQ scales are more reliable and normally distributed than individual items. Finally, the choice to use the experimental Ratcliffe tariffs [27] despite best–worst scaling methods being controversial was driven by the fact these tariffs yielded utility values that were much more in line with those published by Petrou and colleagues [9, 10].

Whilst Transfer to Utility through OLS regression is not the only mapping method available to researchers to derive utility values from disease-specific health status instruments, it is useful for developing an algorithm to transform scores across two quite structurally and conceptually different instruments—but which are seeking to capture quality of life. OLS regression was viewed as the most theoretically defensible, in that it represents a mathematical

transformation based on a representative sample and free of assumptions about how the scales and items of each of the instruments should be related. It is also practical and transparent in implementation.

There is significant scope for additional research in this area building upon the preliminary algorithms from this study. Most notably, this includes replication of the current study in larger data sets, which will allow exploration of item-based algorithms, utilisation of other preference-based quality of life instruments (HUI3, 17D, EQ-5D-Y), direct administration of the utility instrument to children and the comparison of alternative statistical mapping methods.

Conclusion

The successful development of robust algorithms for estimating preference-based utility values from the SDQ opens up a large body of literature and large databases of service data for health economic evaluation and the opportunity to have that data inform policy and resource allocation in child and adolescent mental health.

Acknowledgements The authors would like to acknowledge the School of Nursing and Midwifery at the University of South Australia for the research development grant that enabled this study to take place as well as Southern Mental Health—Child and Adolescent Mental Health Service (SMH-CAMHS) for their support in hosting ongoing health economic projects within their service.

References

1. Department of Health & Ageing. (2012). Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee. *Pharmaceutical Benefits Advisory Committee*. Accessed January 10, 2013, from http://www.pbs.gov.au/industry/listing/elements/pbac-guidelines/PBAC4.3.2.pdf.
2. Rabin, R., & De Charro, F. (2001). EQ-5D: A measure of health status from the EuroQol group. *Annals of Medicine, 33*(5), 337–343.
3. Ungar, W. J. (2011). Challenges in health state valuation in paediatric economic evaluation: Are QALYs contraindicated? *PharmacoEconomics, 29*(8), 641–652.
4. Romeo, R., Byford, S., & Knapp, M. (2005). Annotation: Economic evaluations of child and adolescent mental health interventions: a systematic review. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 46*(9), 919–930.
5. Hsia, R. Y., & Belfer, M. L. (2008). A framework for the economic analysis of child and adolescent mental disorders. *International Review of Psychiatry, 20*(3), 251–259.
6. Kilian, R., Losert, C., Park, A.-L., McDaid, D., & Knapp, M. (2010). Cost-effectiveness analysis in child and adolescent mental health problems: An updated review of literature. *International Journal of Mental Health Promotion, 12*(4), 45–57.
7. Dalziel, K., & Segal, L. (2009). Economic evaluation in child protection: What are the special challenges? In W. Ungar (Ed.), *Economic evaluation in child health* (pp. 133–164). London: Oxford Press.
8. Kromm, S. K., Bethell, J., Kraglund, F., Edwards, S. A., Laporte, A., Coyte, P. C., et al. (2012). Characteristics and quality of pediatric cost-utility analyses. *Quality of Life Research, 21*(8), 1315–1325.
9. Petrou, S., Johnson, S., Wolke, D., Hollis, C., Kochhar, P., & Marlow, N. (2010). Economic costs and preference-based health-related quality of life outcomes associated with childhood psychiatric disorders. *The British Journal of Psychiatry, 197*, 395–404.
10. Petrou, S., & Kupek, E. (2009). Estimating preference-based Health Utilities Index mark 3 utility scores for childhood conditions in England and Scotland. *Medical Decision Making, 29*(3), 291–303.
11. Mortimer, D., & Segal, L. (2008). Comparing the incomparable? A systematic review of competing techniques for converting descriptive measures of health status into QALY-weights. *Medical Decision Making, 28*(1), 66–89.
12. Petrillo, J., & Cairns, J. (2008). Converting condition-specific measures into preference-based outcomes for use in economic evaluation. *Expert Review of Pharmacoeconomics and Outcomes Research, 8*(5), 453–461.
13. Brazier, J. E., Yang, Y., Tsuchiya, A., & Rowen, D. L. (2010). 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, 11*(2), 215–225.
14. Chuang, L. H., & Whitehead, S. J. (2012). Mapping for economic evaluation. *British Medical Bulletin, 101*(1), 1–15.
15. National Institute for Health and Clinical Excellence. (2008). *Guide to the methods of technology appraisal*. London: National Institute for Health and Clinical Excellence.
16. Segal, L., Day, S. E., Chapman, A. B., & Osborne, R. H. (2004). Can we reduce disease burden from osteoarthritis? *The Medical Journal of Australia, 180*(5 Suppl), S11–S17.
17. Goodman, R. (2007). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 38*(5), 581–586.
18. Stevens, K. J. (2008). *Working with children to develop dimensions for a preference based generic paediatric health related quality of life measure*. Health Economics and Decision Science Discussion Paper 08/04. http://www.shef.ac.uk/scharr/sections/heds/discussion.html.
19. Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160.
20. Rutter, M. (1967). A children's behaviour questionnaire for completion by teachers: Preliminary findings. *Journal of Child Psychology and Psychiatry, 8*, 1–11.
21. Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*(11), 1337–1345.
22. Hawes, D. J., & Dadds, M. R. (2004). Australian data and psychometric properties of the Strengths and Difficulties Questionnaire. *Australian and New Zealand Journal of Psychiatry, 38*, 644–651.
23. Goodman, A., & Goodman, R. (2009). Strengths and Difficulties Questionnaire as a dimensional measure of child mental health. *Journal of the American Academy of Child and Adolescent Psychiatry, 48*(4), 400–403.
24. Stevens, K. (2009). Developing a descriptive system for a new preference-based measure of health-related quality of life for children. *Quality of Life Research, 18*(8), 1105–1113.
25. Stevens, K. (2011). Assessing the performance of a new generic measure of health-related quality of life for children and refining

it for use in health state valuation. *Applied Health Economics and Health Policy, 9*(3), 157–169.
26. Stevens, K. (2012). Valuation of the child health utility 9D index. *PharmacoEconomics, 30*(8), 729–747.
27. Ratcliffe, J., Flynn, T., Terlich, F., Brazier, J., Stevens, K., & Sawyer, M. (2012). Developing adolescent specific health state values for economic evaluation: An application of profile case best worst scaling to the Child Health Utility-9D. *Pharmacoeconomics, 30*, 713–727.
28. Ratcliffe, J., Stevens, K., Flynn, T., Brazier, J., & Sawyer, M. (2012). An assessment of the construct validity of the CHU9D in the Australian adolescent general population. *Quality of Life Research, 21*(4), 717–725.
29. Ratcliffe, J., Stevens, K., Flynn, T., Brazier, J., & Sawyer, M. (2012). Whose values in health? An empirical comparison of the application of adolescent and adult values for the CHU-9D and AQOL-6D in the Australian adolescent general population. *Value in Health, 15*(5), 730–736.
30. Drummond, M. (2001). Introducing economic and quality of life measurements into clinical studies. *Annals of Medicine, 33*(5), 344–349.
31. Canaway, A. G., & Frew, E. J. (2012). Measuring preference-based quality of life in children aged 6–7 years: A comparison of the performance of the CHU-9D and EQ-5D-Y-the WAVES Pilot Study. *Quality of Life Research, 22*(1), 173–183.
32. Ratcliffe, J., Couzner, L., Flynn, T., Sawyer, M., Stevens, K., Brazier, J., et al. (2011). Valuing Child Health Utility 9D health states with a young adolescent sample: A feasibility study to compare best-worst discrete choice experiment, standard gamble and time trade off methods. *Applied Health Economics and Health Policy, 9*(1), 15–27.