

Original Research Article

Child-rated and parent-rated quality of life in childhood intermittent exotropia: findings from an observational cohort study

ABSTRACT

Purpose: To use the Pediatric Quality of Life Inventory (PedsQL™) to describe generic quality of life (QOL) in children with intermittent exotropia [X(T)], to examine changes in scores, and to compare scores in children with X(T) to those of age-matched healthy cohorts.

Methods: PedsQL™ was administered to children and parents as part of the UK Improving Outcomes in Intermittent Exotropia (IOXT) study. Excluding 27 children with co-morbidity, PedsQL data was available from 365 parents and 152 children. Paired-samples t-tests examined change in PedsQL™ scores over time. One-sample t-tests and mean differences compared scores between children with X(T) and healthy UK samples.

Results: Mean parent-rated PedsQL™ scores from the X(T) cohort at baseline were: 90.6 (Physical Health), 78.2 (Emotional Functioning), 88.8 (Social Functioning), 83.4 (School/Nursery Functioning), 83.6 (Psychosocial Summary), 86.2 (Total). Mean baseline child-rated scores were: 78.1 (Physical Health), 76.5 (Emotional Functioning), 73.6 (Social Functioning), 72.2 (School/Nursery Functioning), 74.2 (Psychosocial Summary), 75.5 (Total). X(T) parents rated their child's QOL similar to healthy children's parents, except for poorer School/Nursery Functioning in 2-4 year olds. X(T) children rated their QOL significantly better than age-matched healthy children. There were no significant changes over time.

Conclusions: Using the PedsQL™ we were unable to detect significant effects of X(T) on generic QOL. However, evidence for PedsQL's utility in this condition remains limited without further investigation in larger samples and concurrent control groups.

30 Further qualitative work and consideration of condition-specific measures in UK
31 cohorts are needed before practitioners can better inform parents about psychosocial
32 impacts of X(T).
33

34 INTRODUCTION

35 Intermittent exotropia [X(T)] is one of the commonest forms of childhood strabismus.
36 In X(T), one eye intermittently drifts outwards when the child is looking at distant
37 objects, or is tired or inattentive. Potential functional consequences include loss of
38 stereovision and amblyopia. Parents often have concerns relating to social exclusion
39 and bullying because of the appearance of the strabismus.¹
40

41 Childhood X(T) has been associated with subsequent poor psychological well-being
42 into young adulthood.^{2,3} There may also be more immediate implications for the
43 child's quality of life (QOL) yet relatively little exists in the literature about this.⁴⁻
44 ⁸One group of researchers, from the Mayo Clinic in North America, has studied parental
45 and children's concerns regarding the effect of X(T) on QOL, leading to the recent
46 development of their Intermittent Exotropia Questionnaire (IXTQ).⁶⁻¹⁰ As far as we are
47 aware, there have been no other reports of paediatric QOL in this specific
48 condition and the relevance of this group's work to a different population and
49 healthcare system in the UK is unestablished.
50

51 QOL estimates of patients, including those with ophthalmologic conditions, often
52 differ significantly from those of physicians.¹¹ Capturing child and
53 parental perspectives is important since parental insight may not correspond with that
54 of their children,¹² and there is growing recognition that each needs to be taken into

55 consideration. Parent-child agreement on QOL ratings in X(T) has been found
56 to be poor.¹³

57

58 A widely-used and accepted generic QOL instrument is the Pediatric Quality of Life
59 Inventory (PedsQLTM), which has strong evidence of reliability and validity in healthy
60 children and those with a range of acute and chronic illnesses.¹⁴ However, evidence of
61 its usefulness in assessing generic QOL in children with X(T) is very limited. Here we
62 describe parent- and child-rated PedsQLTM scores from a UK observational study of
63 X(T) (the Improving Outcomes in Intermittent Exotropia (IOXT) study).¹⁵⁻¹⁷ The
64 primary aim of this paper is to examine change over time in self- and parent-rated
65 PedsQLTM scores. A secondary aim is to compare baseline PedsQLTM scores to those
66 of unaffected age-matched cohorts.

67

68 **METHODS**

69 **Participants**

70 The IOXT study involved 26 children's eye clinics / orthoptic departments.¹⁵ Between
71 May 2005 and December 2006, collaborating departments approached parents/carers
72 of children under 12 years, diagnosed with X(T) within the previous 12 months,
73 without significant coexisting ocular pathology (e.g. cataract). In addition to clinical
74 assessment, demographic details were recorded at enrolment including age, gender,
75 and general health. Treatment status was documented (observation only, conservative
76 treatment for squint such as glasses/patching, eye muscle surgery, or treatment for
77 visual acuity only). Clinical and QOL follow-up ended in December 2008 (December
78 2009 for surgery patients). As this was an observational study, clinicians managed
79 children according to their usual clinical criteria. Informed consent was obtained from

80 parents/guardians. The study received a favourable opinion from UK North West
81 Multi-Centre Research Ethics Committee.

82

83 **Outcome measures**

84 Generic QOL data were collected from parents and children using the PedsQL™ 4.0
85 Generic Core Scales(<http://www.pedsql.org/index.html>). We used the age-
86 appropriate ‘proxy’ (parent/carer) version for those with children aged 2-4 years
87 (toddler version), 5-7 years and 8 years or older, and the child’s self-rated version for
88 those aged 5-7 and 8 years or older. The PedsQL assesses QOL on four individual
89 scales: Physical Health (8 items), Emotional Functioning (5 items), Social
90 Functioning (5 items) and School Functioning (5 items, or ‘Nursery/Day Care
91 Functioning’ for the toddler version which has 3 items). These four scales can be
92 amalgamated to form a Total score (23 or 21 items) and 2 summary scores: Physical
93 Health summary and Psychosocial Health summary. We followed developers’ scoring
94 instructions by reversing and linearly transforming scores to a 0-100 scale (higher
95 scores indicating better QOL), and, for missing data, imputation of the mean of
96 completed items in a scale when 50% or more are completed.

97

98 PedsQL™ questionnaires were administered in the clinic setting as part of the IOXT
99 study assessment in all except 2 cases. In accordance with PedsQL™ administration
100 guidelines, child versions were self-completed by those aged 8 or older. If the child
101 was 5-7 years, instructions and questions were read to the child word for word by an
102 orthoptist who then circled his/her response. Parent versions were self-completed. A
103 written protocol and training video were produced for use by study orthoptists to
104 ensure PedsQLs were administered in a standardised way. This generic instrument has
105 been widely validated in numerous conditions and in healthy groups, and we found

106 that its psychometric properties specifically in children with X(T) were generally
 107 good.¹³
 108
 109 Baseline PedsQL™ data were scheduled to be collected within 12 weeks of
 110 enrolment. Follow-up appointments were arranged on a 3-monthly basis during the
 111 first year of the study and 6-monthly thereafter for an average period of 2 years. These
 112 visits involved both clinical and QOL review. Outcome QOL is the latest PedsQL™
 113 completed, but at least 6 months after the initial QOL assessment.

114

115 Data analysis

116 Data were analysed using SPSS version 19.0. Summary statistics (mean scores,
 117 standard deviations (SD)) are reported for each PedsQL™ scale. Analyses are
 118 performed across the whole cohort of children with X(T), i.e. for all age groups
 119 combined, and (see Supplemental Material) repeated for age-specific versions of the
 120 PedsQL™. Paired-samples t-tests were performed to examine change over time.

121

122 Baseline scores are compared with age-matched published scores^{18,19} from healthy
 123 UK samples: mean differences and one-sample t-testswere used.Comparisons with
 124 healthy UK samples are possible for those aged between 2 years and 4 years 11
 125 months¹⁸ and for those aged between 5.5 and 8.5 years.¹⁹All scales/sub-scales are
 126 available for use in the 2-4 year old comparison. Only the Total, Physical Health and
 127 Psychosocial Summary scores can be used in the older age-group comparison, as
 128 individual scores for Emotional, Social and School/Nursery Functioning were not
 129 published for the 5.5 to 8.5 year-old healthy group.

130

131 One-way analysis of variance was used to examine whether severity of X(T) at
 132 baseline was associated with QOL at initial assessment. Severity was determined by
 133 scores on the Newcastle Control Score (NCS)²⁰ with a score of 1 to 3 denoting mild, 4
 134 to 6 moderate, and 7 to 9 severe X(T). One-way analysis of variance was used to
 135 explore any association between treatment and PedsQLTM scores (see Supplemental
 136 Material). Given the scarcity of reports of QOL in X(T), and to place this study in
 137 context, further descriptive comparison was made of PedsQLTM scores in this sample
 138 and those reported in the American study.⁸

139

140 Co-morbidity was present in 27 children, the most common co-existing health issues
 141 being asthma and developmental delay. We have excluded those with co-morbidity
 142 from our analyses as they had significantly poorer QOL compared to those without
 143 other health conditions (see Supplemental Material).

144

145 **Statement of Ethics**

146 We certify that all applicable institutional and governmental regulations concerning
 147 the ethical use of human volunteers were followed during this research.

148

149 **RESULTS**

150 **Participant characteristics**

151 Baseline PedsQLTM data were available from 365 parents and 152 children (excluding
 152 27 children with comorbidity). Mean (SD) age at time of initial parent PedsQLTM was
 153 52 (22) months. Mean (SD) age at time of initial child PedsQLTM was 74 (16) months.
 154 57% of the children were female.

155

156 **Baseline PedsQLTM scores**

157 Mean (SD) baseline parent-rated scores for all 365 parents were 86.2 (12.8) on Total
 158 score, 90.6 (12.8) on Physical Health, 78.2 (18.8) on Emotional Functioning, 88.8
 159 (15.5) on Social, 83.4 (17.6) on School and 83.6 (14.7) on Psychosocial Summary.
 160 Mean (SD) baseline scores for all 152 children with QOL data were 75.5 (14.7) on
 161 Total score, 78.1 (17.4) on Physical Health, 76.5 (19.5) on Emotional Functioning,
 162 73.6 (23.2) on Social, 72.2 (19.6) on School and 74.2 (16.2) on Psychosocial
 163 Summary. Age-specific PedsQL™ scores are provided in the Supplemental Material.

164

165 **Paired data**

166 Mean interval between baseline and outcome PedsQL™ was 24 months (range 6 to
 167 48) for parents and 20 months (6 to 48) for children. Paired data (baseline and
 168 outcome scores) were available from 301 parents and 128 children: Figure 1 shows
 169 their baseline and corresponding outcome scores. For parent-ratings there was a small,
 170 statistically significant deterioration over time on Total (mean difference 1.5, 95% CIs
 171 0.1 to 2.9, $p=0.036$), School/Nursery (mean difference 2.5, 95% CIs 0.3 to 4.8,
 172 $p=0.025$), and Psychosocial Summary (mean difference 1.9, 95% CIs 0.3 to 3.5,
 173 $p=0.023$) scales. Paired data within age-specific versions are provided in the
 174 Supplemental Material. For child-ratings (Figure 1) there were no significant changes
 175 in self-rated scores.

176

177 Treatment received was not significantly associated with PedsQL™ scores in any of
 178 the parent- or child-rated PedsQL™ scales at baseline or outcome (see Supplemental
 179 Material). Mean parental PedsQL™ scores were higher at outcome on each scale
 180 (except Physical Health) for those with a better surgical outcome, defined as a NCS of
 181 between 0 and 2. However, this was not statistically significant (see Supplemental
 182 Material).

183

184 **Comparisons with unaffected children (UK sample)**

185 *2-4 year olds*

186 Parents of 2 to 4 year-old children with X(T) rated their child's QOL similarly to
187 those of age-matched healthy children in terms of Total and summary (Physical
188 Health and Psychosocial) scores (Table 1). Further details are provided in the
189 Supplemental Material.

190

191 *5.5 to 8.5 year olds*

192 In the 5.5 to 8.5 year-old group, there were no significant differences between X(T)
193 parents' and healthy parents' ratings on Total or summary scores (Table 1). Children
194 with X(T) rated their own QOL as significantly better than the age-matched healthy
195 sample on the Total, Physical and Psychosocial Summary scales (Table 1).

196

197 **Cross-cultural comparisons**

198 Figure 2 shows that baseline parental PedsQLTM scores were very similar to those in a
199 recent North American study of QOL in X(T).⁷ In contrast, child-rated scores were
200 significantly lower amongst our sample compared to their American counterparts with
201 X(T) (one-sample t tests, $p=0.005$ to $p<0.001$).

202

203 **Squint severity**

204 Although children with severe X(T) generally had poorer mean parent- and self-rated
205 PedsQLTM scores than those with mild or moderate X(T), this was not statistically
206 significant. Details are reported in the Supplemental Material.

207

208 **DISCUSSION**

209 Drawing on data from a large UK multicentre observational study of children with
 210 X(T), we have described child-rated and parent-rated generic QOL using the widely-
 211 used and validated PedsQL™. We found that parent-rated PedsQL™ scores for
 212 children with X(T) were similar to those of unaffected cohorts in the UK, while
 213 the self-rated scores of the children themselves were better than those of unaffected
 214 children. These findings differ from those of a research group from the Mayo Clinic in
 215 North America (the only other group to have used the PedsQL™ specifically in this
 216 condition).⁸ In their study of 51 children with X(T) and 47 controls, they found that
 217 parent-rated scores were poorer in children with X(T) than in controls, while child-
 218 rated scores were similar between affected and unaffected children.⁸ The differences
 219 between the UK and North American studies are most likely due to variations in the
 220 way the healthy samples were recruited given that the latter benefitted from a
 221 concurrent control group. Age is also a likely factor, since the age of the healthy
 222 sample in the American study ranged from 5 to 16 years whereas those in the present
 223 study ranged from 2 to 8.5 years.

224
 225 We found no effects of treatment on PedsQL™ scores, and no significant deterioration
 226 in self-rated PedsQL™ scores over an average follow-up period two years. While
 227 some parent-rated PedsQL™ scores did worsen slightly over time within the youngest
 228 age group, the size of these changes is unlikely to be meaningful.²¹

229
 230 Generic measures such as the PedsQL™ are important because of their applicability
 231 to a diverse range of conditions, enabling comparisons between various groups
 232 including healthy populations and thereby placing scores into context. However, they
 233 may not be sensitive enough to demonstrate effects associated with strabismus
 234 generally or X(T) in particular, or its treatment. A particularly informative way to

235 determine any effect of X(T) on QOL is through qualitative methods. Indeed
 236 qualitative research, also by the Mayo Group in North America, reveals the most
 237 frequently mentioned QOL ‘concerns’ of children with X(T) was worry.⁶ From the
 238 parents’ perspective, “comments from other people” was the most frequently
 239 mentioned concern.

240

241 Although no paediatric strabismus-specific measures were available at the start of our
 242 study, the qualitative findings in America have recently informed the development of
 243 the Intermittent Exotropia Questionnaire (IXTQ).⁷ The advantage of such condition-
 244 specific measures is that they should be responsive to any change over time, for
 245 example in those who undergo treatment for their condition, while the main
 246 disadvantage is that they can not be used across different conditions or to compare
 247 with unaffected populations. Thus there will remain a need for both condition-specific
 248 and generic patient-reported outcome measures, depending on the context of the
 249 study. A popular solution is the use of a ‘battery’ of outcome measures incorporating
 250 both specific and generic components, although this can be more time-consuming and
 251 may result in respondent burden.

252

253 An important limitation of this paper is that we did not have a concurrent control
 254 group with which to compare the PedsQLTM scores of children with X(T) and we
 255 therefore used published, age-matched scores from unaffected UK children. The
 256 published data for the older control group were collected from one region of the UK.
 257 Consequently, there may have been geographical and socio-economic differences
 258 between this group and our national IOXT cohort. The data for unaffected toddlers,
 259 however, were collected from a wider geographical area and from a variety of
 260 sources. Another important limitation is that, due to the young age of the children in

261 the IOXT study, there were less baseline child-rated PedsQL™ data with which to
 262 evaluate change over time from their own perspective.
 263
 264 The findings from the IOXT study show that UK children with X(T) did not appear to
 265 fare any worse than unaffected peers in terms of PedsQL™ scores, and that there was
 266 no detrimental effect on their PedsQL™ scores over time. These findings do not
 267 necessarily imply a lack of effect of X(T) on QOL. However, they do add to the sparse
 268 literature on the condition itself and on the utility of the universally-accepted
 269 PedsQL™ as an outcome measure specifically in ophthalmology. While the
 270 PedsQL™ has demonstrated reasonable psychometric properties in X(T),¹³ the
 271 evidence regarding its direct relevance to children with eye conditions remains
 272 limited. The implications for the management of this condition in the UK are difficult
 273 to establish, although the lack of any significant difference between children with
 274 X(T) and those without it should be reassuring to worrying parents. Further
 275 investigation of the comparison of generic and condition-specific instruments in larger
 276 cohorts will further help clinicians.
 277
 278 In conclusion, parent- and child-rated PedsQL™ scores were similar to or better than
 279 those of unaffected cohorts in the UK. Scores were not altered by treatment or time.
 280 Local and international evidence about the impact of intermittent exotropia on a
 281 child's quality of life remains limited. Further qualitative work should be pursued
 282 together with further investigation of established QOL measures in larger samples,
 283 and with concurrent control groups, in order to inform researchers, practitioners and
 284 parents about the specific effects of childhood X(T) and in turn how to evaluate its
 285 treatment.
 286

287 REFERENCES 288

- 289 1.Horwood J, Waylen A, Herrick D, Williams C, Wolke D. Common visual defects
290 and peer victimisation in children. *Invest Ophthalmol Vis Sci.* 2005;46:1177-1181.
291
- 292 2.McKenzie JA, Capo JA, Nusz KJ, Diehl NN, Mohny BG.Prevalence and sex
293 differences of psychiatric disorders in young adults who had intermittent exotropia as
294 children. *Arch Ophthalmol.* 2009;127:743-747.
295
- 296 3.Mohny BG, McKenzie JA, Capo JA, Nusz KJ, Mrazek D, Diehl NN. Mental
297 illness in young adults who had strabismus as children. *Pediatr.* 2008; 122: 1033-
298 1038.
299
- 300 4. Wen G, McKean-Cowdin R, Varma R, Tarczy-Hornoch K, Cotter SA, Borchert M,
301 et al. General health-related quality of life in preschool children with strabismus or
302 amblyopia. *Ophthalmol.* 2011; 118: 574-580.
303
- 304 5.Archer SM, Musch DC, Wren PA, Guire KE, Del Monte MA. Social and emotional
305 impact of strabismus surgery on quality of life in children. *J AAPOS.*2005; 9: 148-
306 151.
307
- 308 6.Hatt SR, Leske DA, Adams WE, Kirgis PA, Bradley EA, Holmes JM. Quality of
309 life in intermittent exotropia: child and parent concerns. *Arch Ophthalmol.*2008; 126:
310 1525-1529.
311

- 312 7. Hatt SR, Leske DA, Yamada T, Bradley EA, Cole SR, Holmes JM. Development
313 and initial validation of quality-of-life questionnaires for intermittent
314 exotropia. *Ophthalmol.* 2010; 117: 163-168.
315
- 316 8. Hatt SR, Leske DA, Holmes JM. Comparison of quality-of-life instruments in
317 childhood intermittent exotropia. *J AAPOS.* 2010;14:221-226.
318
- 319 9. Yamada T, Hatt SR, Leske DA, Holmes JM. Spectacle wear in children reduces
320 parental health-related quality of life. *J AAPOS.* 2011; 15:24-28.
321
- 322 10. Yamada T, Hatt SR, Leske DA, Holmes JM. Specific health-related quality of life
323 concerns in children with intermittent exotropia. *Strabismus.* 2012; 20:145-151.
324
- 325 11. Stein JD. Disparities between ophthalmologists and their patients in
326 estimating quality of life. *Curr Opin Ophthalmol.* 2004; 15: 238-243.
327
- 328 12. Upton P, Lawford J, Eiser C. Parent-child agreement across child health-related
329 quality of life instruments: a review of the literature. *Qual Life Res.* 2008; 17: 895-
330 913.
331
- 332 13. Buck D, Clarke MP, Powell C, Tiffin P, Drewett RF. Use of the PedsQL in
333 childhood intermittent exotropia: estimates of feasibility, internal consistency
334 reliability and parent-child agreement. *Qual Life Res.* 2012; 21: 727-736.
335

336 14. Varni JW, Seid M, Kurtin PS. PedsQL 4.0: Reliability and validity of the Pediatric
337 Quality of Life Inventory version 4.0 generic core scales in healthy and patient
338 populations. *Med Care*. 2001; 39: 800-812.

339

340 15. Buck D, Powell C, Cumberland P, Davis H, Dawson E, Rahi J et al. Presenting
341 features and early management of childhood intermittent exotropia in the UK:
342 inception cohort study. *Br J Ophthalmol*. 2009;93:1620-1624.

343

344 16. Buck D, Powell CJ, Rahi J, Cumberland P, Tiffin P, Sloper J et al. The improving
345 outcomes in intermittent exotropia study: outcomes at 2 years after diagnosis in an
346 observational cohort. *BMC Ophthalmol*. 2012;12:1.

347

348 17. Buck D, Powell CJ, Sloper JJ, Taylor R, Tiffin P, Clarke MP. Surgical intervention
349 in childhood intermittent exotropia: Current practice and clinical outcomes from an
350 observational cohort study. *Br J Ophthalmol*. 2012;96: 1291-1295.

351

352 18. Buck D. The PedsQL™ as a measure of parent-rated quality of life in healthy UK
353 toddlers: psychometric properties and cross-cultural comparisons. *J Child Health
354 Care*. 2012;16:331-338.

355

356 19. Cremeens J, Eiser C, Blades M. Factors influencing agreement between child self-
357 report and parent proxy-reports on the Pediatric Quality of Life Inventory 4.0
358 (PedsQL) generic core scales. *Health Qual Life Outcomes*. 2006; 4: 58.

359

360 20. Buck D, Clarke MP, Haggerty H, Hrisos S, Powell C, Sloper J et al. Grading the
 361 severity of intermittent distance exotropia: the revised Newcastle Control Score. *Br J*
 362 *Ophthalmol.* 2008;92:577.

363

364 21. Varni JW, Burwinkle TM, Seid M, Skarr D. The PedsQL 4.0 as a
 365 pediatric population health measure: feasibility, reliability, and validity. *Ambul*
 366 *Pediatr.* 2003; 3: 329-341.

367 **Legends for illustrations**
 368

369 Figure 1 Mean (SD) parent and child PedsQL scores at baseline and outcome

370 Figure 2 Comparison of IOXT study PedsQL™ scores with those from a recent study
 371 of X(T) in North America
 372

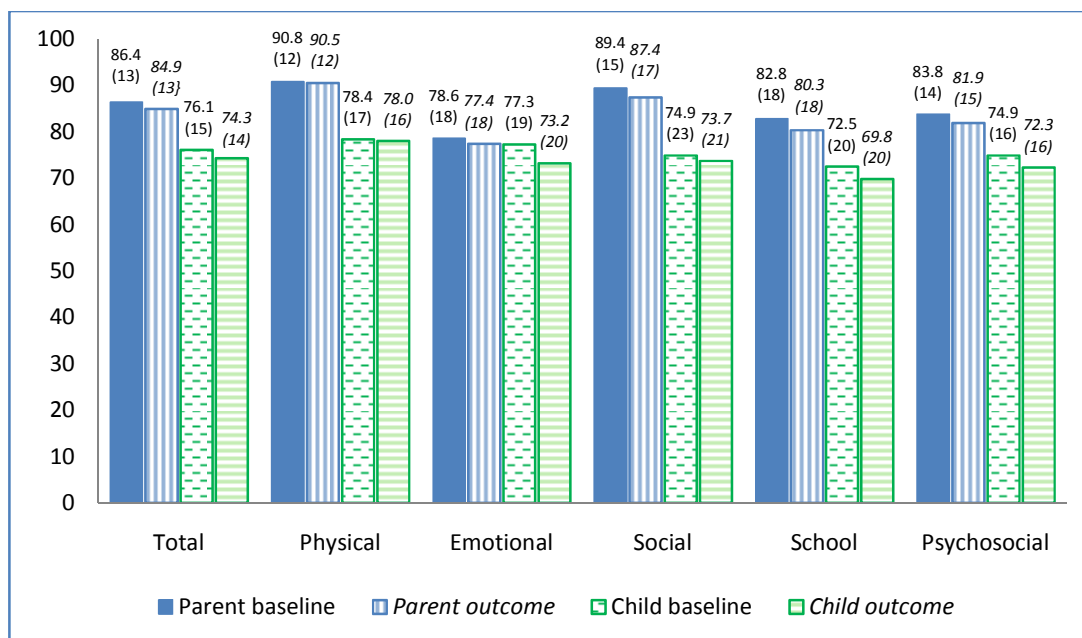
373

374 **Table 1** Mean difference in PedsQL scores: X(T) sample compared to age-
 375 **matched healthy samples**
 376

Mean (SD):				
Parental scores 2 to 4 years	Healthy sample (n=256)	X(T) sample (n=234)	Mean difference; 95% CIs	p value
Total PedsQL	87.8 (8.7)	89.0 (10.2)	1.2; -0.2 to 2.5	0.083
Physical Health	92.6 (9.1)	92.3 (10.3)	-0.3; -1.7 to 0.99	0.62
Psychosocial Summary	84.6 (10.5)	86.8 (12.2)	2.2; 0.63 to 3.8	0.006
Parental scores 5.5 to 8.5 years	Healthy sample (n=149)	X(T) sample (n=82)	Mean difference; 95% CIs	p value
Total PedsQL	79.9 (11.7)	80.8 (15.2)	0.9; -2.4 to 4.3	0.58
Physical Health	86.1 (11.4)	88.0 (15.9)	1.9; -1.6 to 5.4	0.28
Psychosocial Summary	76.7 (13.0)	76.9 (17.0)	0.2; -3.5 to 3.9	0.89
Child scores 5.5 to 8.5 years	Healthy sample (n=149)	X(T) sample (n=92)	Mean difference; 95% CIs	p value
Total PedsQL	71.8 (14.4)	77.8 (13.9)	6.0; 3.1 to 8.9	<0.001
Physical Health	76.4 (14.0)	80.8 (15.7)	4.4; 1.1 to 7.6	<0.001
Psychosocial Summary	68.9 (16.2)	76.3 (15.2)	7.4; 4.2 to 10.5	<0.001

377 *one-sample t-test
 378
 379

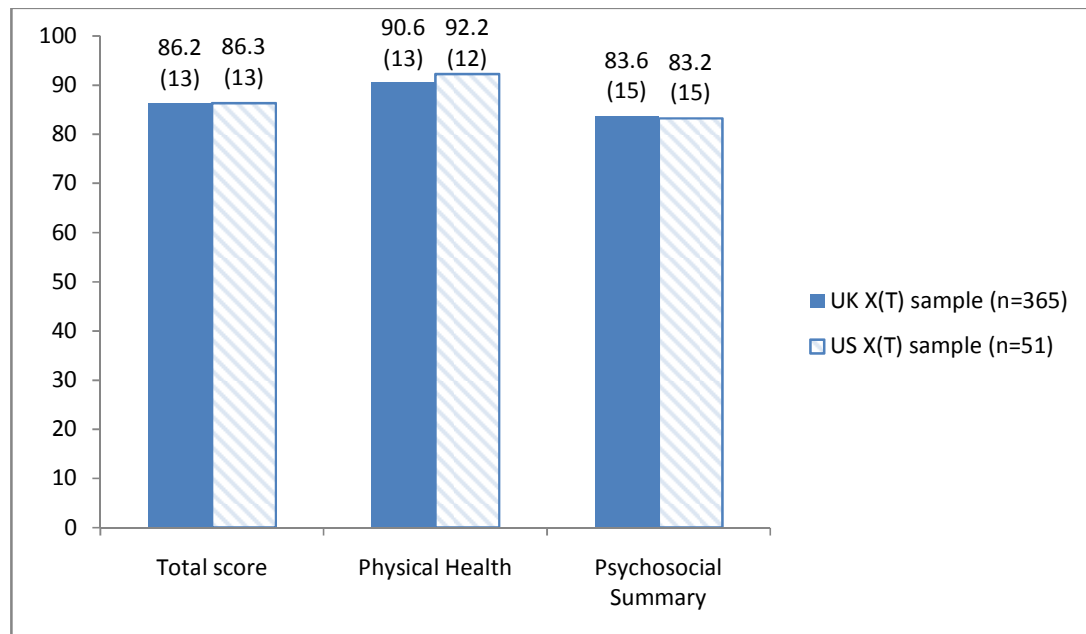
380 **Figure 1 Mean (SD) parent and child PedsQL scores at baseline and outcome**



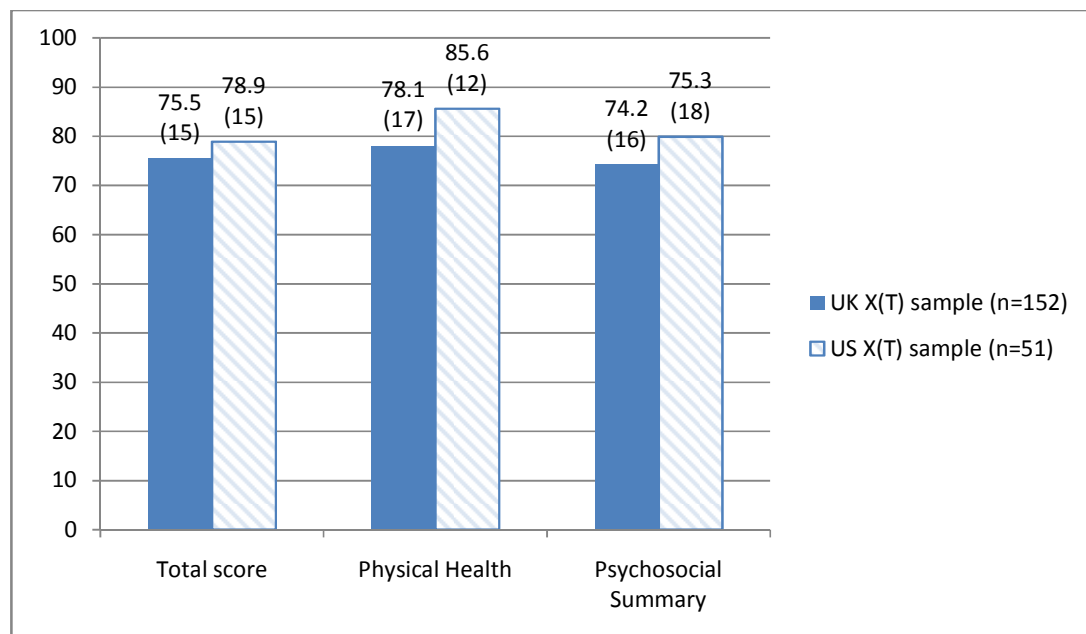
381
382

Figure 2 Comparison of IOXT study PedsQL™ scores with those from a recent study of X(T) in North America

a) Parental scores



b) Child-rated scores



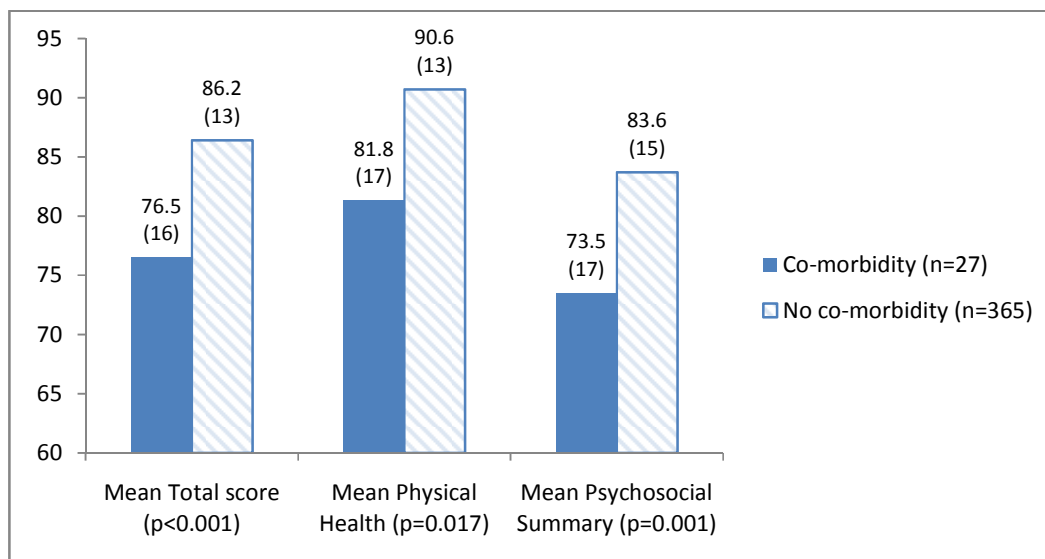
Supplemental material

Co-morbidity

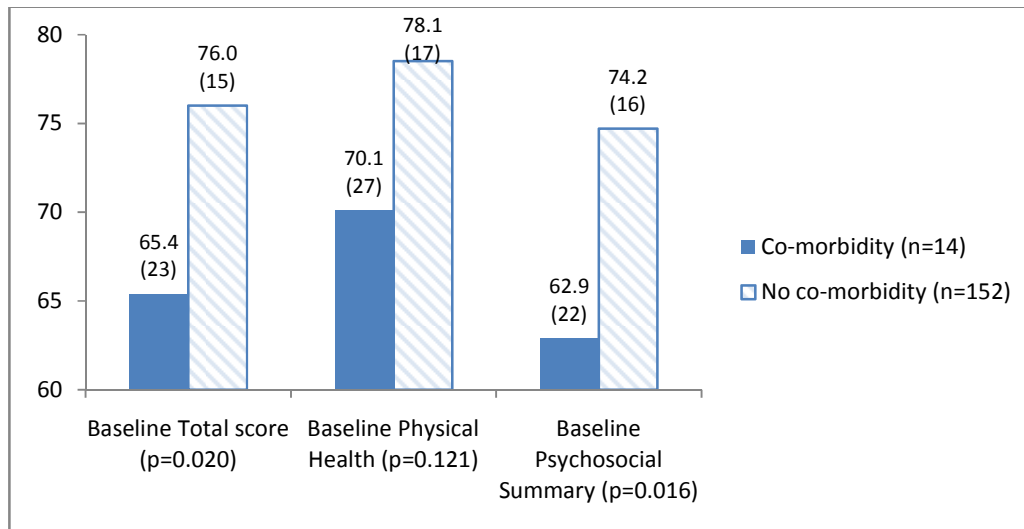
Children who presented with other health conditions had poorer PedsQL™ scores than those without co-morbidities on each parent-rated scale ($p=0.048$ to $p<0.001$), and on the Total and Psychosocial Summary child-rated scales ($p=0.020$, $p=0.016$ respectively; Supplementary Figure 1).

Supplementary Figure 1: Relationship between co-morbidity and baseline PedsQL™ scores

a) Parental ratings



b) Child-ratings



Scores on the age-specific versions of the PedsQL™ in children with intermittent exotropia

Mean (SD) baseline parent-rated scores for all parents of 2-4 year olds (n=253) were: 88.0 (11) on Total score, 92.3 (11) on Physical Health, 81.2 (17) on Emotional Functioning, 92.3 (12) on Social Functioning, 88.0 (15) on Nursery Functioning and 86.9 (12) on Psychosocial Summary score.

Mean (SD) baseline parent-rated scores for all parents of 5-7 year olds (n=95) were: 80.5 (15) on Total score, 87.3 (16) on Physical Health, 72.0 (20) on Emotional Functioning, 80.9 (19) on Social Functioning, 77.5 (18) on School Functioning and 76.8 (16) on Psychosocial Summary score.

Mean (SD) baseline parent-rated scores for all parents of those aged 8 or over (n=17) were: 76.1 (18) on Total score, 83.8 (16) on Physical Health, 66.6 (23) on Emotional Functioning, 80.9 (21) on Social Functioning, 68.2 (23) on School Functioning and 72.0 (20) on Psychosocial Summary score.

446

447 Mean (SD) baseline child-rated scores for all 5-7 year olds (n=134) were: 75.4 (15) on
448 Total score, 77.6 (18) on Physical Health, 76.8 (20) on Emotional Functioning, 73.0
449 (24) on Social Functioning, 72.7 (19) on School Functioning and 74.3 (16) on
450 Psychosocial Summary score.

451

452 Mean (SD) baseline child-rated scores for all of those aged 8 or over (n=18) were:
453 76.4 (14) on Total score, 81.4 (14) on Physical Health, 74.2 (18) on Emotional
454 Functioning, 78.3 (20) on Social Functioning, 68.9 (22) on School Functioning and
455 73.8 (16) on Psychosocial Summary score.

456

457 Supplementary Figure 2 illustrates the mean Total and Summary scores for the age-
458 specific versions of the PedsQL™ at baseline and outcome for those with paired
459 data. Paired-samples t-tests found a small but statistically significant deterioration over
460 time in parents' Total and psychosocial ratings within the 2-4 year old age group. The
461 mean deterioration (95% CIs) was -3 (-5 to -1) on the Total score (p=0.001), and -3.7
462 (-6 to -2) on the Psychosocial summary score (p<0.001). There were no significant
463 changes in Physical Health in this age group. There were no significant changes on
464 any parent-rated scales in the older age groups, and none on any of the child-rated
465 scales (Supplementary Figure 2).

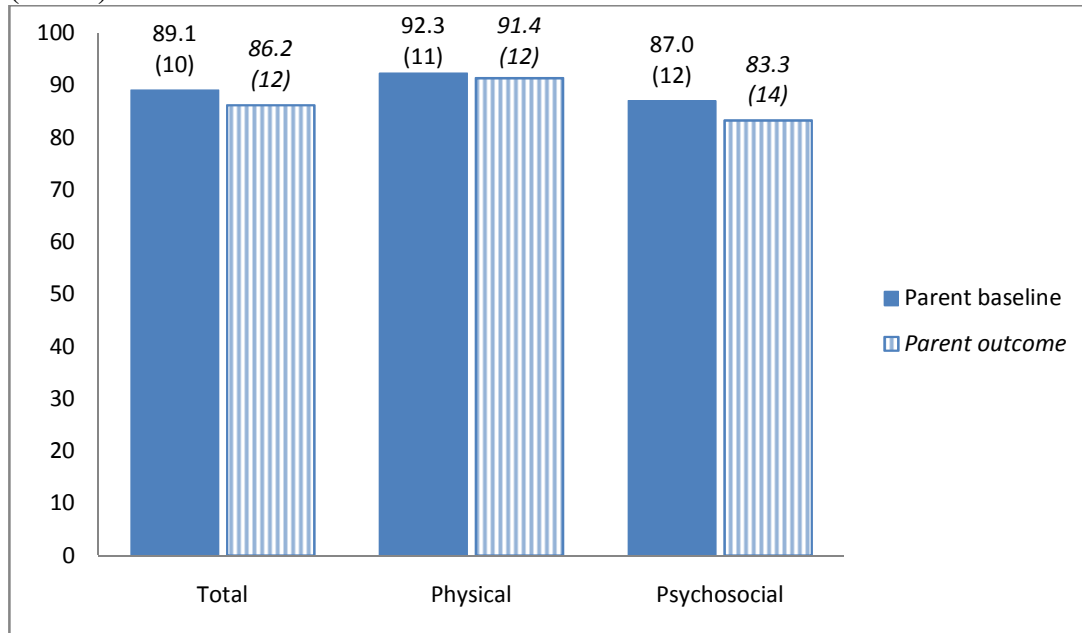
466

467

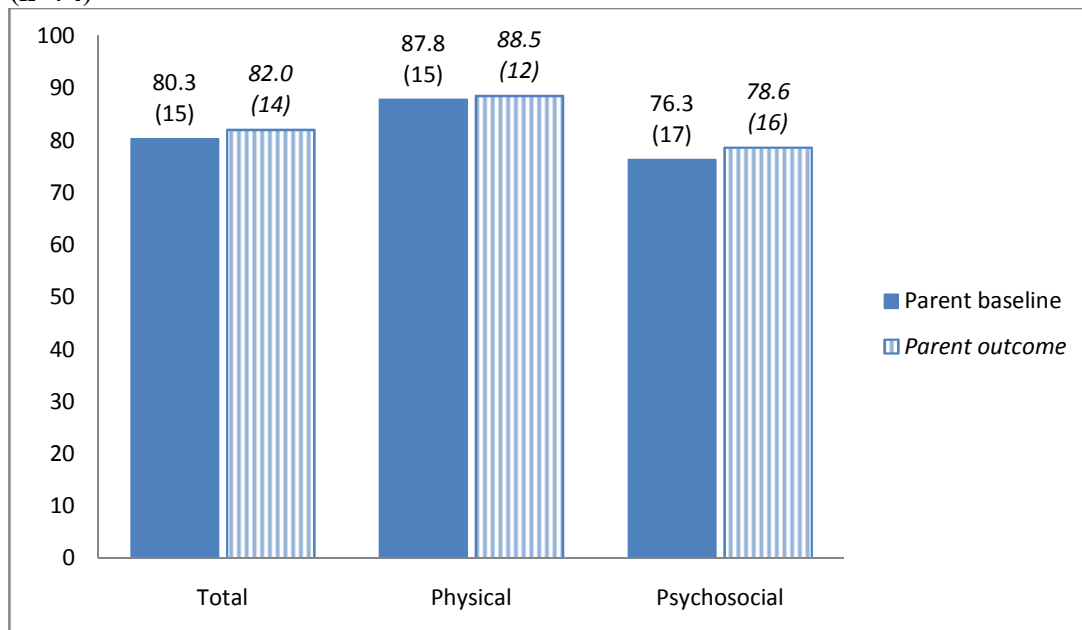
468

Supplementary Figure 2: Mean (SD) parent and child PedsQL scores (age-specific) at baseline and outcome

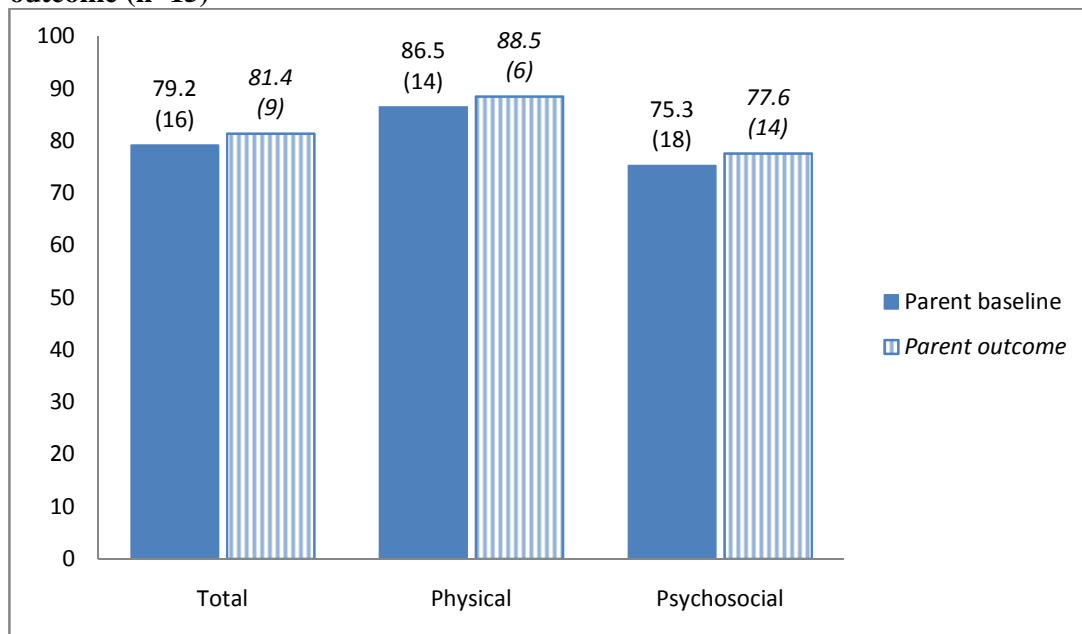
a) Parental ratings for 2-4 year old version (paired data) at baseline and outcome (n=212)



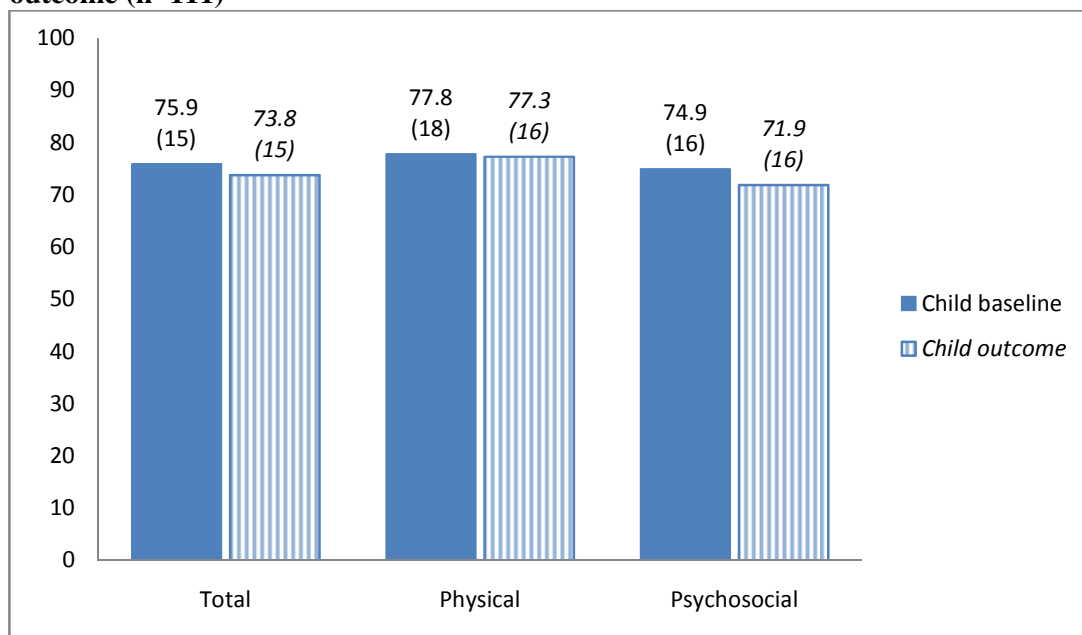
b) Parental ratings for 5-7 year old version (paired data) at baseline and outcome (n=74)



482 **c) Parental ratings for the 8-12 year old version (paired data) at baseline and**
 483 **outcome (n=15)**

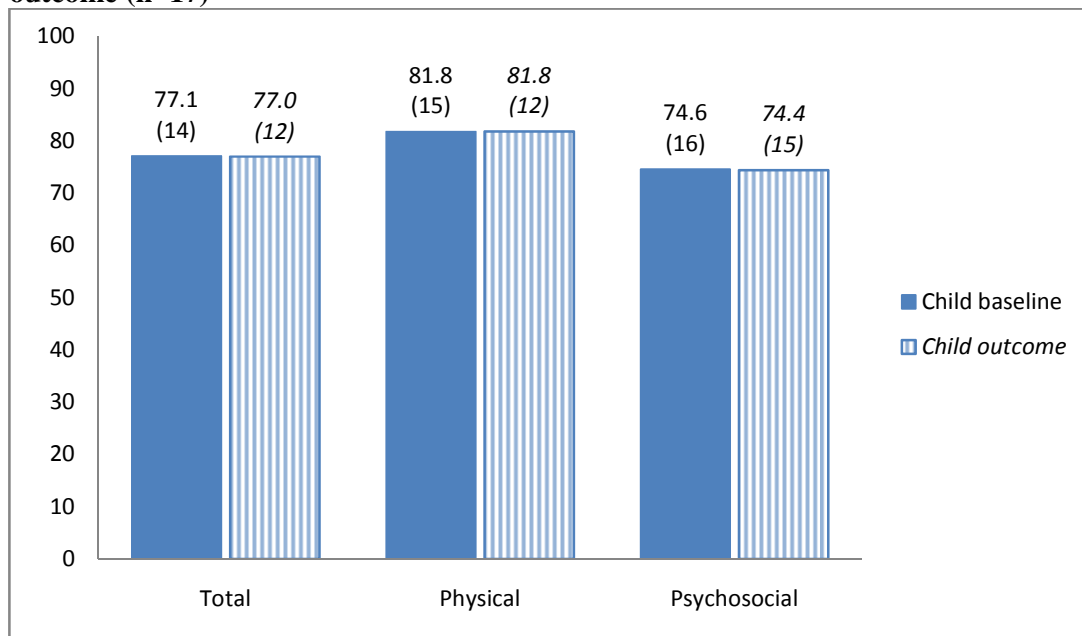


484
 485
 486
 487 **d) Child ratings for the 5-7 year old version (paired data) at baseline and**
 488 **outcome (n=111)**



489
 490
 491

492 **e) Child ratings for the 8-12 year old version (paired data) at baseline and**
 493 **outcome (n=17)**



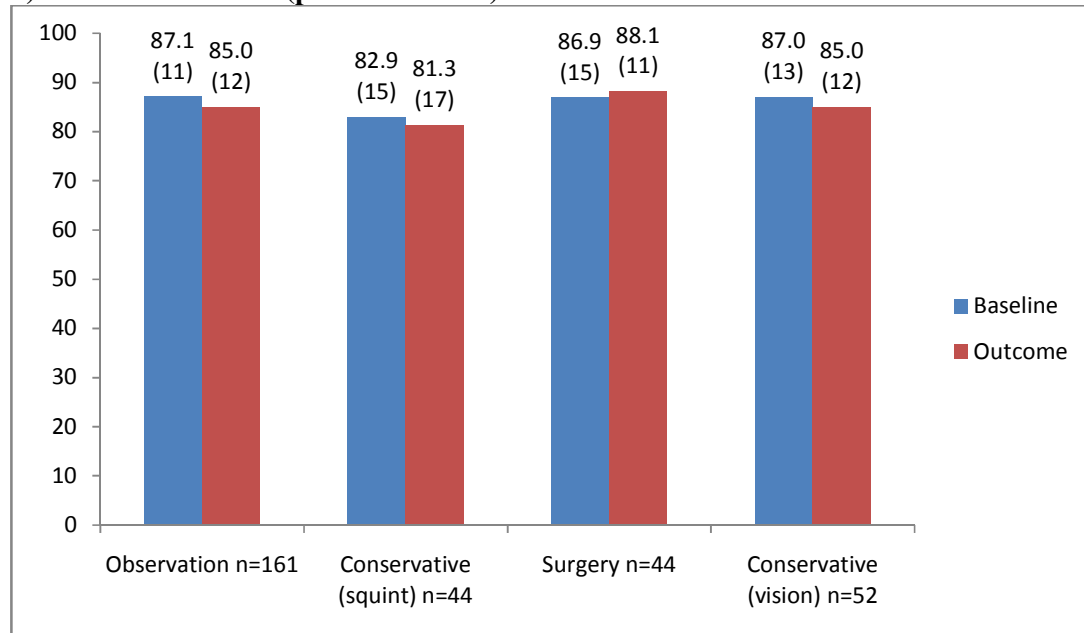
494
 495
 496
 497 **Treatment group**

498 There were no significant differences in parental or child-rated PedsQL™ scores by
 499 treatment group at either baseline or outcome. There was a mean improvement of 7, 4
 500 and 9 on child-rated Total, Physical and Psychosocial scores respectively for those
 501 who had surgery but these changes did not reach statistical significance (paired-
 502 samples t test $p=0.097$ to 0.151). Supplementary Figure 3 shows the mean Total and
 503 Summary scores at both time-points for those with paired data.
 504

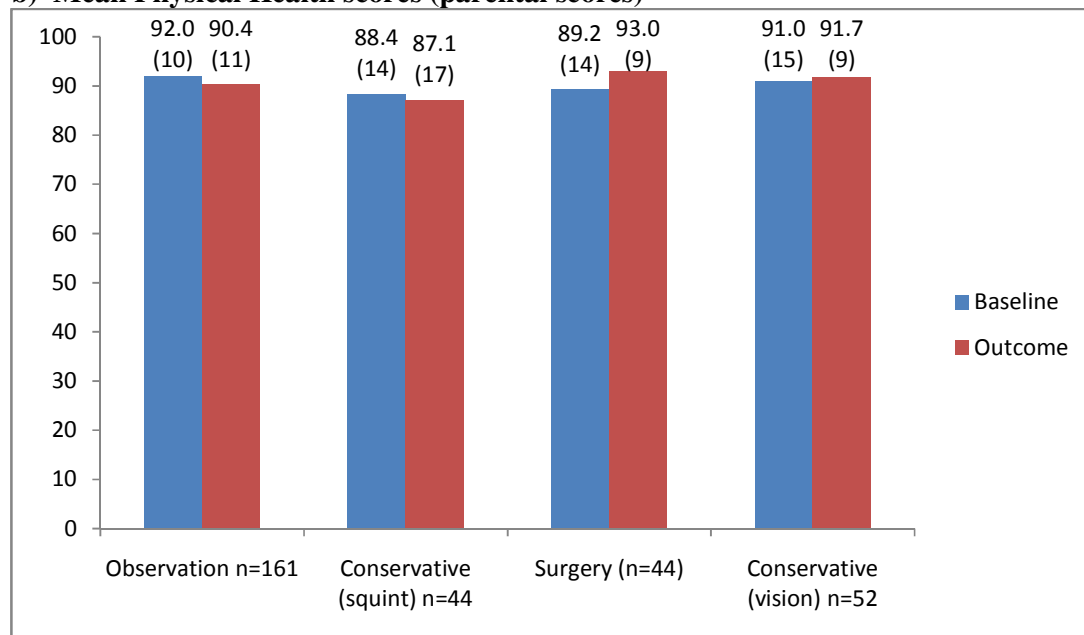
505
 506
 507

Supplementary Figure 3: Mean baseline and outcome PedsQL™ scores by treatment group

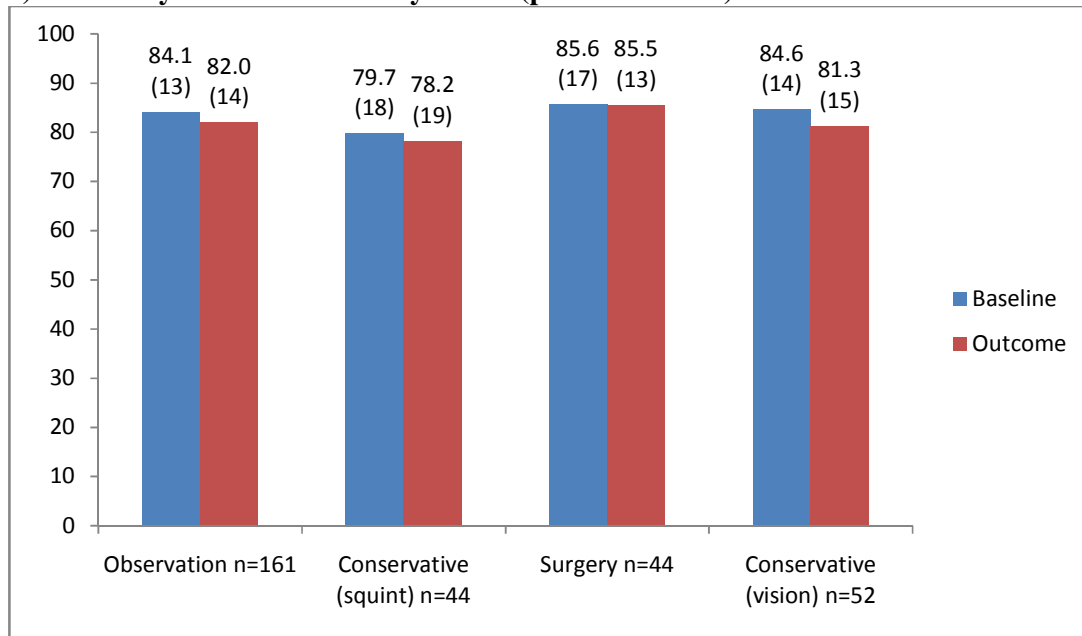
a) Mean total scores (parental scores)



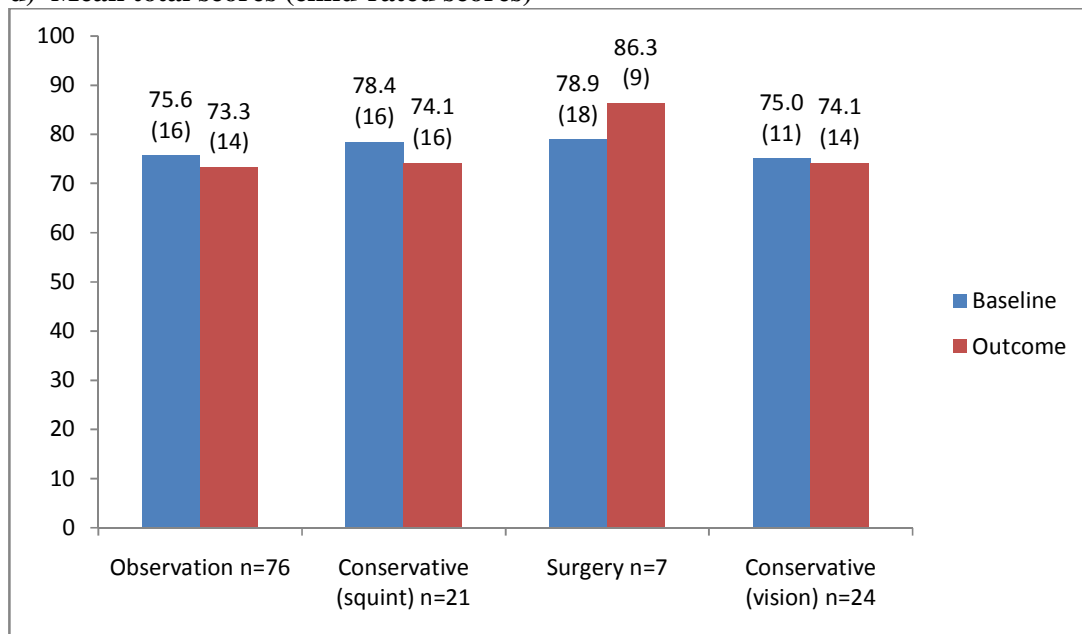
b) Mean Physical Health scores (parental scores)



519 **c) Mean Psychosocial Summary scores (parental scores)**

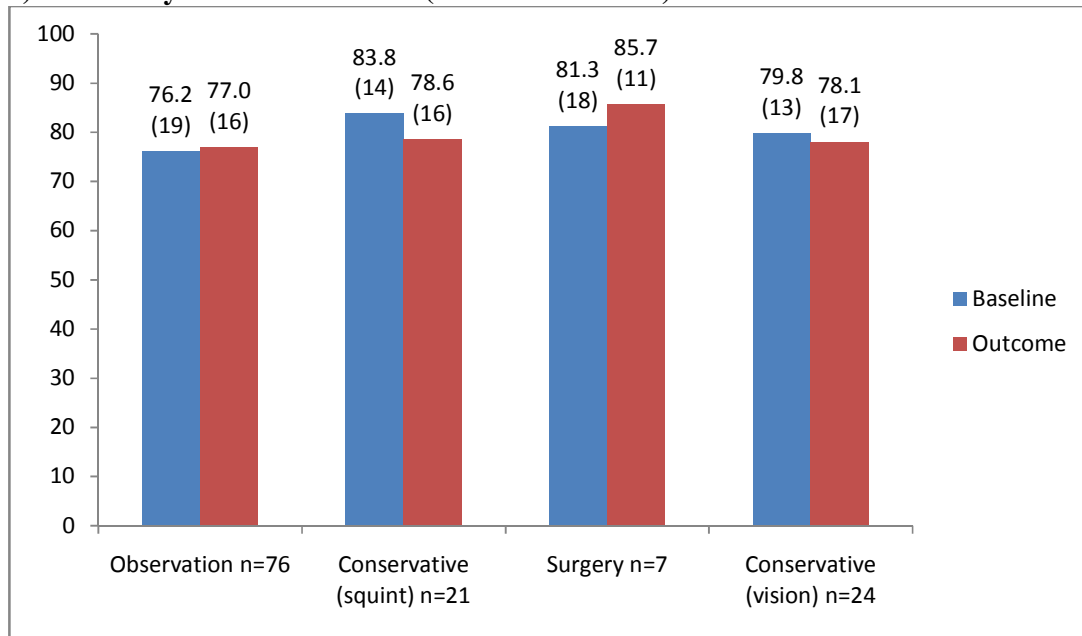


520
521
522
523 **d) Mean total scores (child-rated scores)**

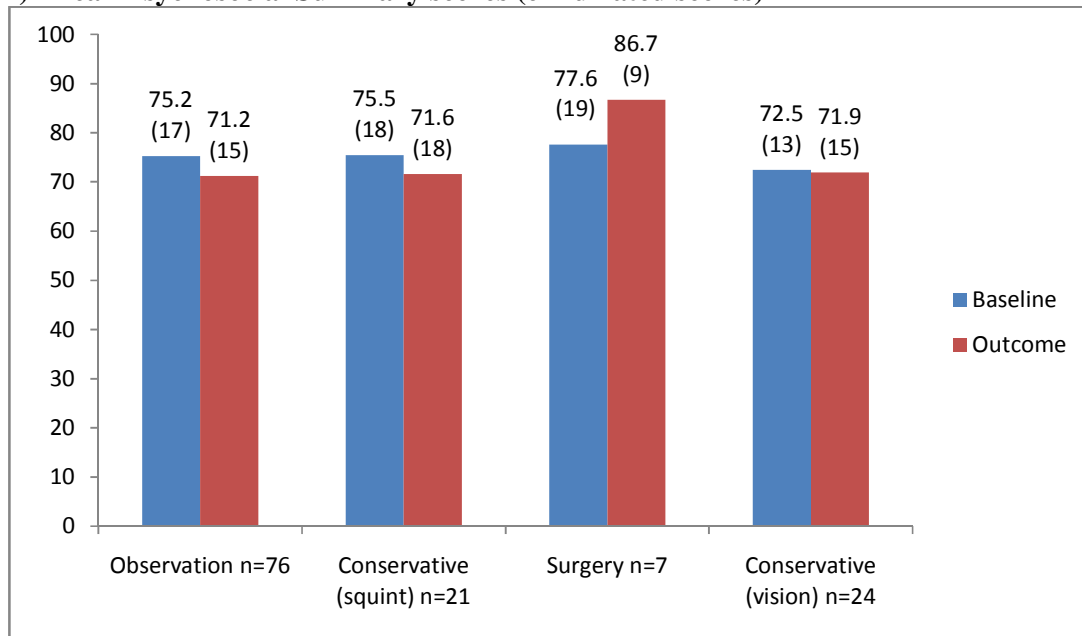


524
525
526

527 **e) Mean Physical Health scores (child-rated scores)**



528
529
530 **f) Mean Psychosocial Summary scores (child-rated scores)**



531
532
533

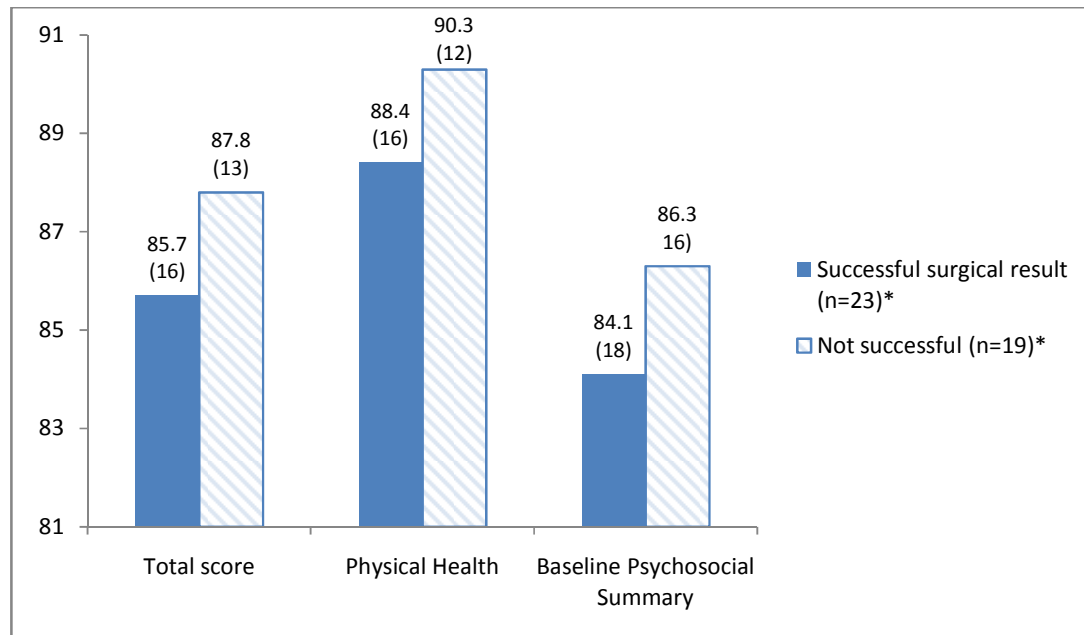
534 **Surgical success and parental PedsQL™ scores**
535

536 Within the surgery group specifically, those subsequently deemed to have a successful
537 surgical result (NCS 0-2) had lower mean parental scores at baseline on each
538 PedsQL™ scale (except Emotional Functioning) than those with less satisfactory
539 control scores or those who were overcorrected. Conversely, mean parental
540 PedsQL™ scores were higher at outcome on each scale (except Physical Health) for
541 those with a better surgical outcome than those with less satisfactory control scores or
542 those who were overcorrected. However, none of these differences were statistically
543 significant: mean (SD) Total score 89.2 (9.2) and 86.0 (12.5) respectively, $p=0.36$;
544 Physical Health 92.0 (9.5) and 93.5 (9.5), $p=0.62$; Emotional Functioning 84.5 (15.5)
545 and 80.3 (16.8), $p=0.42$; Social Functioning 94.8 (9.3) and 89.7 (15.8), $p=0.22$;
546 School Functioning 84.9 (14.7) and 75.5 (17.7), $p=0.08$; Psychosocial Summary 87.6
547 (10.5) and 82.3 (15.2), $p=0.19$. Supplementary Figure 4 illustrates Total and
548 Summary scores at baseline and outcome in terms of surgical success.

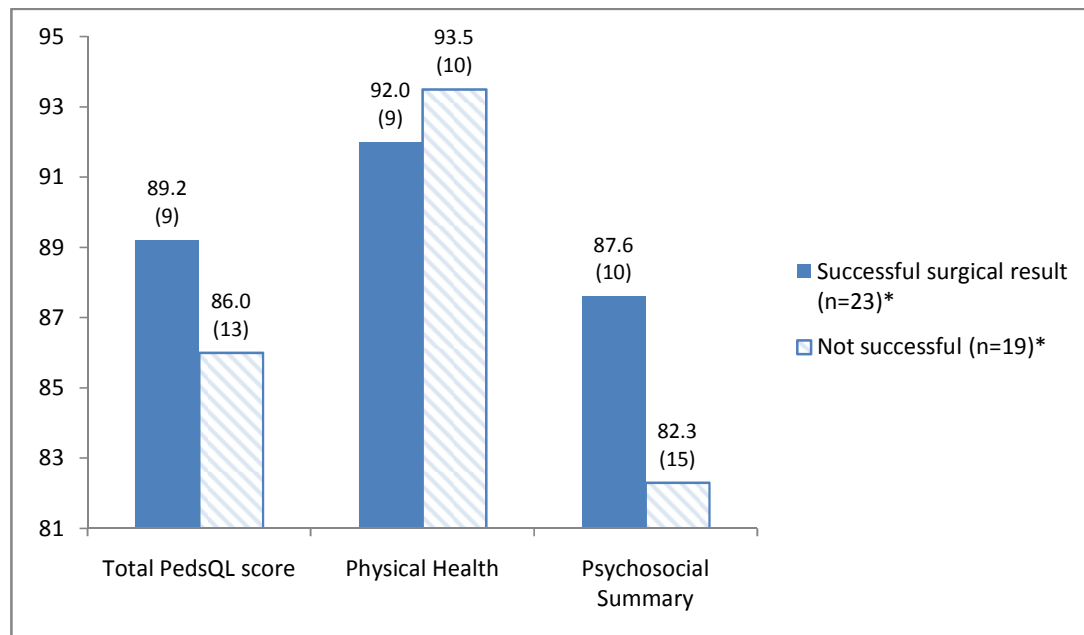
549
550

Supplementary Figure 4 – Surgical group only

a) Mean parental PedsQL™ scores at baseline* by surgical success



b) Mean parental PedsQL™ scores at outcome* by surgical success



* Total NCS at time of latest PedsQL™ assessment missing in 2 cases

Exploring the relationship between surgical success and child-rated scores is

problematic given that there are only 7 in this group, 4 of whom had a successful

surgical outcome.

567

568

569 **Comparison with healthy children**

570

571 As reported in the main Results, parents of 2 to 4 year-old children with X(T) rated
572 their child's QOL similarly to those of age-matched healthy children in terms of the
573 Total and two summary (Physical Health and Psychosocial Summary) scores. Closer
574 examination of individual sub-scales shows that X(T) parents' ratings were
575 significantly higher (better) than the healthy sample on Emotional Functioning (80.9
576 compared to 76.0, $p < 0.001$, mean difference 4.9, 95% CIs 2.6 to 7.1) and Social
577 Functioning (92.3 compared to 89.9, $p = 0.002$, mean difference 2.4, 95% CIs 0.9 to
578 3.9), although they were worse on Nursery Functioning (88.2 compared to 92.3,
579 $p < 0.001$, mean difference -4.1, 95% CIs -6.4 to -1.9).

580

581 **Squint severity**

582

583 Children with severe X(T) generally had poorer mean parent- and self-rated
584 PedsQLTM scores than those with mild or moderate X(T), but this was not statistically
585 significant. Details are provided in Supplementary Table 1.

586

587

588 **Supplementary Table 1 Oneway analysis of variance to examine relationship**
589 **between severity of X(T) and mean PedsQL™ scores**
590

		Mean (SD)	95% CIs for mean	p value**
Parental scores by X(T) severity*				
Total score	Mild X(T)	86.7 (12)	84.9 to 88.5	0.10
	Moderate X(T)	86.2 (12)	84.3 to 88.1	
	Severe X(T)	79.8 (20)	69.5 to 90.1	
Physical Health	Mild X(T)	91.4 (13)	89.5 to 93.2	0.14
	Moderate X(T)	90.2 (12)	88.3 to 92.0	
	Severe X(T)	85.1 (17)	76.3 to 93.9	
Emotional Functioning	Mild X(T)	79.3 (17)	76.8 to 81.8	0.17
	Moderate X(T)	77.5 (20)	74.5 to 80.6	
	Severe X(T)	70.7 (24)	58.4 to 83.0	
Social Functioning	Mild X(T)	88.6 (15)	86.3 to 90.8	0.15
	Moderate X(T)	89.7 (15)	87.5 to 91.9	
	Severe X(T)	82.1 (23)	70.2 to 93.9	
School/Nursery Functioning	Mild X(T)	83.9 (17)	81.2 to 86.5	0.13
	Moderate X(T)	83.8 (18)	80.6 to 86.9	
	Severe X(T)	73.6 (26)	58.0 to 89.2	
Psychosocial Summary score	Mild X(T)	83.9 (14)	81.9 to 85.9	0.13
	Moderate X(T)	83.8 (15)	81.5 to 86.1	
	Severe X(T)	76.5 (22)	65.0 to 87.9	
Child scores by X(T) severity[†]		Mean (SD)	95% CIs	p value**
Total score	Mild X(T)	75.5 (15)	72.4 to 78.7	0.84
	Moderate X(T)	75.8 (14)	72.1 to 79.6	
	Severe X(T)	71.7 (18)	49.8 to 93.7	
Physical Health	Mild X(T)	77.9 (18)	74.2 to 81.8	0.97
	Moderate X(T)	78.3 (17)	73.9 to 82.7	
	Severe X(T)	76.2 (16)	56.9 to 95.6	
			71.6 to 79.8	
			73.1 to 83.4	

Emotional Functioning	Mild X(T)	75.7 (19)	45.3 to 96.7	0.59
	Moderate X(T)	78.3 (19)		
	Severe X(T)	71.0 (21)		
Social Functioning	Mild X(T)	74.1 (23)	69.2 to 79.0	0.64
	Moderate X(T)	73.7 (23)	67.6 to 79.8	
	Severe X(T)	64.0 (21)	38.3 to 89.7	
School/Nursery Functioning	Mild X(T)	72.8 (19)	68.8 to 76.9	0.88
	Moderate X(T)	71.2 (20)	65.9 to 76.4	
	Severe X(T)	73.0 (34)	31.0 to 100	
Psychosocial Summary score	Mild X(T)	74.3 (16)	70.9 to 77.7	0.79
	Moderate X(T)	74.5 (16)	70.3 to 78.8	
	Severe X(T)	69.3 (20)	44.4 to 94.2	

* n=185 had mild X(T), 162 had moderate and 17 had severe X(T), as defined by the Newcastle Control Score (NCS) (categorised as: NCS 1 to 3 = mild; NCS 4 to 6 = moderate, and NCS 7 to 9 = severe). Total NCS missing in one case

† n=89 had mild X(T), 58 had moderate and 5 had severe X(T)

** Oneway anova