Monograph:
Ara, R., Duarte, A., Harnan, S. et al. (4 more authors) (2015) Supporting the collection patient reported outcome measures in the National Clinical Audits for assessing costeffectiveness Work Package 1. What patient reported outcome measures should be used in the 13 health conditions specified in the 2013/14 National Clinical Audit programme? APPENDIX C, INFLAMMATORY BOWEL DISEASE. Report. Policy Research Unit in Economic Evaluation of Health and Care Interventions (EEPRU) . Policy Research Unit in Economic Evaluation of Health and Care Interventions

---

EEPRU

Economic Evaluation of Health and Social Care Interventions Policy Research Unit

RESEARCH REPORT

Supporting the routine collection of patient reported outcome measures in the National Clinical Audits for assessing cost-effectiveness

Work Package 1

What patient reported outcome measures should be used in the 13 health conditions specified in the 2013/14 National Clinical Audit programme?

APPENDIX C, INFLAMMATORY BOWEL DISEASE

Authors: Roberta Ara, Ana Duarte, Sue Harnan, Jo Leaviss, Steve Palmer, Mark Sculpher, John Brazier

Correspondence to: Roberta Ara, HEDS, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA. Email r.m.ara@sheffield.ac.uk

RR0030

March 2015

The University Of Sheffield.

UNIVERSITY of York

The Policy Research Unit in Economic Evaluation of Health and Care interventions is funded by the Department of Health Policy Research Programme. It is a collaboration between researchers from the University of Sheffield and the University of York.

The Department of Health's Policy Research Unit in Economic Evaluation of Health and Care Interventions is a 7 year programme of work that started in January 2011. The unit is led by Professor John Brazier (Director, University of Sheffield) and Professor Mark Sculpher (Deputy Director, University of York) with the aim of assisting policy makers in the Department of Health to improve the allocation of resources in health and social care.

This is an independent report commissioned and funded by the Policy Research Programme in the Department of Health. The views expressed are not necessarily those of the Department.

INDEX

TABLES

FIGURES

Acronym	Definition
AE	Adverse events
ANOVA	Analysis of variance
CAI	Clinical activity index
CD	Crohn's disease
CDAI	Crohn's disease activity index
CDI	Children's depression inventory – Short form
CGQL	Cleveland Global Quality of Life
CHQ	Child Health Questionnaire
CHQ-CF87	Child Health Questionnaire – child form 87
CHQ-PF50	Child Health Questionnaire – Parent form 50
CHU-9D	Child Health Utility 9D
CLIQ	Crohn's Life Impact Questionnaire
DH	Department of Health
EMA	European Medicines Agency
EEPRU	Policy Research Unit in Economic Evaluation of Health and Care Interventions
EQ-5D (-3L)	EuroQol 5 dimensions
EQ-5D-5L	EuroQol-5D-5 levels
EQ-5D-Y	EuroQol-5D youth version
ES	Effect size
FR	Future research
GI	Gastrointestinal
HRQoL	Health related quality of life
HS	Health states
HTA	Health technology assessment
HUI2	Health Utility Index mark 2
IBD	Inflammatory bowel disease
IBDQ-9	Inflammatory bowel disease questionnaire-9
IBDQ-32	Inflammatory bowel disease questionnaire-32
IBDQOL	Inflammatory bowel disease quality of life questionnaire
IBDSI	Inflammatory bowel disease stress index
ICC	Intraclass correlation coefficient
ICF	World Health Organization's International Classification of Functioning, Disability and Health
MTA	Multiple technology assessment
N or n	Number
NCA	National Clinical Audit
NHP	Nottingham Health Profile
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
PCDAI	Paediatric Crohn's disease activity index
PedsQL	Paediatric quality of life inventory
PedsQL MFS	Paediatric quality of life inventory - Multidimensional Fatigue Scale
PedsQL GI	Paediatric quality of life inventory – gastrointestinal module
PR	Parent/carer report

PR	Potential recommendations
PRO2	Patient reported outcome 2
PRO3	Patient reported outcome 3
PROM(s)	patient reported outcome measure(s)
QALY	Quality adjusted life year
R&D	Research and development
RFIPC	Rating Form of IBD Patient Concerns
RCT(s)	Randomised controlled trial(s)
SF-36	Short form 36
SIBDQ	Short Inflammatory Bowel Disease Questionnaire
SR	Self report
SRM	Standardised response mean
STA	Single technology assessment
TA(s)	Technology Appraisal(s)
TAG	Technology Assessment group
TTO	Time trade off
UC	Ulcerative colitis
UK	United Kingdom
UK-IBD-Q	IBD-Control, quality of life
VAS	Visual analogue scale
WP	Work package
WPAI: CD	Work Productivity and Activity Impairment: Crohn's Disease

1. BACKGROUND

The Policy Research Unit in Economic Evaluation of Health and Care Interventions (EEPRU) was approached by Jason Cox (Research and Development (R&D) Division) to prepare a programme of research to support the appropriateness of, and use of, patient reported outcome measures (PROMs) collected for the National Clinical Audit (NCA). The EEPRU programme was informed by a R&D template prepared by Simon Bennett, Steve Fairman and Keith Willett at National Health Service (NHS) England.

The purpose of introducing PROMs into the NCA programme is to be able to 1) compare performance between providers and commissioners in the NHS, 2) compare the cost-effectiveness of alternative providers in delivering the specific services (i.e. linking outcomes and resource use), and 3) assess the cost-effectiveness of alternative interventions and other changes in the NHS. The intention is to introduce PROMs across a range of conditions over the next 3 years commencing with 13 conditions in the 2014/15 NCA programme.

The agreed research programme consists of 3 concurrent work packages (WP) as described in the document submitted to the Department of Health (DH) (8th November 2013). The current document provides details on the objectives, methodology and results for Work Package 1 (WP1): to determine what PROMS should be used in the 13 health conditions specified in the 2014/15 NCA programme.

2. OVERVIEW

WP1 is split into three separate components consisting of:

- WP1.1 To examine whether the EuroQol-5D (EQ-5D) is appropriate in the 13 health conditions specified in the 2013/14 NCA programme.
- WP1.2 To identify what measure could be used when the EQ-5D is not appropriate in the 13 health conditions, taking into account that the proposed measure would be used to generate preference-based utility measures (either directly through existing preference-based weights, or indirectly through existing mapping functions suitable for the proposed measure).
- WP1.3 To identify the evidence required to address questions of cost-effectiveness using the NCA data.

Each component consists of a series of reviews of the literature.

This Appendix provides the detailed results for the condition inflammatory bowel disease (IBD) and should be read in conjunction with both the main report and the methods/search strategy appendices.

3. METHOD

The full detailed methodology used is provided in Appendix A and B, including the search strategy, selection criteria for studies included, and data extraction etc. In summary, a review of the literature was undertaken to assess the appropriateness of the EQ-5D in terms of classic psychometric criteria (WP1.1); where the EQ-5D was not considered appropriate, additional searches were undertaken to identify alternative measures (WP1.2); and finally, existing health technology appriasials were reviewed and data requirements were compared with variables currently collected in the IBD audit (WP1.3).

3.1 **Psychometric properties** (WP1.1)

Assessments reported in the included studies were categorised according to the following definitions:

Acceptability

Data relating to how acceptable the measure was to the person completing it, expressed as the proportion of completed surveys, or the proportion of missing data.

Reliability

There are two main definitions for reliability, a) the degree to which a measure reproduces the same results in an unchanged population and b) the degree to which a measure reproduces the same results when completed by different assessors (e.g. patient and proxy report). In both cases, reliability can be assessed by re-testing, and calculating the correlations or difference between tests. In case a) the comparison may be between the same populations separated by time, where no change in health state was observed (as compared to using an alternative condition specific or generic measure). In case b) the measure may be completed by multiple people (proxies) on the patient's behalf and their responses compared with those of the patient. Where the outcome measure is specifically designed for self-report by patients, this test of reliability may be expected to produce less agreement.

Construct validity

This is an assessment of how well an instrument measures what it intends to measure. Two main definitions are used in this review.

a) Known group validity, where estimates for groups that are known to differ in a concept of interest are compared either qualitatively or statistically. The known groups may be defined using other measures, according to clinical categorisation.

b) Convergent validity assesses the extent to which a measure correlates with other measures of the same or similar concepts. Correlation coefficients were considered low if <0.3, moderate if between 0.3 and 0.5, and strong when >0.5.

Responsiveness

a) Change over time. This is an assessment of whether measurements using the instrument can detect a change over time, where a change is expected. This may be before and after an intervention, or through progression of a disease. Evidence was considered to be good where a t-test was significant, though weaker evidence to support responsiveness was considered where there was a change in the expected direction, but was not statistically significant or not tested. Effect size (ES) and standardised response mean (SRM) were also acceptable assessments of responsiveness.

b) Ceiling and floor effects were also considered to be indicators of responsiveness. Assessments of ceiling effects include the proportion of patients who score full health within a group of patients with known health detriments. A ceiling or floor effect can affect the sensitivity of the measure in detecting changes over time in patients at the extremes of the measure (for example those with severe disease activity and those with just minor symptoms of the condition).

3.2 **Alternative measures** (WP1.2)

As the IBD audit includes paediatrics and the EQ-5D is not designed for use in this population, alternative instruments were reviewed. This entailed both a review of existing guidelines, and a review of primary studies relating to the four prespecified paediatric measures (EQ-5D-Y, Child Health Utility 9D (CHU-9D), Health Utility Index mark 2 (HUI2), Paediatric quality of life inventory (PedsQL)(1-4)), as detailed in section 3.2, Appendix A.

3.3 **Evidence required for economic evaluations** (WP1.3)

The existing Health Technology Assessment (HTAs) were reviewed alongside the variables currently collected in the NCA to determine if clinical or PROM data routinely collected in the NCAs would

suffice to address questions of cost-effectiveness, and to identify any gaps in the evidence that would be required to compare providers, or the cost-effectiveness of interventions or policies.

4. RESULTS FOR INFLAMMATORY BOWEL DISEASE

4.1 Evidence of appropriateness of EQ-5D in Inflammatory Bowel Disease (IBD) (WP1.1)

As no existing review was known to the authors (see Appendix A), and no relevant systematic review was identified by Longworth et al. 2014a, (5)a search was performed to identify primary studies reporting the results of a psychometric assessment of the EQ-5D in patients with IBD (specifically, Crohn's disease (CD) or ulcerative colitis (UC)). The systematic review of primary studies is reported below. All associated Tables are provided in the Appendix.

4.1.1 Results of searches for primary psychometric studies in IBD

Searches for primary studies identified 89 unique titles of potential relevance. 75 studies were excluded upon examination of their title and abstract. The full text of 14 titles was obtained, of which two met the criteria for inclusion in this review. The excluded articles are listed in Appendix B with reasons for exclusion.

4.1.2 Studies included in the systematic review of primary studies for IBD

Two studies assessed the psychometric properties of the EQ-5D in IBD.(22;23) Both studies were performed in Germany and both recruited a mix of adult patients (N=152;(6) N=502(7)) with either CD (80.9%,(6) 53.8%(7)) or UC (19.1%,(6) 46.2%(7)). Konig et al. used the van der Schulenburg (1998) German EQ-5D tariff (range: 0–100) which is not comparable with the United Kingdom (UK) preference-based EQ-5D index.(8) Consequently only the results of the analyses involving the EQ-5D health dimensions (and not the EQ-5D index) are reported below.(6) Stark et al. used a more recent German EQ-5D tariff, and the standard UK EQ-5D tariff.(9;10) The results for analyses conducted on the EQ-5D health dimensions and UK EQ-5D index scores are reported below for this study.

Stark et al. compared the EQ-5D data with the inflammatory bowel disease questionnaire (IBDQ range: 32 (worst) to 224 (best)) and the Crohn's disease activity score (CDAI range: unbound with disease categories: ≤150 quiescent; >150 active; >450 extremely severe).(11;12) Stark et al. used the following sub-groups: CD (n=270) and UC (n=232); patients in remission (CD=57.1%, UC=62%) versus patients with active disease (CD=42.9%, UC=38%); patients who reported no change in health status over time (n=360) versus patients who reported a change in health status over time (worse health n=29, better health n=42) all sub-categorised by CD, UC, remission or active disease status.(7) Konig

et al. compared the EQ-5D with short form -36 (SF-36) health dimensions (physical functioning, role limitations, social functioning, pain, mental health; all range 0-100), the CDAI and the clinical activity index (CAI).(12;13) The latter two instruments are validated clinical measures in IBD and were used to sub-group patients by severity of symptoms or disease activity and changes over time.(12;13) Konig et al. used the following sub-groups in their analyses: inpatients (n=31) versus outpatients (n=113); patients with active disease (n=58) versus patients in remission (n=94); patients who reported no change in health status over time (n=52).(6) It should be noted that some of the sample sizes of the sub-groups were relatively small, and results from these comparisons may not be robust.

The cohorts in the two studies were similar in terms of age at diagnosis (all were adults ≥17 years old), duration of disease, number in remission and number with active disease (Table A2, Appendix). Small numbers of patients were missing from some analyses in both studies.

4.1.3 Psychometric properties reported in the primary studies for IBD

All evidence related to adults (≥17 years of age).

Acceptability: Konig et al. examined the acceptability of the EQ-5D by assessing the proportion of missing responses and found it to be well accepted with less than 3.5% of EQ-5D responses missing (compared to 9% for the IBDQ)(Table A4, Appendix).(6) Stark et al. did not assess acceptability as the sample analysed only included respondents who completed the EQ-5D (N=502).(7)

Reliability: Both studies reported results for test/retest reliability for patients who reported no change in health status over time (Table A4, Appendix). The proportion of agreement in outpatients who reported no change (n=52/66) over an average period of 19.5 days ranged from 80.4% for the health dimension anxiety/discomfort (kappa=0.61, p-values not reported) to 100% for the health dimension self-care (kappa=1.0).(6) Similar results were reported for sub-groups of patients (CD n=195, UC n=166) who reported no change at follow-up in the second study with substantial or almost perfect agreement for all five health dimensions (kappa>0.60, p-values not reported).(7)

Construct validity (known group): When comparing the results for outpatients (n=120) and inpatients (n=31), there was no significant difference in responses for the health dimensions pain/discomfort (p=0.5994) or anxiety/depression (p=0.4394). Assuming outpatients would have a better health related quality of life (HRQoL) than inpatients this result is not as expected. However,

the number of inpatients was small, possibly resulting in underpowering.(6) When comparing patients with UC who were in remission with those with active disease, the mean EQ-5D preference based scores and the responses for the health dimensions were significantly better for those in remission ($p<0.001$), with the exception of the responses for the dimension self-care ($p=0.20$).(7) Comparing sub-groups, the responses for all five health dimensions were significantly better for those in remission compared to those with active disease. For example, 96.1% of patients in remission indicated they had no problem with mobility compared to 66.0% of patients with active disease, $p<0.0001$).(6)

Construct validity (convergent): Correlation between the EQ-5D preference-based index and the activity indices were good (Spearman rank: $r=-0.75$ for CDAI, $r=-0.65$ for CAI, $p<0.0001$) (Appendix).(7) Assessing the median responses on the SF-36 and IBDQ dimensions with the median responses on comparable EQ-5D health dimensions (for example physical functioning on the SF-36 was compared with mobility on the EQ-5D, emotional functioning on the IBDQ was compared with anxiety/depression on the EQ-5D), with the exception of the EQ-5D dimension self-care, the results were ordered appropriately according to the responses on the EQ-5D dimensions and results were significantly different between the groups ($p<0.0001$) (Table A4, Appendix).(6) Stark did not assess this property.(7)

Responsiveness (change over time): When assessing responsiveness to change in health status over time (Table A5, Appendix), statistically significant changes in the expected direction were observed for the EQ-5D preference-based index (change in mean EQ-5D: -0.09, $p<0.05$ for worse health ($n=26$); 0.095, $p<0.0001$ for improved health ($n=41$)).(7) Stark et al. also performed a number of sub-group analyses, including analysing patients with CD, UC, those with active disease, or those who were in remission separately. In these sub-group analyses, the direction of effect was as expected, but not all sub-group analyses were statistically significant. The results for all sub-groups who reported an improvement in health were statistically significant (except for the remission sub-group ($n=19$)), though again the small numbers may cause underpowering. Of the results for the sub-groups who reported a deterioration in health, the analyses for patients in remission ($n=9$), and those with UC ($n=16$) were statistically significant ($p<0.05$).

In those reporting an improvement in health, the direction of effect was as expected in all groups, but a strong effect (defined as SRM >0.8) was only seen in UC ($n=20$, SRM=0.83) and those in remission ($n=19$, SRM=1.10). Similarly, for those who reported a deterioration in health, all effect

sizes were in the expected direction, but only the results for patients in remission (n=9, SRM=-0.82) and those with UC (n=10, SRM=-0.84) showed a strong effect.

Responsiveness (ceiling effect): Both studies assessed potential ceiling effects on the EQ-5D (Table A5, Appendix). Some evidence of a ceiling effect was reported with 19.4% and 30.1% of inpatients and outpatients respectively scoring no problems on any of the five health dimensions,(6) and 43% and 31% for patients with UC or CD respectively scoring full health on the EQ-5D preference-based index (EQ-5D=1).(7) These results indicate the EQ-5D may not be able to discriminate between different health states in patients with less severe disease or who are in remission, and may be less responsive to improvements over time in these groups. However, it was noted in both studies that the distribution of patients across disease severity was skewed towards those who were less severely affected by the conditions; for example the median IBDQ total score was 174 (observed range 66-224) in one study,(6) and just 0.4% and 2.2% of patients with CD and UC respectively were reported as having severe disease activity in the second study.(7) As such, the studies may be biased towards overestimating a ceiling effect in a routine NHS IBD population.

4.1.4 Conclusion of appropriateness of EQ-5D in IBD

The evidence base assessing the performance of the EQ-5D in IBD is currently small, with only two studies satisfying the inclusion criteria for this review and only one of these assessing EQ-5D UK tariff data.(7) All studies were in adults. The EQ-5D was well accepted with less than 3.5% of EQ-5D responses missing (compared to 9% for the IBDQ). Construct validity was generally good on all EQ-5D health dimensions (except self-care in patients with UC) when assessed against the disease activity indices, and when discriminating between those with active disease and those in remission. Some problems were observed in the health dimensions pain/discomfort and anxiety/depression for outpatients relative to inpatients but the sample sizes were very small for some comparisons. There was some evidence of the EQ-5D being a responsive measure in those with active disease (further away from full health) who reported a deterioration in health. Lack of statistically significant results suggest there may be problems with the EQ-5D's responsiveness when detecting changes in less severely ill patients and those in remission. Small numbers of patients in the sub-group analyses used to assess responsiveness made it difficult to draw robust conclusions, however, ceiling effects were also observed. Stark et al. did not include a representative proportion of severely ill patients, and the authors recommended that further studies are conducted in this group of patients.(7)

In conclusion, the evidence supporting the appropriateness of the EQ-5D in adults with IBD is good but limited and there is a lack of evidence relating to how appropriate the EQ-5D is for surgical health states (Table 1). Evidence is required in paediatrics, and in patients with more severe disease, both of which are relevant for the IBD NCA population (see Section 3.2).

4.2 **Alternative measures in IBD** (WP1.2)

The IBD NCA includes paediatrics as well as adults and the EQ-5D adult version is not recommended for use in paediatrics. There is a youth version of the EQ-5D called the EQ-5D-Y, which is based on a minor modification of the wording of the items in the original measure. However, this is a relatively new version and is not widely used as yet. Consequently, searches were conducted to identify alternative measures for use in paediatrics and adults. This comprised a search for existing guidelines and recommendations (section 4.2.1) and a separate search for primary studies relating to EQ-5D-Y, CHU-9D, HUI2, and PedsQL (section 4.2.2).(1-4)

4.2.1 Alternative measures for IBD

Eleven documents were identified by the initial searches as described in Section 3 and Appendix A. Four documents were from the Royal College of Paediatrics and Child Health, describing various results from the IBD NCA reviewed below (Section 4.3), and thus were excluded from further analysis.(14-17) A further five documents related to irritable bowel syndrome and were excluded,(18-22) leaving two reports of relevance to this analysis.(23;24) One was a research guideline from the European Medicines Agency (EMA), produced by an expert panel with a stakeholder consultation period.(23) The report related to CD, and contained a very brief section on paediatrics where the paediatric version of the CDAI (the PCDAI) was recommended as a measurement of efficacy.(23) The final and most recent report (2014) provided a synopsis of the literature available for a particular pharmaceutical intervention in severe, active UC. This document listed several clinical and HRQoL measures used in randomised controlled trials (RCTs) in adults (no paediatric evidence) with UC but no attempt was made to evaluate these.(24)

In addition to the reports described in the previous paragraph, evidence presented in three manuscripts known to the authors is worthy of consideration.(25-27) While these sources are of relevance, it should be noted that they were not found through a systematic search process, and consequently other relevant evidence may not have been identified.

The first source of evidence is a systematic literature review which compared the contents of IBD specific PROMs relative to the World Health Organization's International Classification of Functioning, Disability and Health (ICF).(26) Four databases (Medline®, EMBASE, PsycINFO, CINAHL and CENTRAL by MC) were searched (terms: 'Colitis, Ulcerative', 'Crohn Disease', 'Inflammatory Bowel Disease' in title and abstract, and MESH terms (if available)). Articles were limited to those written in English which were published between 1999 and 2009. 46 studies were selected from the initial hits identified (n=9728). These studies used eight IBD-specific PROMs including: Cleveland Global Quality of Life (Faszio Score) (CGQL),(28) Inflammatory Bowel Disease Quality of Life Questionnaire (IBDQOL),(29) Inflammatory bowel disease questionnaire (IBDQ-32),(30) Inflammatory bowel disease stress index (IBDSI),(31) Inflammatory Bowel Disease Questionnaire - short form (IBDQ-9),(32;33) Rating Form of IBD Patient Concerns (RFIPC),(34) Short Inflammatory Bowel Disease Questionnaire (SIBDQ),(35)and Work Productivity and Activity Impairment: Crohn's Disease (WPAI: CD).(36) The IBDQ-32,(30) was used in 60% (36/46) of the studies with the next most common being IBDQOL (6/46), followed by the RFIPC (5/46) and SIBDQ (5/46). The measures were not reviewed in any detail other than their ability to capture ICF and they cannot currently be used to generate utility values.

The second source of evidence is a conference abstract which presented an exploratory analysis examining whether items on the CDAI (pain, stool frequency, general well-being, recorded over a 7 day period) diary card (recorded over 7 days) could be adapted to be used as a PROM in adults with Crohn's disease.(25) Two potential measures were explored: the Patient Reported Outcome-2 (PRO2) score (range: 0-35) which used abdominal pain and stool frequency; and the Patient Reported Outcome-3 (PRO3) score (range: 0-50) which used abdominal pain, stool frequency and general well-being. While only summary results are currently in the public domain, these are promising. Estimates of effects were similar in the PROMs and the full CDAI scores (range: 0-400) across various sub-groups in an RCT. The PRO3 appeared to have a slightly higher correlation with the end of study CDAI score (r=0.89 vs. r=0.76), and change in scores (r=0.71 vs. 0.51) than the PRO2.

The third source of evidence was produced by the UK IBD Standards Group in 2013, and draws on data collected in the first three rounds of the UK IBD Audit, the National Royal College of Nursing Inflammatory Bowel Disease Nursing Audit and other national reports funded by government and charities. This report provides some information relating to the newly formed UK IBD register (June 2013) which is a repository of anonymised data from adult and paediatric patients with IBD.(37) These data are to be used for prospective audit and research purposes and it is believed the register

will be linked to the current IBD audit. The priorities identified by the authors included: 'Developing agreed clinical and patient-reported outcome measures that will support ongoing monitoring of the quality, safety and cost-effectiveness of IBD services.' The authors also referred to two PROMs-related projects (discussed below) which were ongoing at the time of publication.(38;39)

The first project involved the IBD-Control©, a PROM which was developed to capture outcome data in adult patients with IBD in a busy clinical setting.(38) The objective was to inform face-to-face consultations, or telephone/virtual clinics. Using a recall period of the previous two weeks, and 13 questions, the measure captures treatment concerns and the physical, emotional and social impact of IBD. Responses to eight of the questions provide an overall summary score (IBD-Control-8 range: 0-16 where 0=worse control). The measure has been validated, using the EQ-5D, the IBD-Control, quality of life questionnaire (UK-IBD-Q), and several clinical measures, in a study including adults with CD (n=160) and UC (n=139). Bodger et al. concluded the IBD-Control was a rapid (mean completion time =1 min 15 secs), reliable (Intraclass correlation coefficient (ICC) test-retest using 2 week repeat ≥0.96), valid (moderate to strong correlations with UK-IBD-Q and EQ-5D (r =0.52-0.86), analysis of variance (ANOVA) for detecting difference in remission, mild, moderate or severe disease ($p<0.001$)), and sensitive (ES: 0.76-1.44) measure.(38)

The second project involved the Crohn's Life Impact Questionnaire (CLIQ), designed to capture the impact of Crohn's disease from the patient's perspective, and the first PROM specifically designed for adult patients with CD.(39) The CLIQ consists of 27 items relating to quality of life and 9 relating to activity limitations; producing two corresponding unidimensional scales (range: 0-25 with 25 being the most severely affected). In a validation survey (n=273) using the Nottingham Health Profile (NHP) measure as a comparator, internal consistency was reported to be good and both CLIQ scales were shown to distinguish between self-reported severity of CD, general-health and work status. While both CLIQ scales correlated with the NHP ($p<0.01$), the relationship was strongest between the CLIQ QoL and the NHP Emotional reaction scores (r=0.80); and the CLIQ activity limitation and the NHP physical mobility scores (r=0.65).

4.2.2 Measures for paediatrics with IBD

As the results of the review above (Section 4.2.1), and the review performed under WP1.1 (Section 3.1) did not provide conclusive evidence to support the use of a particular measure in paediatrics, an additional search was conducted with the aim of identifying any literature describing the psychometric properties of the four pre-specified preference-based measures in IBD.

Results of searches: The searches identified 49 unique references. Full papers of 10 studies identified as being potentially relevant when the abstract and title were assessed against the inclusion criteria were obtained. An additional manuscript was identified from the references lists of these 10 papers,(40) Of the 11 papers, one did not present the psychometric properties of any measure,(41) and two did not report sufficient details for the IBD subgroup (Table A6, Appendix).(42;43) A total of eight articles, representing seven studies, satisfied the inclusion criteria and were included in this review.(40;44-50)

Results of review: Across the seven included studies, the psychometric properties of both IBD-specific measures and generic measures were reported (Appendix). No studies were identified which presented evidence relating to the use of the CHU-9D, HUI2, or EQ5D-Y in paediatrics with IBD.

Abdovic et al. assessed the reliability and validity of the IMPACT-III questionnaire in Croatian children with IBD.(44) While primarily examining the translated version of the IMPACT-III questionnaire, as the comparator was the generic PedsQL™ measure, the results are potentially relevant for other settings. Duffy 2011, described several HRQoL measures used in children with either juvenile idiopathic arthritis or IBD but provided limited information on the psychometric properties of the instruments.(45) Marcus et al. examined fatigue and HRQoL (IMPACT-III, PedsQL^{TM}) in children (aged 10-17 years) with a clinical diagnosis of IBD (CD=52, UC=13) compared with healthy controls (n=157).(46) Perrin et al. evaluated the IMPACT Questionnaire against the PedsQL v4 in children (aged 8-18 years) with either UC (n=59) or CD (n=161), using the Paediatric Crohn's disease activity index (PCDAI) to categorise severity.(47) Upton assessed the UK translation of the PedsQL^{TM} v4 in a relatively large sample of children (n=1,399) with IBD and their parents (n=970).(48) Ogden used a sample of children with IBD (n=97) to assess the psychometric properties of the IMPACT-III (UK), using the Child Health Questionnaire (CHQ) as the comparator.(40) Finally, Varni et al. and Lane et al. (conference abstract only) presented results from the same study which described a de-novo gastrointestinal (GI) module for the PedsQL. The module was compared with the PedsQL^{TM} v4 and was assessed in children (aged 3-18 years) with either CD (n=192) or UC (n=67).(49;50)

No studies were identified in this population for the three paediatric generic preference-based measures (EQ-5D-Y, CHU-9D, HUI2). A brief synopsis of the measures used and the psychometric properties of the measures reported within the seven studies are provided below, sub-grouped by

IBD-specific measures and generic HRQoL measures. It should be noted that not all these measures are PROMs, and the clinician reported measures (generally used as a comparator for the HRQoL measures, or to define severity-based subgroups) are included for completion only.

IBD-specific measures used in paediatrics with IBD

Two IBD-specific measures that had been used in paediatrics with IBD were identified: the IMPACT questionnaire (available as versions I, II and III)(51-53) and the Paediatric Crohn's Disease Activity Index (PCDAI).(54)

The *IMPACT Questionnaire* is a self-report measure developed for children (9 years and over) with IBD.(51;52) The IMPACT-III version (an update of versions I and II) includes 35 questions covering six domains: bowel symptoms, systemic symptoms, social functioning, body image, treatment/interventions, and emotional functioning. The responses (each question has a five-point Likert scale response) are used to generate a total score (range 35 to 175), where higher values indicate a better quality of life.(45) The IMPACT-III does not have a preference-based tariff and thus cannot be used to generate quality adjusted life years (QALYs) in economic evalutions without a validated relationship with a preference-based measure. Three studies included in this review reported data relating to the psychometric properties of IMPACT III.(40;44;45) Two studies included in this review reported data relating to the psychometric properties of IMPACT II but as this version is superceded by IMPACT III, the evidence is not reported here.(45;47)

Acceptability: Abdovic reported acceptability was good with all patients (n=104) completing the IMPACT-III questionnaire compared to 94% of patients completing the PedsQL^{TM} v4.(44) Similarly, 93/94 of children in the Ogden study believed the questions were easy to understand, and 87/94 indicated, if requested, they would complete the IMPACT-III again.(40)

Reliability: Test-retest reliability (n=50) over a period of 4-8 weeks was reported to be good in the validation study (ICC between 0.66 and 0.84).(40) Internal reliability was reported to be very good (Cronbach's α=0.74–0.88) for all domains in the validation study.(40) Internal consistency was high for the IMPACT-III total score (Cronbach's α=0.91) and with the exception of the treatment interventions sub-score (Cronbach's α=0.33) all reliability coefficients were greater than 0.61 in a second study.(44)

Construct validity (known group): Significant differences across sub-groups categorised by severity (severe vs. moderate vs. inactive/mild symptoms) groups were observed for the embarrassment scale (63.7* vs. 81.0 vs. 81.2, *p<0.05), symptom scale (45.0** vs. 64.2* vs. 80.6, *p<0.05, **p<0.01), and the energy scale (46.4* vs. 62.1* vs. 77.7, *p<0.05) in the validation study.(40) While the IMPACT-III was shown to distinguish between those with active (n=45) and inactive (n=59) disease, the difference in mean scores when comparing moderate/severe activity (n=8) with mild activity (n=37) was not statistically significant (p>0.05) in a second study.(44) However, these sample sizes were extremely small.

Construct validity (convergent): Using comparable domains on the IMPACT-III and the Child Health Questionnaire (CHQ), convergent validity was confirmed in the validation study with significant correlations (p<0.001) for all comparisons presented (e.g. energy compared with physical function, r=0.63, p<0.001).(40) With the exception of the associations between the IMPACT-III domain worry about stool and the PedsQL^{TM} emotional, social, psychosocial and school subscales (r≤0.305 for all), the correlations between the IMPACT-III and PedsQL domains were moderate to strong (r>0.4, p<0.001 for all).(44) Strong relationships were reported between the total IMPACT-III scores and self-reported PedsQL^{TM} v4 scores (r=0.74, p<0.001), the PedsQL^{TM} fatigue scores (r=0.63, p<0.001), and the PCDAI scores (r=-0.52, p<0.0001).(46)

The Paediatric Crohn's Disease Activity Index (PCDAI)(54) used in Marcus(46) and Perrin(47). The PCDAI is a clinician-completed measure used to determine the severity of the condition which uses a combination of clinical history, physical examination and laboratory results. Possible scores range from 0 to 100, with larger scores indicating more disease activity. It has been reported to have good reliability, and to be responsive to change in paediatric patients with CD.(55;56) The PCDAI cannot be used as a PROM at the moment. While it is possible that the exploratory research exploring the feasability of adapting items in the adult CDAI to use as a PROM (PRO2 or PRO3) may be exended to the paediatric version (see Section 4.2.1), this is unlikely to happen in the near future. In addition, the PCDAI cannot be used to generate utility scores without a validated relationship with a preference-based measure.

Evidence on generic- measures used in paediatrics with IBD

Amongst the included studies, data relating to the psychometric properties of three generic measures (with three variations of one measure) were identified. These were: children's depression inventory – short-form (CDI); child health questionnaire (CHQ); the paediatric quality of life inventory (PedsQL); The PedsQL Multidimensional Fatigue Scale (PedsQL MFS); and the PedsQL gastrointestinal (PedsQL GI) module.

Children's Depression Inventory – short-form (CDI)(57) was used in Marcus et al., and referred to in Duffy 2011.(45;46) The CDI is a self-report screening measure for symptoms of depression in children aged 7-17 years. Items relate to thoughts and behaviours over the previous two weeks and total scores range from 36-100 with scores greater than 63 being indicative of clinically significant symptoms of depression. There was no significant difference in mean scores (44.8 vs. 43.9) when comparing patients with IBD (n=70) and healthy controls (n=157), or in the percentage of respondents with clinically significant depressive symptoms (1.4% vs. 1.3%). However, an inverse relationship was shown between the CDI short-form scores and the PedsQL MFS, indicating a direct relationship between fatigue and symptoms of depression.(46) The CDI does not measure all aspects of HRQoL and cannot be used to generate QALYs in economic evalutions.

Child Health Questionnaire (CHQ)(58) used in Ogden

Originally designed by Landgraf in the 1990's, the CHQ is designed to measure wellbeing, functional health status, and health outcomes in children (4-19 years) and is a widely used and accepted measure. The parent/proxy (CHQ-PF50)(59) is the most widely used version while the Child Health Questionnaire – child form 87 (CHQ-CF87) is completed by adolescents (age 10-19 years).(60) The questionnaire includes 87 items covering the following domains: behaviour, bodily pain, general health, mental health, physical functioning, parent impact-time, parent impact-emotional, role-emotional⁄behavioural, role-physical, and self-esteem. There are two overall summary scores: physical and psychosocial (range 0-100 with 100 being better health). This measure was used as the comparator in the study by Ogden et al. when assessing the psychometric properties of the IMPACT-III (UK), as discussed above. As the literature searches were not designed to identify evidence for this measure, and no evidence was found on the psychometric properties, additional searches and a review of all evidence on the measure used in IBD would be required before it could be recommended for inclusion in the NCA.

The Paediatric Quality of Life Inventory (PedsQL)(49)

Data relating to the psychometric properties of PedsQL were reported in Abdovic et al., Duffy et al., Marcus et al., Perrin et al, and Upton et al. The PedsQL, a generic measure of HRQoL, has been reported as being one of the most thoroughly developed measures available for paediatrics.(60) It takes 4 minutes to complete and is either self-completed (5-18 years), or completed by a parent/caregiver (2-18 years), and comes in 3 forms designed for the patients age (5-7, 8-12, 13-18 years). The measure (version 4) covers 23 items describing four domains: emotional (5 items), social (5 items), physical (8 items), and school (5 items). Items are answered on a five-point Likert scale (0=''never a problem'' to 4=''almost always a problem''). The scores from these are used to derive summary scores in physical health (8 items) and psychosocial health (15 items), and an overall total score. These are all standardised (0-100) where higher scores indicate better HRQoL. The PedsQL does not have a preference-based tariff which could be used to generate QALYs in economic evaluations, and the main researcher and originator of the measure has no immediate plans to conduct research in this area.(49) [personal communication, Varni June 2014]

Acceptability: Acceptance was very good in one study (CD=74, UC=30) with just 5.8% (6/104) of children (aged ≥9 years) not completing the PedsQL v4 compared to 100% who completed the IMPACT III questionnaire.(44)

Reliability: Internal consistency was good for the PedsQL total score (cronbach's α=0.91,(44) α=0.89(46), and summary scores ($\alpha \geq 0.74$ (44), $\alpha \geq 0.90$.(46) Internal reliability of the UK version of the PedsQL sub-scales was reported to exceed 0.70 for self-report (total n=1,399, IBD n=76) and proxy-reports.(total n=970, IBD n=87).(48)

Construct validity (convergent): Comparing domains on the PedsQL with similar domains on the IMPACT III questionnaire, the correlations were strong ($r \geq 0.5$, $p<0.001$) or moderate ($r \geq 0.3$) for the majority of comparisons.(44) Weak correlations included: PedsQL school vs. IMPACT III concerns ($r=0.269$, $p<0.01$), PedsQL school vs. IMPACT III worry about stool ($r=0.286$, $p<0.01$), PedsQL™ emotional vs. IMPACT III worry about stool ($r=0.262$, $p<0.01$), PedsQL social vs. IMPACT III worry about stool ($r=0.206$, $p<0.042$).(44) The PedsQL™ total and subscales were also correlated with IBD specific factors on the Impact questionnaire (n=220), namely the well-being symptoms ($r>0.52$, $p<0.001$), and the total scale score ($r>0.54$, $p<0.001$). The relationship was less strong (but significant, $p<0.001$) for the Impact questionnaire factors: emotional functioning ($0.46 \leq r \leq 0.64$),

social interactions (0.37≤r≤0.49), and body image (0.36≤r≤0.51).(47) The PedsQL was also able to detect CD activity, as evidenced by the relationships with the PCDAI scores (total score: r=0.52 p<0.0001; physical health: r=0.47 p<0.001); psychosocial health: r=0.51 p<0.0001; emotional functioning: r=0.46 p<0.001; school functioning: r=0.53 p<0.0001).(46)

Construct validity (known group): The total PedsQL score was statistically significantly lower in patients with IBD (n=70), compared with healthy controls (n=157) (76.69 vs. 85.93, p<0.0001) and the ESs for the PedsQL 4.0 dimensions were reported to range from small (emotional functioning, ES=0.32; social functioning, ES=0.30) to large (total score, ES=0.89; physical health, ES=1.50; school functioning, ES=1.13).(45;46) Similarly, the mean scores on the PedsQL total score, and the sub-scales were lower in children with IBD (n=76) than in healthy controls (n=1032) for both self-report and proxy-report.(48) With the exception of the self-report social functioning scale, all differences were statistically significant (p<0.05).(48)

The PedsQL Multidimensional Fatigue Scale,(51) was used in Marcus et al., and referred to in Duffy et al.(45;46)

The PedsQL MFS was designed to measure the perception of fatigue in children, and has been validated in paediatric patients with cancer and rheumatological diseases.(51) With 18 items in total, it can be completed in less than 5 minutes, has self-report and parent-proxy versions (8-12 years), plus an adolescent version (13-18 years). The measure describes three fatigue domains including general (6 items), sleep/rest (6 items) and cognitive (6 items). The scores are transformed onto a 0-100 scale with higher scores indicating less fatigue. Two of the studies identified report some psychometric properties of the PedsQL MFS in paediatrics with IBD.(45;46) Comparing paediatrics with IBD (n=70) against healthy controls (n=157), the mean scores on the PedsQL Total score and three individual domains were lower in the IBD subgroup, and ESs ranged from 0.35 for cognitive function to 0.84 for general fatigue in the self-report data, and from 0.72 for cognitive fatigue to 1.96 for general fatigue in the proxy-report scores.(46) A direct relationship was reported between the total PedsQL 4.0 scores and the MFS (r=0.80, p<0.001).(46) There was also an inverse relationship between age and total MFS (r=-0.16, p=0.02), and a direct relationship was also reported between disease activity (measured using the PCDAI) and fatigue (range: r-0.40 to -0.48, p<0.01 for all).(46) Although described here as it is used as a comparator in two of the studies in this review, the PedsQL MFS is not considered a candidate measure for inclusion in the NCA as it does not capture all aspects of HRQoL associated with IBD.

The PedsQL GI module used in Varni 2014

There are several disease-specific modules of the PedsQL (for asthma, arthritis, cancer, cardiac disease, diabetes) which are designed to be used in conjunction with the core modules. A recent publication provides initial results from a module designed for use in gastrointestinal (GI) conditions including IBD (CD and UC). In development since 2008, this module has both parent-proxy (ages 2-4, 5-7, 8-12, 13-18 years) and self-report versions (ages 5-7, 8-12, 13-18 years). The module includes 74 items describing 24 different scales relating to GI-specific symptoms: stomach pain and hurt (6 items), stomach discomfort when eating (5 items), food and drink limits (6 items), trouble swallowing (3 items), heart burn and reflux (4 items), nausea and vomiting (4 items), gas and bloating (7 items), constipation (14 items), blood in poop (2 items), diarrhoea (7 items), worry about going poop (5 items), worry about stomach aches (2 items), medicines (4 items), and communication (5 items).(49) As with the core modules, the scales are transformed onto 0-100 scale with higher scores indicating fewer problems and less severe symptoms.(49) This measure is considered to be a candidate measure for the NCA when used in conjunction with the core PedsQL measure.

A relatively large sample (n=584 children aged 5–18 years, n=682 parents of children aged 2–18 years) of patients with either functional gastrointestinal disorders (FGIDs) or organic gastrointestinal diseases (including CD and UC) were used to assess the measurement properties of the PedsQL GI module.

Acceptability: The module was well accepted with just 1.69% and 1.84% of item responses missing on the child self-report and parent proxy-reports respectively.

Reliability: The ICC statistics between the self-report and parent proxy-report showed agreement for the vast majority of PedsQL GI scales, with exceptions being communication (ICC=0.37) and trouble swallowing (ICC=0.49). Test-retest reliability was not tested.

Construct validity (known group): The tests for known group validity (subgrouped by seven GI disorders with CD=192 and UC=65) showed the expected differences in mean scores ($p<0.001$) for both the child self report and the parent proxy-report, however, the sample sizes for the subgroups were extremely small in some cases.

Responsiveness: Although there was no evidence of floor effects, ceiling effects (less GI symptoms) were observed on a number of the individual scales (e.g. >60% scored the highest value for 'blood in poop' and 'trouble swallowing'), which may infer insensitivity to improvement in paediatrics with less severe symptoms. Responsiveness to changes over time was not tested.

Summary and conclusion of review of literature on paediatrics with IBD

While primary research in this area appears to be growing with evidence of development of several PROMs targeted at paediatrics with IBD, the evidence identified which could be used to compare PROMs in this population was limited. The searches, although limited in scope due to the time constraints of the project, did not identify any evidence which could be used to generate QALYs directly from PROMs in this population. The most likely target measures for inclusion in the IBD NCA are the IMPACT-III and the PedsQL™ v4. Based on the evidence reviewed, the target age group, and alternative responder versions, the PedsQL™ is recommended over IMPACT-III measure. However, research is required to generate an associated preference-based tariff for the PedsQL™ (or a mapping function to one of the alternative preference-based generic measures) which could be used to generate utility values for use in cost-effectiveness evaluations.

Table 1: Summary of evidence on PROMs for IBD

Measure (N)	Target Age (years)	Target Responder	Acceptability	Reliability	Construct (KGV; Convergent)	Responsiveness (Change over time; Ceiling effects)
Adults						
EQ-5D (2)	-	-	Good	Good	Good; Good	Mixed; Mixed
Acceptable but requires additional validation (n studies =2) particularly in patients with severe IBD and those undergoing surgical procedures. Is not appropriate for paediatrics with IBD.						
PRO2, PRO3 (1)	-	-	Very limited evidence available (n studies =1)			
Acceptable but requires additional validation and is only suitable for adults with CD (not for UC).						
Paediatrics						
IMPACT-III (3)	≥9	SR	Good	Good	Mixed; Good	No evidence
Acceptable but requires additional validation (n studies = 3) and cannot be used to generate QALYs						
PedsQL (5)	2-18	SR;PR	Good	Good	Good	No evidence
Acceptable (n studies = 5) but cannot be used to generate QALYs						
PedsQL GI module (1)			Good	Good	Good; No evidence	No evidence; poor
Acceptable, but very limited evidence (n studies = 1), would need to be used as an adjunct to the PedsQL core measure, and cannot be used to generate QALYs						

KGV = known group validity; CE=ceiling effect; N = number, SR=self-report, PR=Parent/carer-report

4.3 Evidence for economic evaluations in IBD (WP1.3)

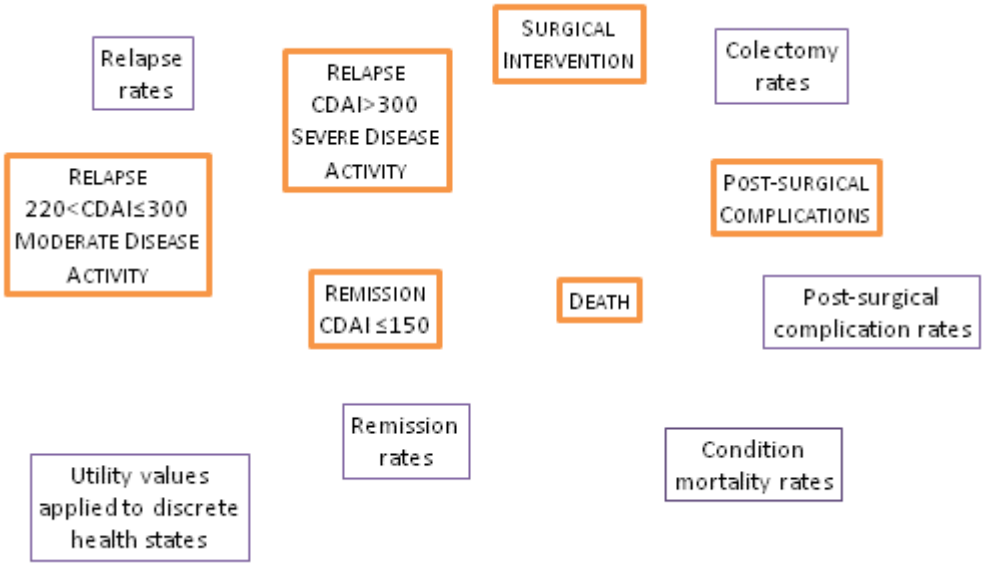
4.3.1 Cost-effectiveness modelling approach used in recent HTAs in IBD

Three technology appraisals (TAs) relating to IBD were identified from the searches. One was superseded by a more recent publication,(61) leaving one multiple technology assessment (MTA),(62) and one single technology assessment (STA) (Table 2).(63) The MTA compared several pharmaceutical interventions for the treatment of moderate to severe CD or fistulising CD in both adult and paediatric populations.(62) The STA compared surgery (colectomy) with rescue therapy (standard care or alternative pharmaceutical interventions) for avoidance or delay of surgery and symptom free remission in hospitalised patients with acute exacerbation of UC.(61) Both of these interventions are reflective of those provided to patients in the IBD NCA which gathers information from secondary care settings.

The MTA used a Markov model with discrete health states defined by remission and surgery (Table 2, Figure 1), with both clinical severity and remission informed by the CDAI (e.g. CDAI<150: quiescent disease or remission; 220<CDAI<300: moderately active disease; CDAI>300: severe disease).(64) The STA used an initial decision tree for the clinical trial data and extrapolated beyond the trial horizon using a Markov model.(61) The clinical pathway was represented by discrete health states (e.g. achieved remission, failed treatment, colectomy, post-surgery complications etc). Response was defined as: a clinic-activity score <10 on two consecutive days and a drop of at least three points (N.B. no additional details or reference was provided for the clinical-activity score). Health state transitions for remission, relapse and post-surgical complications for both TAs were derived from RCT data. Although adverse events for anti-TNF agents are a potential problem, these were not modelled explicitly, but were incorporated into withdrawal from the therapies. Prevalence and changes in concomitant medications such as corticosteroids were sourced from clinical studies.

Figure 1 provides a synopsis of the health states (orange framed boxes with uppercase text) and evidence (purple framed boxes with lower case text) used in the TA for CD.

Figure 1: Modelling approach used in CD HTA



Both studies quality adjusted survival by assigning mean utility values to the discrete health states. For CD, published utility values elicited using time trade-off (TTO) methods were initially used.(65) The utilities used in the paediatric assessment, (62) and the surgical health states in both TAs were based on adult values and assumptions respectively due to the absence of more suitable data.

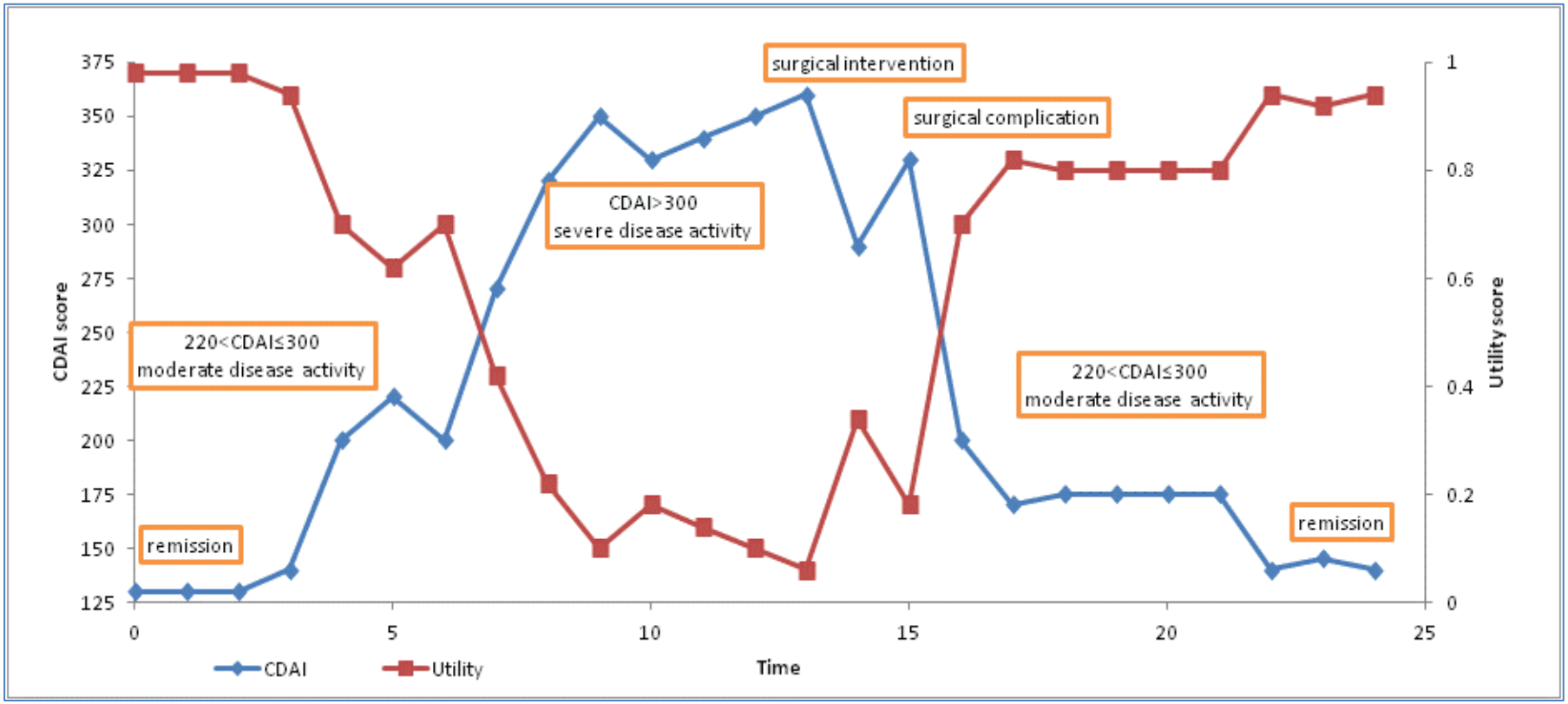
Table 2: Summary of existing models used in IBD TAs

Model method, clinical effect	Method used to model utilities
MTA (TA187): Crohn's disease - infliximab (review) and adalimumab (review of TA40); 2010	
TAG Markov model Four discrete health states: remission‡, relapse (severe: CDAI>300; moderate: 220<CDAI<300), surgery, post-surgery remission‡, death Effectiveness: intervention specific rates for remission/relapse Source: RCTs used for clinical effect	Utility: non-preference values obtained using TTO (65); mean values assigned to discrete HS Source: published literature (adults), assumptions AEs: not specifically modelled
STA (TA163): Ulcerative colitis (acute exacerbations) - infliximab; 2008	
Decision tree followed by Markov model Four discrete health states: remission, active ulcerative colitis, surgical (colectomy) remission, surgical complications Effectiveness: individual intervention rates for colectomy, remission, surgical complications Source: RCTs used for clinical effect	Utility: EQ-5D; mean values assigned to discrete HS Source: survey of patients with ulcerative colitis AEs: assumed the disutility associated with post-surgery complications were equivalent to the utility of active ulcerative colitis

HS: health states; AE: Adverse Events; MTA: Multiple Technology Appraisal; STA: Single Technology Appraisal; TAG: Technology Appraisal Group; TA: Technology Appraisal; TTO: time trade-off; RCT: randomised controlled trial. ‡remission: defined as a CDAI score <150

A subsequent critique of the economic model indicated that a published statistical model describing a continuous relationship between the CDAI and EQ-5D may be more appropriate than the discrete values applied to the individual health states.(62;66) Figure 2 illustrates the potential continuous relationship between the proxy measure (CDAI) and the utility values (EQ-5D) required to generate QALYs. The orange boxes represent the discrete health states used in the existing economic model while the blue (diamond marker) line and red (square marker) line show the changes in disease severity (measured using the CDAI) and utilities (measured using EQ-5D) over time respectively. Modelling a continuous relationship between utilities and a clinical measure of function or symptoms is now widely accepted as the most appropriate approach in chronic conditions characterised by periods of flares and remission.(67;68)

Figure 2: Alternative approach describing utilities by proxy measure (CDAI in CD)



Legend: the orange boxes represent the discrete health states used in the CD cost-effectiveness model, CDAI: Crohn's disease activity index

In summary, the following evidence would be required to compare providers or the cost-effectiveness of interventions for IBD:

- Condition severity (repeatedly measured over time using CDAI for CD, or CAI for UC)
- Surgical rates (type of intervention, success rate, post-surgical complication, length of stay)
- Pharmaceutical interventions (type of intervention, concomitant medications, remission rates, relapse rates, adverse events)
- Utility values (collected alongside condition severity scores and surgical interventions)
- Death rates (IBD related, all cause)

The majority of this evidence would need to be linked through timings of collection.

4.3.2 Fields collected in the IBD NCA

The objective of the UK IBD NCA is to improve the quality and safety of care delivered in secondary care for patients (any age including paediatrics) with IBD throughout the UK. The biennial audit collects information on processes and outcomes relating to both inpatient and outpatient services for IBD from each hospital participating in the audit. Different levels of information are collected on each of three sub-categories: CD, UC, IBD unspecified, depending on the particular round of the audit. For example, in the fourth round, the Inpatient, and Organisational audits were only completed for patients (up to a maximum of 50[1] patients per hospital per year) with UC. To be eligible for inclusion, the patients (of any age) had to have been admitted for treatment or surgery for UC, and had to have remained in hospital for longer than 24 hours. Patients were not eligible for inclusion in the audit if the primary reason for admission was not treatment of UC, or they had a day case procedure (e.g. infusion, endoscopy or day surgery), or they were discharged within 24 hours of admission. The clinical information is collected via three mandatory questionnaires (depending on audit round) completed by NHS staff: Inpatient audit, Organisational audit, and Biologic audit[2]. There is also a postal patient questionnaire of inpatient experience which includes individual patient unique identifiers and is completed following discharge by either the patient or their parent/carer (for paediatrics).

[1] Multiple admissions for same person are treated as independent for the audit. Hence a total maximum of 50 entries may have less than 50 individual patients (multiple admissions linked via the system generated identifier)

[2] As recommended by NICE, audit of the use of biologics (adalimumab, infliximab) is mandatory in all patients treated with these.

The Inpatient Care Audit is completed for individual patients and covers areas such as patient demographs, admission and mortality, the extent (or severity) of the condition/symptoms and any comorbidities, medical and surgical interventions, discharge arrangements, and any outpatient care prior to admission (Appendix). The Organisational audit is completed once for the Trust/hospital and provides total numbers and organisation information in the following areas: patient and IBD staff demographs, patient experience and involvement in the IBD service, clinical quality (direct and extended IBD team, available access to specialists and diagnostic services, MDT processes, surgery, inpatient facilities and care etc), provision and support for patient's choice of care, research involvement, and education provision and support (Appendix). The Biologic audit is completed for each patient on anti-TNF therapy and contains information on: patient demographs, IBD disease details, initial anti-TNF therapy, current anti-TNF therapy including dose and continued use etc, treatment selection, reviews of treatment, any adverse reactions to the therapies, and a disease severity score (severe, moderate, mild). The self-completed patient questionnaire, **which includes the EQ-5D questionnaire,** gathers information on the patient's experience of the health care services provided by the hospital and background information such as the reason for hospital admittance (elective, scheduled), the type of ward(s) they stayed on, toiletry facilities, hygiene, dietary requirements and standard of food received, the clinicians and nurses involved in providing care, their personal care and treatments, levels of pain, operations and procedures, and information on medications provided during hospitalisation and on discharge (Appendix).

4.3.3 Comparing fields in IBD NCA with variables used in existing HTAs

Based on the existing HTA models, the key clinical information required to inform a standard economic evaluation comparing interventions (either anti-TNF agents, or surgical procedures) in IBD in the secondary care setting would be: condition severity, therapy regimens, the associated remission, relapse and withdrawal rates, the rates and types of surgical interventions and complications, and preference-based utility values. With the exception of the utility data, in the existing TAs the evidence required was sourced from clinical trial data. While the current IBD audit collects some information on the majority of these areas for the individual patients and for the hospitals taking part, there are some obvious omissions and one of the key issues is the timing of the data collection.

Looking at the evidence that could be used to compare providers, the Organisational audit would provide aggregate numbers on surgeries performed per hospital and the Inpatient care audit would provide some data on the indication for surgery, the surgical procedure and surgical complication

rates. However, it not clear if there is sufficient information to adjust for case-mix (for example severity of condition, described using a standardised clinical measure) which could affect surgical success rates. The mandatory biologic audit would provide some of the information require to compare these pharmaceutical interventions (for example withdrawal rates and the reason for discontinuation together with the dates of these), but it is not clear if all data needed is collected (for example remission or relapse rates and associated dates), and again some validated measure of severity of condition, such as the CDAI, would be required to use these data in an economic model.

As mentioned earlier, the EQ-5D is collected in the patient questionnaire. However, this instrument asks patients to describe their health related quality of life 'today'. As there is no other PROM or indicator of disease severity in the questionnaire, it will not be possible to link these responses to health states defined within an economic model such as remission, relapse or treatment related adverse events. In addition, as the evidence is collected post discharge, the patients will not provide responses relating to surgery or surgical complications. It may be possible to use a published statistical relationship to predict utility values from the reported CDAI scores for patients with CD.(62) While the CDAI is currently collected in the IBD audit, the timing of collection of the EQ-5D and CDAI differ, consequently the evidence will be of limited value for informing the utility values required for an economic evaluation. For these data to be useful for economic modelling purposes, they need to be collected at the same time, particularly as IBD is characterised by periods of 'flares' and remission.

While it is believed that the EQ-5D will be retained in the IBD patient questionnaire, it is not known if there are any planned or ongoing studies directly related to the inclusion of additional PROMs in the IBD audit.[personal communication Kajal Mortier, project coordinator, May 2014] The recently reported initial results for the CD PROM (PRO2 and PRO3) derived from items within the CDAI are promising and this may provide an alternative worthy of consideration (see section 4.3).(25;69) However, this would still leave gaps in the evidence base required for surgical procedures and pharmaceutical related adverse events. To our knowledge, there is also no equivalent PROM or published relationship between a clinical measure and preference-based utility measure for patients with UC, or for paediatric patients at the moment.

4.4 Recommendations for IBD

Table 3 summarises the recommendations and associated future research for IBD. In summary, the EQ-5D appears to be appropriate in adults with IBD, and the current IBD audit collects much of the information required to conduct economic evaluations. However there are caveats associated with these conclusions which require consideration. The PedsQL appears to be the most appropriate measure for paediatrics, but there are limitations with the usefulness of this measure for economic evaluations. The issues and corresponding potential recommendations (PR) and areas for future research (FR) are discussed below. All suggested future research areas are indicative and would require a discussion and detailed proposal if required.

The conclusion that the EQ-5D is appropriate in adults was informed by just two studies involving patients with either CD or UC, and only one of the studies assessed the EQ-5D UK preference-based index. The evidence used in the TAs indicated that there was a dearth of robust EQ-5D evidence in this population, particularly in adults at the more severe end of the disease spectrum, in paediatrics, and in patients undergoing surgical interventions related to their condition. EQ-5D data collected in the current IBD patient questionnaire could greatly enhance the evidence base in this area if it could be linked in some way to clinical severity (see PR.4 below). This would reduce uncertainty in future economic evaluations used to inform policy decisions in the UK and in particular, would enable comparisons of biologics as used in routine clinical practice.

The evidence suggested there may be a ceiling effect in the EQ-5D in adults with less severe disease. However, these patients are unlikely to be among those hospitalized for treatment of their IBD condition. The inclusion of the new five level tool (EQ-5D-5L) could potentially reduce the observed ceiling effects,(70) once the preference-based weights have been confirmed. However, the psychometric properties of this questionnaire have not been assessed in patients with IBD and this would require additional research (PR.1, FR.1). This would involve the concurrent collection of a measure against which the EQ-5D could be assessed, together with additional information such as patient demographs, recent surgical procedures and outcomes, current medications etc.

The IBD NCA includes patients of all ages (including paediatrics) while the EQ-5D is specifically targeted at adults (over age 18 years). The results of the literature review for paediatrics suggests the PedsQL^{TM} is the most appropriate measure for inclusion in the NCA, augmented with the PedsQL^{TM} GI module once validated (PR.2). However, there is no existing method to generate utilities from this instrument, so its usefulness for economic evaluations is limited. Research to

generate an associated preference-based measure would require collaboration with the developers of the PedsQL™. This could be directly through preference-weights for the PedsQL™, or indirectly using a mapping function from the PedsQL™ to one of the alternative paediatric preference-based measures. This is worth considering given that the PedsQL™ has different versions for different age groups and also has both patient report and proxy-report versions. Alternatively, the inclusion of a validated preference-based measure specifically designed for paediatrics (such as the CHU-9D or HUI2 for younger children and the EQ-5D-Y for adolescents)(1-3) is an option which might be considered in the interim to ensure that the NCA data can be used to inform economic evaluations (PR.3). Again, the psychometric properties of the measures included would need to be assessed in this population (FR.3).

As mentioned previously, a review of the latest TA conducted by the National Institute for Health and Care excellence (NICE)'s Decision Support Unit, suggested that the use of a mechanism to map from condition severity to preference-based utility measures could potentially improve the methodology (as described in Figure 2) and reduce the uncertainty in the results generated from CD cost-effectiveness models.(62) There is at least one published function which could be used to map between the CDAI and the EQ-5D-3L, suitable for adults with CD, but no known functions for UC or paediatrics.(66) This methodology would require concurrent collection of the EQ-5D and the clinical measure (e.g. the CDAI and the PCDAI for CD, and equivalent measures for UC) (PR.4). An alternative would be to identify a suitable PROM for inclusion in the patient questionnaire. As discussed earlier, the PRO2 or PRO3, which could potentially be used within a cost-effectiveness model, might be suitable for CD, but no known equivalents are available for either UC or paediatrics. For this evidence to be used in economic modelling, research would be required to generate mapping functions between the clinical and preference-based measures in adults and in paediatrics separately for patients with CD and patients with UC (FR.5).

The IBD audit collects a wealth of information on the clinical status of patients admitted to hospital for treatment of their condition, and the associated interventions and care received whilst in hospital and on discharge. However, it is not clear if there is sufficient detailed mandatory information on variables such as response to treatments, relapse and clinical activity, to inform all parameters required for an economic model. Additional mandatory fields to capture this information would considerably increase the flexibility of secondary use of the data (PR.6). Formal detailed recommendations of which fields to include would require additional detailed inspection of the exact data collected in the current IBD audit (FR.6).

Finally it is recommended that the proposed links between the IBD audit and the new IBD register (see Section 4.3.3) are utilised to make full use of the clinical and PROM data that will be available (PR.7).

Table 3: Recommendations and associated future research for IBD

PR.1	*Include the new version of the EQ-5D (EQ-5D-5L) in future adult patient questionnaires*
FR.1	*Assess the psychometric properties of the EQ-5D-5L in adults with IBD using data collected in the audit*
PR.2	*Include the PedsQL™ (and the PedsQL™ GI module) in future paediatric patient questionnaires*
FR.2	*Investigate potential collaboration with the developers of the PedsQL™ with a view to developing a methodology to generate preference-based utility measures directly or indirectly (via mapping to alternative measure) from the PedsQL™*
PR.3	*Include age related paediatric preference-based HRQoL instrument (e.g. CHU-9D, HUI2 and EQ-5D-Y) in future paediatric patient questionnaires*
FR.3	*Assess the psychometric properties of the paediatric preference-based tools in IBD using data collected in the audit*
PR.4	*Synchronise the timing of collection of a clinical measure (such as the CDAI for patients with CD, or the CAI for patients with UC) and the HRQoL measure*
FR.4	*Conduct analyses to generate mapping functions between the suggested clinical and preference-based measures to enable the evidence to be used in economic models*
PR.5	*Include an additional PROM to capture disease severity, such as the PRO2 or PRO3 (and equivalent measures for UC and paediatrics), in the patient questionnaire*
FR.5	*Assess the validity of the PRO2/PRO3 using data collected in the audit*
FR.6	*Conduct research to generate equivalent condition severity PROMs in adults and paediatrics with UC*
PR.6	*Include additional mandatory fields in the IBD audit such as response to current treatment, relapse and current disease activity (linked by time to HRQoL)*
FR.7	*Detailed analyses of fields currently collected in the IBD audit to identify recommendations for future mandatory fields*
PR.7	*Utilise links between the IBD audit and the new IBD register*

5. SUMMARY

5.1 Summary of evidence used to inform the conclusions for WP1.1 and WP1.2

In summary, a review of primary studies (n=2) provides evidence of acceptability, reliability, and known group/convergent validity for the EQ-5D in adults with IBD (Table 4). However the evidence on the responsiveness of the EQ-5D is mixed with some ceiling effects and potential insensitivity to changes over time reported. While the EQ-5D is considered to be acceptable, additional validation is required particularly in patients with severe IBD and those undergoing surgical procedures. A review of evidence of PROMs for paediatrics provides evidence of acceptability, reliability and known group/convergent validity for the PedsQL™ (5 studies) in paediatrics with IBD (Table 4). The PedsQL™ does not currently have an associated preference-base tariff, but it has both self-report and parent/carer versions and covers the full age spectrum for paediatrics (2-18 years). Additional preference-based measures are also recommended for use in paediatrics with IBD.

Table 4: Summary of evidence currently available for recommended measure(s)

Measure	N	Acceptability	Reliability	Construct		Responsiveness		Overall
				KGV	Convergent	Change over time	Ceiling Effect	
EQ-5D	2	Good	Good	Good	Good	Mixed	Mixed	Acceptable but not appropriate for paediatrics
PedsQL	5	Good	Good	Good	NE	NE	NE	Acceptable
PedsQL GI module		This measure is currently being validated and will be available shortly						

N= number of studies used to inform conclusions, KGV: known group validity; NE: no evidence
[a] consider the PedsQL GI module as an adjunct to the core measure

5.2 Summary of evidence required for use in economic evaluations (WP1.3)

The EQ-5D is currently collected in the IBD audit, but as it is not collected at the same time as other key variables used in the economics (for example, surgery or flares in symptoms), its usefulness in comparing interventions is limited. It may be possible to use a clinical variable (for example the CDAI in patients with CD) and an existing relationship between the CDAI and EQ-5D to enable the NCA data to be used in economic evaluations. Despite the issue with the timing of collections, the EQ-5D would be useful when comparing providers and if the timings of data collection could be synchronised with the clinical data, then it could be used in standard economic evaluations. While the audit collects much of the information required to conduct economic evaluations, for example the aggregate numbers of surgeries and surgical complications could be used to compare providers, it is not clear if there is sufficient evidence to adjust for case-mix. There are also areas where

additional evidence, if mandatory, would be beneficial for future economic evaluations. These include details of pharmaceutical interventions and associated response and relapse data collected at the same time as a clinical variable such as the CDAI, surgical rates including type of intervention, success rate and associated complications.

APPENDIX: INFLAMMATORY BOWEL DISEASE

The tables in this Appendix provide additional information for the reviews (WP1.1, 1.2 and 1.3) conducted for IBD.

Table A1: Characteristics of studies included in the systematic review of primary studies for IBD

Study ref Author, Year	Country	Disease/treatment stage	Treatment (if any)	Study type (e.g. cross sectional, RCT, cohort)	Study objective
Konig, 2002(6)	Germany	Patients with inflammatory bowel disease (either Crohn's disease or ulcerative colitis)	No treatment, single cohort Questionnaires given 2 times, two weeks apart	Consecutive patients attending outpatient appointments	To analyse the construct validity, criterion validity, test-retest reliability and responsiveness of the EQ-5D
Stark, 2010(7)	Germany	Patients with inflammatory bowel disease (either Crohn's disease or ulcerative colitis), 18 years or older	Treatment not reported	Random sample of members of German IBD association	To assess validity, reliability, and responsiveness of EQ-5D, especially the responsiveness to meaningful differences in patient reported changes in health status

RCT, randomised controlled trial; IBD, Inflammatory bowel disease; CD, Crohn's disease; UC, ulcerative colitis

Table A2: Participant characteristics studies in the systematic review of primary studies for IBD

Study ref Author, Year	Number of participants recruited	Age in years mean (sd); range	male %	Ethnicity	Other characteristics	Missing data (patients completing study) include reasons for non-completion if given
Konig, 2002(6)	152	41.4 (12.6); 17-73	52%	NR	Crohn's disease: 80.9% Ulcerative colitis: 19.1% Age at diagnosis: 27.1 (sd 11.3) yr Duration of disease: 13.9 (sd 8.5) yr Remission: 62% Active disease: 38%	Some data are missing for 6 patients with ileostomy
Stark, 2010(7)	502	42 (11); 17-83	41%	NR	Crohn's disease: 53.78% Ulcerative colitis: 46.21% Age at diagnosis: 29 (sd 11) yr Duration of disease: 14 (sd 8) yr Remission: 59.6% Active disease: 40.4%	447 patients returned follow-up questionnaires Some (n<5) data missing from some analyses Reasons NR

SD, standard deviation; yr, year; NR, not reported; CD, Crohn's disease; UC, ulcerative colitis

Table A3: Valuation and descriptive methods used in the systematic review of primary studies for IBD

Study ref Author, Year	GENERIC MEASURES Descriptive system	Tariff used	Mean (SD); 95% CI	OTHER MEASURES USED Condition-specific HRQL measures	Clinical measures	Qualitative questions	Missing data; completion rates of measures; etc.
Konig, 2002(6)	EQ-5D SF-36 health dimensions	German,(8) (range 0-100)		None	IBDQ CDAI CAI	none	Acceptance of EQ-5D assessed by proportion of missing responses on EQ-5D
Stark, 2010(7)	EQ-5D	UK (10) and German (9)	At baseline **UK tariff** **CD:** 0.77 (SD 0.24; median 0.8) **UC:** 0. 84 (SD 0. 18; median 0. 85)	None	CDAI CAI	None	NR

IBDQ, Inflammatory Bowel Disease Questionnaire; CDAI, Crohn's Disease Activity Index; CAI, Clinical Activity Index; SD, standard deviation; CD, Crohn's disease; UC, ulcerative colitis

Table A4: Acceptability, reliability and validity assessment in IBD

Author, Year	Method of measuring validity Type of validity, how (e.g. known group/convergent)?	Validity results Group A(n) vs. Group B(n)[ø] Mean EQ-5D; mean difference in EQ-5D	Authors' conclusions/notes
Konig, 2002(6)	Acceptability, assessed by proportion of missing responses	EQ-5D missing range: 0.7 to 3.3% IBDQ missing range: 2.0% to 9.2%	As shown by the low proportions of missing responses, the EQ-5D was well accepted in this population
	Reliability (test-retest), kappa statistic	80.5% of outpatients completed the 2nd EQ-5D 52 (79%) reported no change, 12 (18%) reported an improvement **EQ-5D health dimension (n): agreement (%), Kappa statistic** Mobility (52): 47 (90.4%), 0.39 Self-care (52): 52 (100.0%), 1.00 Usual activities (52): 46 (88.5%), 0.73 Pain/discomfort (50): 43 (86.0%), 0.74 Anxiety/depression (51): 41 (80.4%), 0.61	The EQ-5D index showed ceiling effects hence it may not be able to discriminate health states in patients with less severe disease The construct/concurrent validity results were good on the whole The test-retest results were good
	Construct validity (convergent), Spearman rank correlation	EQ-5D and CDAI correlation = -0.48* EQ-5D and CAI correlation = -0.66**	
	Construct validity (convergent), Spearman rank correlation	Correlations between EQ-5D and IBDQ ranged from 0.52 to 0.62 (p<0.0001 for all)	
	Construct validity (known group), remission vs active disease, various statistical tests used	**Remission** (% no; moderate; extreme problems) **Mobility:** 96.1; 3.9; 0 **Self-care:** 98.7; 1.3; 0 **Usual activities:** 90.7; 9.3; 0 **Pain/discomfort:** 58.1; 41.9; 0 **Anxiety/depression:** 70.3; 27.0; 2.7 **Active** (% no; moderate; extreme problems) **Mobility:** 66; 29.8; 4.3, p<0.0001[a] **Self-care:** 83.0; 12.8; 4.3, p=0.0019[b] **Usual activities:** 42.6; 42.6; 14.9, P<0.0001[a] **Pain/discomfort:** 23.9; 67.9; 8.7, NR **Anxiety/depression:** 44.7; 51.1; 4.3, p=0.005[a]	
	Construct validity (known group), Inpatients vs.	**Outpatients** (% no; moderate; extreme problems)	

	outpatients, various statistical tests used	**Mobility:** 90.0; 10; 0 **Self-care:** 96.7; 3.3; 0 **Usual activities:** 74.0; 24.4; 1.7 **Pain/discomfort:** 44.0; 54.3; 1.7 **Anxiety/depression:** 59.3; 36.4; 4.2 **Inpatients** (% no; moderate; extreme problems) **Mobility:** 64.5; 29.0; 6.5, p=0.0013[b] **Self-care:** 77.4; 16.1; 6.5, 0.0015[b] **Usual activities:** 41.9; 38.7; 19.4, 0.0007[a] **Pain/discomfort:** 38.7; 48.4; 12.9, 0.5994[a] **Anxiety/depression:** 51.6; 41.9; 6.5, 0.4394[a] Significantly better response levels for outpatients (compared to inpatients) observed for: mobility, self-care, usual activities. For pain/discomfort and anxiety/depression difference was small and not significant.	
Stark, 2010(7)	Reliability (test-retest), patients who reported no change in health. Simple kappa to test categorical variables; ICC using 2-way ANOVA to test continuous variables	**Kappa statistic** Substantial to almost perfect agreement for all EQ-5D health dimensions **UK index ICC** All: 0.76 CD: 0.76 UC: 0.73 **Same health:** the mean EQ-5D score in those reporting the same health was higher in all groups and significantly higher in the overall group, those with CD, and those with active disease	The reliability of the EQ-5D index scores in test–retest was good, but the ICCs and thus the reliability of the EQ-5D index scores were lower than that of the VAS score.
	Construct validity (convergent), Spearman rank correlations	**UK index** EQ-5D and CDAI in CD patients: r= -0.75 (p<0.0001) EQ-5D and CAI in UC patients: r= -0.65 (p<0.0001)	The construct validity of the EQ-5D in IBD subjects, i.e., its agreement with accepted disease criteria, was good, and it was able to discriminate between those with active disease and those in remission. UK index scores of subjects in remission were significantly better than those with active disease
	Construct validity (known group), active disease vs remission. Statistical test type not reported.	**EQ-5D UK index** Mean (SD) CD remission: 0.89 (0.13), range (0.26-1.00) CD active: 0.61 (0.29), range (-0.18–1.00), p <0.0001 UC remission: 0.91 (0.14), range (0.23–1.00)	

		UC active: 0.71 (0.18), range (0.09–1.00), p <0.0001 All EQ-5D health dimensions showed significant differences between remission and active disease except self-care in UC

n, number; IBDQ, Inflammatory Bowel Disease Questionnaire; CDAI, Crohn's Disease Activity Index; CAI, Clinical Activity Index; CD, Crohn's disease; UC, ulcerative colitis; ICC, intraclass correlation coefficient; ANOVA, analysis of variance; [x]p<0.0001, Spearman rank; [**]p<0.001, Spearman rank; [a]Chi squared test with categories 'moderate problems' and 'extreme problems' collapsed into one category due to small expected number of observations (<5) in 'extreme problems' category.
[b]Fisher's exact test with categories 'moderate problems' and 'extreme problems' collapsed into one category due to small expected number of observations (<5) in 'moderate problems' and 'extreme problems' categories.

Table A5: Responsiveness assessment in IBD

Author, Year ref	Method of measuring responsiveness (e.g. effect sizes, statistical significance)	Responsiveness results	Authors' conclusions/notes
Konig, 2002(6)	Responsiveness (ceiling effect), % scoring full health	Ceiling effect observed on EQ-5D (59.7% rated at least 4/5 items on EQ-5D as 'no problems' – 25% scored EQ-5D =1	
Stark, 2010(7)	Responsiveness (ceiling effect), % reporting full health	31% of CD and 43% of UC patients reported full health	Indicative of a ceiling effect
Stark, 2010(7)	Patients who reported a change in health in the transition question. T-test of difference in means.	**Mean Difference (Mean (SD (n)) for worse; same; improved** **UK Index)** **All:** -0.09 (0.168* (26)); 0.019 (0.14* (357)); 0.095 (0.142*** (41)) **CD:** -0.102 (0.206 (16)); 0.026 (0.155* (194)); 0.070 (0.134* (21)) **UC:** -0.07 (0.083*(10)); 0.011 (0.123 (163)); 0.123 (0.148**(20)) **Active:** -0.098 (0.209 (15)); 0.047 (0.20** (121)); 0.148 (0.135***(20)) **Remission:** -0.084 (0.102* (9)); 0.007 (0.092 (195)); 0.038 (0.134 (19)) **Improved health :** Changes in EQ-5D index scores were statistically significant (p<0.001) for patients reporting improved health except in those in remission. **Worse health:** The mean EQ-5D score in those reporting worse health was lower in comparison to subjects reporting stable health in all subgroups, but it was only significantly different in the overall group, UC subjects, and subjects in remission	EQ-5D failed to respond in some subgroup analyses, though overall changes were seen. This may reflect the low power in these groups (n: 9–21) or it may be that the change in health status is not captured by the questions of the EQ-5D.. The EQ-5D index score was most responsive (large ES and SRM) for subjects with active disease who reported improved health and for subjects in remission who reported worse health
Stark, 2010(7)	Patients who reported a change in health in the transition question. Standardised response mean (SRM).[a]	**Standardised Response Mean (worse; same; improved)** **UK index** **All**: -0.53; 0.13; 0.67 **CD**: -0.49; 0.17; 0.52 **UC**: -0.84; 0.09 0.83 **Active**: -0.47; 0.24; 1.10 **Remission**: -0.82; 0.07; 0.28 **Improved health:** direction of effect as expected in all groups, but only a strong effect (SRM>0.8) in UC and active disease **Worse health:** direction of effect as expected in all groups, but a strong effect (SRM>0.8) only seen in UC and those in remission	

	Patients who reported a change in health in the transition question. Effect size (ES).[b]	**UK index (worse; same; improved health)** All: -0.33; 0.09; 0.55 CD: -0.32; 0.11; 0.43 UC: -0.68; 0.06; 0.66 Active: -0.38; 0.18; **1.06** Remission: -0.63; 0.05;0.22 All had effect sizes in the expected direction, but only data in patients with active disease who reported improved health showed a strong effect size (>0.8)

SD, standard deviation; n, number; CD, Crohn's disease; UC, ulcerative colitis; SRM, Standardised response mean; ES, effect size.

[a] calculated by dividing the difference of the means at the 2 timepoints by the SD of the differences of scores between the 2 timepoints

[b] calculated by dividing the difference of the means at the 2 timepoints by the standard deviation (SD) of the baseline mean which relates the change to the baseline SD

*P < 0.05; **P < 0.01; ***P < 0.0001.

Table A6: Measures reviewed or used in the seven studies included in the paediatric systematic review in IBD

	Abdovic 2013(44)	Duffy 2011(45)	Marcus 2009(46)	Perrin 2008(47)	Upton 2005(48)	Ogden 2011(40)	Lane 2012 (50)	Varni 2014(49)
Study objective and population	Children (≥ 9 years) CD=74 UC=30	Description of HRQoL measures used in children with IBD	Examined fatigue in children (X-x years) with IBD (n=70) compared to controls (n=157)	Evaluated the Impact Questionnaire in children (8-18 years) with UC=59 CD=161	Assessed the UK version of the PedsQL v4 in a mixed sample including IBD Total (IBD) children=1399(76) Total (IBD) parents =970 (67) Age 8-18 years	Validate the IMPACT-III (UK) in British children (8-17 years) with IBD CD=64 UC=12 IC=21	Conference abstract of Varni 2014	Assess psychometric properties of PedsQL GI module in paediatrics (3-18 years) with a broad range of GI disorders CD=192 UC=67
IMPACT-III (51;52;56)	Yes	Yes	Yes	-	-	Yes	-	-
IMPACT-II questionnaire(51;52)	-	Yes		Yes	-		-	-
PCDAI (severity)(54)	-	-	Yes	Yes	-		-	-
CDI (57)	-	Yes	Yes	-	-		-	-
CHQ(58)	-	-	-	-	-	yes	-	-
PedsQL fatigue(46;51;71)	-	Yes	Yes	-	-		-	-
PedsQL 4.0 [varni 1999, (4;72-75)	Yes	Yes	Yes	Yes	Yes		Yes	Yes
PedsQL GSM (49)	-	-	-	-	-		Yes De-novo GI module	Yes De-novo GI module

[#]Croatian adaptation of the IMPACT-III, PCDA Paediatric Crohn's disease activity index, IC indeterminate colitis, CDI Children's Depression Inventory
CHQ Child Health Questionnaire, PedsQL Paediatric Quality of Life Inventory

Table A7: IBD specific and generic measures used in the studies included in the paediatric systematic review in IBD

	IBD specific measures		Generic measures				
	IMPACT-III	PCDAI	Child health questionnaire (CHQ)[waters 2009]	PedsQL 4.0 generic core scales [Varni 2001]	PedsQL Multi-dimensional Fatigue Scale [Griffiths 2009]	PedsQL GI module [Varni 2014]	Children's depression inventory (CDI) [kovacs 2003
Age(years)	≥9 years	paediatrics	4-19	2-18	2-18	2-18	≥9 years
Respondent	Child	Clinician	Child/parent	Child/parent	Child/parent	Child/parent	Child
Items	35		87	23	18	74	35
Domains	6	3	10	4	3	24	6
Total summary score	35-175	0-100	2x summary scores 0-100	2x summary scores 0-100	3x summary scores 0-100	Scores range 0-100	35-175

Table A8: Mandatory fields collected in the IBD NCA (Inpatient care)

*INPATIENT **UC** CARE AUDIT TOOL (all questions mandatory)*
Separate adult and paediatric tools, Questions below taken from adult version
PRE-SECTION PATIENT DEMOGRAPHICS
Patient audit number (automatically allocated), Patient's age at admission, Sex
SECTION ONE: ADMISSION/MORTALITY
Admission: Date of admission to this hospital, What was the primary reason for admission (elective admission for established active UC, emergency admission for established active UC, transferred from another site for surgery or further medical treatment, Elective admission for surgery, new diagnosis of UC, other)
Discharge/Mortality: Was the patient (dates): discharged home, transferred for surgery or further medical management , deceased, Was death UC related
SECTION 2: ASSESSING THE EXTENT OF ULCERATIVE COLITIS
IBD team/ward: When was the patient first seen by a member of the IBD team? Was the patient: seen by an IBD Nurse specialist during admission, transferred to a specialist gastroenterology bed
Patient history: What was the extent of the colitis? (proctitis, left sided, extensive, unknown), Has the patient had previous admissions with UC in the two years prior to this admission? If yes (how many times, Has there been a related admission within the last 30 days? Patient already been included in this audit?
Comorbidity: Did the patient have any significant comorbid diseases (none, diabetes, cardiovascular disease, liver disease, respiratory, active cancer, renal failure, other)
Severity of Disease: How many loose or bloody stools were passed in the first full day following admission, Date a stool sample sent for Standard Stool Culture, Stool culture positive (Y/N)
SECTION 3: MEDICAL INTERVENTIONS
Venous thromboembolism: Was the patient given prophylactic heparin, Did the patient have a thrombotic episode during this admission
Weight assessment and Dietetic support during admission: Nutritional risk assessment undertaken, Dietitian see the patient during admission, Patient's weight measured, Dietary treatment initiated
Steroid therapy: Were corticosteroids prescribed during admission, If yes, which (IV /oral corticosteroids)
Which other therapies were started during the admission: Ciclosporin , Anti-TNF, Clinical trial or significant other medical therapies, Name of trial or therapy, Decision to treat discussed at MDT meeting
SECTION 4: SURGICAL INTERVENTIONS
Surgical therapy: Did patient have surgery (date), Indications for surgery (e.g. failure of medical therapy, high grade dysplasia, abscess, closure of stoma, obstruction etc), Seen by a stoma nurse
Surgical complications: e.g. no complications, deep vein thrombosis, wound infection, small bowel obstruction, respiratory, stoma complications, other etc
SECTION 5: DISCHARGE ARRANGEMENTS
If the patient was discharged on steroids was bone protection prescribed, Was patient on immunosupressives on discharge or was there a clear plan to start, Plan for maintenance Anti-TNF on discharge, Was the plan for follow up documented in the notes, If yes, how was the follow- up specified
SECTION 6: OUTPATIENT CARE PRIOR TO ADMISSION
What was the date of the last clinic review, Was disease active at last OPD appointment, If yes, was patient admitted to hospital at this time, If not admitted, was treatment changed, If yes: for 5 ASA, Steroids, Topical, Immunosuppressant: Started/Increased, Stopped/Decreased, Not changed
Prolonged steroid use: Has the patient been prescribed steroids for > 3 months during past 12 months, If yes, what steroid sparing strategies were tried , What was the outcome of the steroid sparing strategy
Anaemia: What was the patient's Hb on admission, If patient was anaemic how long prior to admission was this known, If iron deficient, what treatment was provided, Did the patient tolerate this treatment

Table A9: Mandatory fields collected in the IBD NCA (Organisational)

ORGANISATIONAL AUDIT TOOL (separate adult and paediatric tools; text (from adult text) indicative of areas covered; all questions mandatory; all refer to one year audit period unless stated otherwise)

SECTION ONE: DEMOGRAPHICS

*Number of IBD patients (split by CD, UC, IBDU, adult/paediatric): T*otal service, New, Readmitted < 30 days of discharge, newly-started on Infliximab (adalimumab), admitted primarily for treatment of IBD died during that admission

Number of ileo-anal pouch surgery performed on site

Staff: How many WTE staff in IBD team (e.g. gastroenterologists, colorectal surgeons, IBD nurse specialists, stoma nurses, detitians, administrators)

SECTION 2: PATIENT EXPERIENCE

IBD information provided: how to access IBD services, follow up, educational material, 'patient education' session, regular education opportunities for all IBD patients and their families, clear guidance on how patients can seek a second opinion, Rapid access to specialist advice such as telephone, email, or face to face review for relapse patients *,* exercise choice between treatments, written information about IBD and a range of treatments, access to a translator for all face to face and telephone contacts, information is available that is appropriate to the age, understanding and communication needs of the patient, A selection of written information is available for patients in languages other than English

Patient involvement: actively involved in management decisions about care, clear structured pathway for patient to discuss treatment with MDT, IBD patient panel, Involvement of patients in service planning and improvement , patients given opportunity to provide feedback on their care, Reporting, followed by action planning and change implemented as a result of the patient feedback of care

Education of patients and support groups: Newly-diagnosed patients offered education with an IBD nurse/dietitian, Regular education opportunities, open forum meeting which meets at least annually, Information and support for patient organisations, local patient support groups

SECTION 3: CLINICAL QUALITY

The IBD team: Levels of staff and access to specialists*:* Clinical lead, consultant gastroenterologist, IBD and stoma nurse, dietitian, consultant colorectal surgeon; Clear pathway for referral to rheumatologist, support from: radiologist, pharmacist, defined access to ophthalmologist; per 250,000 population has (WYE): 0.5 administrative, 2 consultant gastroenterologists, 2 consultant colorectal surgeons, 1.5 IBD and 1.5 stoma nurse specialists 0.5 WTE gastroenterology dietitians

Inpatient monitoring: On admission (>50%, >60%, >75%, >90%) patients have weight and nutritional risk assessment, stool sample sent for standard stool culture, regular stool chart documented

Mental health services: IBD inpatients can receive specialist mental health assessment within the acute service (< 48 hour), information available how to access counselling support, can be referred for specialist Clinical Psychological support. Secure funding and a clear referral pathway is in place for referral to clinical psychology or a counsellor

Sexual and reproductive health: Written information: IBD in pregnancy, effects on fertility, sexuality and body image; pregnancy clinic (or named obstetrician) for all pregnant IBD patients on current medical treatment, agreed clinical care pathway for shared care between the women's health and IBD services

Multidisciplinary working: MDT meeting where complex IBD cases can be discussed, joint or parallel clinics for patients requiring joint medical and surgical care, Decisions from MDT are documented in patient notes and fed back. Meetings attended by: gastroenterology dietitian, pharmacist, administrator

Access to nutritional support and therapy: (>30%, >60%, >75%, all) of IBD patients are reviewed by a dietitian during inpatient stay if required, IBD patients can be referred to a dietitian experienced in the dietary management of IBD, Enteral nutrition as a primary treatment is available to patients with Crohn's disease, Information given to all new IBD patients includes nutritional advice, Nutrition MDT available to IBD inpatients, All new patients have malnutrition screening, The nutrition NDT includes: specialist dietitian and nutrition support nurse, consultant gastroenterologist or consultant colorectal surgeon

Arrangements for use of immunosuppressives, Prior to starting biological therapies screening for tuberculosis, Assessed for risk of infections, Counselled about the risk of malignancy and sepsis, written local protocols for administration of biologicals, White blood count measured >= 3 monthly, Clinicians have access to a pharmacist with specialist knowledge / interest, Local protocols for administration of biological include pre-treatment, actions for infusion reactions and accelerated infusions, There is a clear guidance written on action if white cell counts are low, etc.

Surgery for IBD: Informed consent (risks/benefits)*,* Patients put on Association of Coloproctology of Great

Britain and Ireland (ACPGBI) Ileal Pouch Registry, Formal regular governance to review surgical morbidity and mortality including review audit of postoperative complications, Facilities/trained surgeons for laparoscopic / laparoscopically-assisted surgery, Complex surgical procedures undertaken, Patients considered for pouch surgery referred for expert pathological, Nominated lead for IBD surgery, Pouch failure (and salvage) managed in, or referred to agreed regional specialist unit, Annual review of IBD surgical service with review of activity, mortality and morbidity with regularly reviewed action plan

Inpatient facilities: Identifiable gastroenterology ward, intensive therapy unit, mixed medical/surgical high dependency unit, gastroenterology and colorectal surgical facilities are on the same site, IBD / suspected IBD patients usually triaged to the gastroenterology ward on admission. At least one toilet per 6 (4,3 IBD) patients

Access to diagnostic services available for: gastrointestinal pathologist assessment before surgery, and referral of complex cases to a nationally recognised expert, Ultrasound/CT/contrast studies for inpatients, within 24 hours, Routine plain abdominal x-ray on admission, Urgent access to endoscopy (<72 hours), histological reports available (<5 days), Urgent histology biopsies (<2 days), Abscess drainage can be performed by interventional radiology, Outpatient access to ultrasound/CT/contrast studies and endoscopic (<4 weeks), Small bowel MRI available, Consultant radiologist who primarily reports all gastrointestinal radiology, Recent audit of reporting and waiting times for CT/MR and endoscopy

Inpatient care: >(30%, 50%, 75%) patients seen by IBD specialist (<24 hour admission), >(50%, 65%, 75%, 90%) compliance with risk assessment and prescribing of thromboprophylaxis, >(50%, 65%, 75%, 90%) patients receiving discharge steroids placed on steroid reduction programme/covered with bone protection agents, Named pharmacist available for inpatient drug reviews, >90% medication history reconciled by a pharmacist shortly after their admission, Access to IBD nurse during admission

There are Trust/Health Board guidelines for the management of acute severe colitis: >75% IBD patients placed in gastroenterology /named surgical ward (<24 hours admission)

There is an acute pain management team available on site, Pain scores are routinely included in nursing observations, usual practice to refer inpatient with severe pain to the acute pain management team

All patients due to have, or have a stoma can be seen by a stoma nurse during admission if required

SECTION 4: ORGANISATION AND CHOICE OF CARE

Referral of suspected IBD patients: Newly-referred patients can go to gastroenterology/surgical clinics, agreed referral pathway (between GP's / secondary care) for urgent OPD referrals, All urgent referrals seen < 4 weeks (more rapidly if necessary), Guidance developed for GP's for referral/identification of symptomatic patients in whom IBD is suspected

Supporting patients to exercise choice between care strategies for outpatient management: All patients under secondary care are reviewed annually, Stable patients referred back to primary care are given a clear plan about what to do in the event of flare up, GP routinely given clear instructions on need/criteria for annual review (colorectal cancer surveillance, renal function, bone densitometry), Patients offered choice of annual review (hospital clinic, telephone clinic, review in primary care)

Outpatient care: The following are documented for all patients at clinic review: number of liquid stools per day, abdominal pain, weight loss. Systems in place to ensure all patients currently under hospital review are identified and are offered surveillance colonoscopy, Steroid usage recorded to ensure all patients identified who have ≥ 3 months continuous steroid, The following are documented in outpatient review : number of liquid stools per day, abdominal pain or mass, general well-being, psychological concerns, weight loss, smoking status. Bone densitometry offered routinely to all patients (≥3 months continuous steroid), Annual data is collected /presented: % patients who remain on steroids (≥3 months), % these patients discussed at MDT, % start additional therapy (eg immunosuppressives, anti-TN, surgery)

*Care of patients aged 16 years and younger within adult services: D*efined access to a consultant paediatric gastroenterologist/consultant paediatrican with interest in gastroenterology, working with an adult consultant gastroenterologist with interest in adolescents, Inpatients are looked after in an age-appropriate environment, Patients have access to IBD nurse specialist with suitable paediatric experience, The team providing care for patients ≤ 16 years, work within a paediatric clinical network, Paediatric patients undergo endoscopy in an age-appropriate environment, carried out by someone with training or extensive experience in paediatric endoscopy, Team providing care have access to a surgeon, anaesthetist with appropriate paediatric training, Defined access to dietitian with paediatric experience

There is defined access to a radiologist with suitable paediatric experience

Transitional care: Transitional care service for young people to support their transfer to adult services by 18-19 years, Named coordinator responsible for preparation/oversight of transitional care, IBD service has a joint transition clinic with paediatric services, Direct referral (not via GP) available for specialist

endocrinology review (concerns about growth and/or pubertal status), IBD service has a specific paediatric to adult transition policy, Staff can refer to psychological services, Close working relationship with psychological services in clinics/ward, Each young person with IBD has individual transition plan, Age-appropriate written and verbal advice provided on day to day management of symptoms/treatment, Support education provided on sexual health in young people with IBD

SECTION 5: RESEARCH, EDUCATION AND AUDIT

Register of patients under the care of the IBD service: IBD service has a searchable database or registry of adult IBD patients, Database is updated: clinical data about IBD patients receiving hospital care, patients on biological therapy, patients on all immunosuppressants (including biological therapies), clinical data about all patients with a diagnosis of IBD

Participation in audit: Service participates in: national IBD audit, in the national IBD audit and results are fed back to the service. An action plan is completed, Patient surveys are carried out annually, All IBD inpatient deaths are reviewed by the IBD team, an action plan is formulated, action plan implementation reviewed at least annually, Service participates in the national IBD audit, completes an action plan and ensures monitoring of actions or changes, Mortality/morbidity meetings attended by MDT to discuss deaths and outcomes of surgery, Regular patient survey, action plan produced, required changes completed

Training and education: Education opportunities for all medical/nursing staff, IBD team provides IBD training for primary care on an ad hoc basis, Advanced nursing practitioners within IBD team have regular, multidisciplinary training schedule, Attendance is audited, protected time for training provided, Primary care practitioners wishing to provide IBD services are named members of the IBD team

Research: IBD service is: part of a clinical trials network (UKCRN), has enrolled patients in IBD trial (<two years), All service members encouraged to participate in research (monetary support, flexible working)

*Service development: A*nnual review of IBD service carried out, IBD team in one or more clinical network arrangements with neighbouring IBD services, Annual review is attended by a MDT of relevant professionals, Annual action plan completed and achievement of the actions reviewed

Table A10: Fields collected in the IBD NCA (Biologics)

BIOLOGIC AUDIT QUESTIONNAIRE (all questions mandatory) *Six questionnaires: CD(A), CD(I), UC(A), UC(I), IBDU(A), IBDU(I), plus follow-up questionnaire* *Extracts below taken from Crohn's disease Adalimumab (CD(A)) questionnaire*
PATIENT DEMOGRAPHS
Surname, Forename, Gender, Date Of birth, NHS number (or Community Health Index Number, or Health and social care number), Postcode
IBD DISEASE DETAILS
Diagnosis (Crohn's Disease) Maximal disease distribution at the time of decision to initiate biologic therapy (Terminal ileum, colonic, ileocolonic, none of these), Any part of the gut proximal to the terminal ileum (Y/N), Perianal involvement (Y/N), Date of diagnosis
INITIAL ANTI-TNF TREATMENT
Is the patient a new starter or already established on ant-TNF treatment for IBD (new starter/already established), Informed consent to receive anti-TNF treatment taken (Y/N), Initial anti-TNF treatment type (infliximab), If new starter, date of decision to start, Date of initial loading dose, Clinical indication for this treatment (severe perianal CD, active luminal CD, not known, other), Patient receiving any concomitant therapies for the management of IBD at the time of this treatment (Y/N), If yes select from (list (methotrexate, antibiotics, steroids etc), Has the patient previously failed to respond or are intolerant to immunosuppressive drugs/corticosteroids (Y/N), If yes select from list (anti-TNF, methotrexate, antibiotics, steroids etc)
ADALIMUMAB TREATMENT
Induction dose, Planned maintenance dose, Any acute reactions to injections during induction regime (Y/N), If yes select from (list (fever, nausea, rash etc), Disease severity score,* Disease severity (severe, moderate, mild)
BIOLOGIC AUDIT Generic follow-up questionnaire (all questions mandatory)
PATIENT IDENTIFIER
NHS, CHI or HSCN number
TREATMENT SELECTION
Date of initial loading dose, Was the patient: seen for follow up, lost to follow up, transitioned to adult care, transferred to another service, deceased
ADALIMUMAB REVIEW DETAILS
Date of Adalimumab review, Review of treatment plan (continue/stop Adalimumab treatment), If continue treatment (every week/every other week), If continue treatment dose (80mg/40mg), If stop treatment (treatment effective and discontinued, loss of response, poor response, side effects/adverse events, patient choice, patient became pregnant, other)
INFLIXIMAB INFUSION DETAILS
Date of Infliximab infusion, Current Infliximab dose number, Infliximab dose at this infusion (5 or 10 mg/kg, other), Continued Infliximab treatment plan (continue/stop Infliximab treatment), If stop treatment (treatment effective and discontinued, loss of response, poor response, side effects/adverse events, patient became pregnant, patient choice, other)
ADDITIONAL SECTION FOR BOTH ADALIMUMAB AND INFLIXIMAB
Were any acute infusion/injection reactions recorded (Y/N), If yes select from list (e.g. fever, itching, nausea etc), Is patient currently receiving any other medication for the management of their IBD (list of alternative medications), Adverse events since last review (Y/N), If yes select from list (e.g. death, malignancy, infection, drug-induced lupus etc), Disease severity score (severe, moderate, mild)

Table A11: Fields collected in the IBD Patient questionnaire

ADMISSION TO HOSPITAL
Was your most recent hospital stay planned in advance or an emergency
HOSPITAL AND WARD
While in hospital, did you ever stay in a specialist ward that cared mainly for patients with bowel conditions (a "gastroenterology" ward)? When you were first admitted to a bed on a ward, did you share a sleeping area, for example a room or bay, with patients of the opposite sex? During your stay in hospital, how many wards did you stay in? While staying in hospital, did you ever use the same bathroom or shower area as patients of the opposite sex? When you needed to use a toilet or bathroom, was there a suitable one located close by? For most of your stay, what type of room or ward were you in? Were you given enough privacy while you were on the ward In your opinion, how clean was the hospital room or ward that you were in? How clean were the toilets and bathrooms that you used in hospital? Did you see posters or leaflets on the ward asking patients and visitors to wash their hands or to use hand wash gels? Were hand-wash gels available for patients and visitors to use?
FOOD
How would you rate the hospital food? Was the hospital food appetising? How much food were you given? Were you offered a choice of food? Do you have any special dietary requirements (e.g. vegetarian, diabetic, food allergies)? Was the hospital food suitable for your dietary needs? Did you get enough help from staff to eat your meals? During your stay in hospital, did you have a visit from a dietitian? Were you given extra nutritional supplements to take (e.g. special drinks or foods) at any time during your admission to help maintain or gain weight? Did you receive any special feed via a tube (e.g. placed through the nose) or directly into your veins during your admission?
DOCTORS
Was there one doctor in overall charge of your care?, During your stay in hospital, did the doctor in overall charge of your care (consultant) arrange for you to be seen by another specialist (i.e. a different medical or surgical specialist), When you had important questions to ask a doctor, did you get answers that you could understand? If you had any worries or fears about your condition or treatment, did a doctor discuss them with you? Did you have confidence and trust in the doctors treating you? How would you rate the courtesy of your doctors? In your opinion, did the doctors who treated you know enough about your condition or treatment? As far as you know, did doctors wash or clean their hands between touching patients?
NURSES
When you had important questions to ask a nurse, did you get answers that you could understand? If you had any worries or fears about your condition or treatment, did a nurse discuss them with you? Did you have confidence and trust in the nurses treating you? In your opinion, were there enough nurses on duty to care for **you** in hospital? If you ever needed to talk to a nurse, did you get the opportunity to do so? Apart from the regular nursing staff on the ward did you receive a visit from a specialist nurse while you were in hospital (eg. 'IBD Nurse', 'Clinical Nurse Specialist', 'Nurse Consultant' or 'Stoma Nurse') How would you rate the courtesy of your nurses? In your opinion, did the nurses who treated you know enough about your condition or treatment? As far as you know, did nurses wash or clean their hands between touching patients?
YOUR CARE AND TREATMENTS
Sometimes in a hospital, a member of staff will say one thing and another will say something quite different. Did this happen to you? Were you involved as much as you wanted to be in decisions about your care and treatment? How much information about your condition or treatment was given to **you**? While you were in hospital, were you told your diagnosis (explanation of what was wrong with you)? Was your diagnosis explained to you in a way that you could understand? If your family or someone else close to you wanted to talk to a doctor, did they have enough opportunity to do so? Did you find someone on the hospital staff to talk to about your worries and fears? Were you given enough privacy when discussing your condition or treatment? Were you given enough privacy when being examined or treated?
PAIN
Were you ever in any pain? Do you think the hospital staff did everything they could to control your pain? When you had pain, was it usually severe, moderate or mild? During your stay in hospital, how much of the time were you in pain? Did you ever request pain relief medication? Overall, how much pain relief medication did you get?

OPERATIONS & PROCEDURES

During your stay in hospital, did you have an operation or procedure? If yes: Beforehand, did a member of staff explain the risks and benefits of the operation or procedure in a way you could understand, Beforehand, did a member of staff explain what would be done during the operation or procedure? Beforehand, did a member of staff answer your questions about the operation or procedure in a way you could understand? After the operation or procedure, did a member of staff explain how the operation or procedure had gone in a way you could understand?

LEAVING HOSPITAL

Did you feel you were involved in decisions about your discharge from hospital? Were your family or someone close to you given enough notice about your discharge? Did a member of staff explain the purpose of the medicines you were to take at home in a way you could understand? Did a member of staff tell you about medication side effects to watch for when you went home? Were you told how to take your medication in a way you could understand? Were you given clear written or printed information about your medicines? Did a member of staff tell you about any danger signals you should watch for after you went home? Did hospital staff take your family or home situation into account when planning your discharge? Did the doctors or nurses give your family or someone close to you all the information they needed to help care for you? Do you feel that you received enough information from the hospital on how to manage your condition after your discharge? Did you receive copies of letters sent between hospital doctors and your family doctor (GP)?

OVERALL

Overall, did you feel you were treated with respect and dignity while you were in the hospital? How would you rate how well the doctors and nurses worked together? Overall, were you treated with kindness and understanding while you were in the hospital? Overall, how would you rate the care you received? Would you recommend this hospital to your family and friends? During your hospital stay, were you ever asked to give your views on the quality of your care?

ABOUT YOU OR YOUR CHILD

Are you male or female? What was your year of birth? How old were you when you left full-time education? In the 12 months before this admission, how many days of (paid or unpaid) work or school have you had to miss as a result of your ulcerative colitis? Please enter the number in the box below

ADOLESCENT SPECIFIC SECTION (AGED 13 TO 18 YEARS OF AGE)

In your opinion, was the ward you stayed on suitable for a person of your age? Did the hospital staff involve you personally (not your family) enough in making decisions about your care? In your opinion, did the doctors know enough about how your condition affects people of your age? In your opinion, did the nurses who treated you know enough about how your condition affects people of your age? Did any member of staff give you advice about how to manage your IBD either at school or at work after you left hospital?

YOUR OWN HEALTH STATE TODAY

EQ-5D questionnaire: possible responses include (no problems, some problems, extreme problems) on Mobility, Self-care, Usual activities, Pain/discomfort, Anxiety/depression. Do you have any of the following long-standing conditions in addition to IBD?

WHO HAS COMPLETED THIS SURVEY

I completed the questionnaire myself and I am aged 12 years or over, A parent/guardian/carer has completed the questionnaire on behalf of child who is under the age of 12 years

OTHER COMMENTS

If there is anything else you would like to tell us about your experiences in the hospital, please do so here. Was there anything particularly good about your hospital care? Was there anything that could be improved? Any other comments?

Reference List

(1) Wille N, Badia X, Bonsel G, Burstrom K, Cavrini G, Devlin N, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. Qual Life Res 2010 Aug;19(6):875-86.

(2) Stevens K. Working with children to develop dimensions for a new preference-based, generic, pediatric, health-related quality of life measure. Qualitative Health Research 2010;20:340-51.

(3) Feeny D, Furlong W, Barr RD. Multiattribute approach to the assessment of health-related quality of life: Health Utilities Index. Med Pediatr Oncol 1998;Suppl 1:54-9.

(4) Varni JW, Seid M, Rode CA. The PedsQL: measurement model for the pediatric quality of life inventory. Med Care 1999 Feb;37(2):126-39.

(5) Longworth L, Yang Y, Young T, Mulhern B, Hernandez AM, Mukuria C, et al. Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: A systematic review, statistical modelling and survey. Health Technol Assess 201418(9):1-224. Available from: URL: http://www.journalslibrary.nihr.ac.uk/__data/assets/pdf_file/0008/108368/FullReport-hta18090.pdf

(6) Konig HH, Ulshofer A, Gregor M, von TC, Reinshagen M, Adler G, et al. Validation of the EuroQol questionnaire in patients with inflammatory bowel disease. Eur J Gastroenterol Hepatol 2002 Nov;14(11):1205-15.

(7) Stark RG, Reitmeir P, Leidl R, Konig HH. Validity, reliability, and responsiveness of the EQ-5D in inflammatory bowel disease in Germany. Inflamm Bowel Dis 2010 Jan;16(1):42-51.

(8) von der Schulengurg JM, Claes C, Greiner W, Uber A. Die deutsche Version des EuroQol-Fragebogens. Z Gesundheitswissenschaften 1998;6:3-20.

(9) Greiner W, Claes C, Busschbach JJ, von der Schulenburg JM. Validating the EQ-5D with time trade off for the German population. Eur J Health Econ 2005 Jun;6(2):124-30.

(10) Dolan P, Gudex C, Kind P, Williams A. The time trade-off method: results from a general population study. Health Econ 1996 Mar;5(2):141-54.

(11) Borgaonkar MR, Irvine EJ. Quality of life measurement in gastrointestinal and liver disorders. Gut 2000;47:444-54.

(12) Best WR, Becktel JM, Singleton JW, Kern F, Jr. Development of a Crohn's disease activity index. National Cooperative Crohn's Disease Study. Gastroenterology 1976 Mar;70(3):439-44.

(13) Rachmilewitz D. Coated mesalazine (5-aminosalicylic acid) versus sulphasalazine in the treatment of active ulcerative colitis: a randomised trial. BMJ 1989 Jan 14;298(6666):82-6.

(14) Royal College of Physicians. Report of the results for the national clinical audit of adult inflammatory bowel disease inpatient care in Scotland. Round 3 Scotland. London; 2012.

(15) Royal College of Physicians. Report of the results for the national clinical audit of paediatric inflammatory bowel disease inpatient care in the UK. National Report Version 2. London; 2012.

(16) Royal College of Physicians. The inaugural national report of the results for the primary care questionnaire responseds. Part of the UK inflammatory bowel disease audit 3rd round. London; 2012.

(17) Royal College of Physicians. National clinical audit of inpatient care for young people with ulcerative colitis. UK IBD audit. 2014.

(18) Herts Valley Clinical Commissioning Group, East and North Hertfordshire Clinical Commissioning Group. Diagnosis and management of irritable bowel syndrome in primary care. www hertsvalleysccg nhs uk/uploads/file/Pharmacy/Local%20Decisions/IBS%20Pathway%20v4 pdf 2014 [cited 14 A.D. May 3];Available from: URL: www.hertsvalleysccg.nhs.uk/uploads/file/Pharmacy/Local%20Decisions/IBS%20Pathway%20v4.pdf

(19) European Medicines Agency. Guideline on the evaluation of medicinal products for the treatment of irritable bowel syndrome. European Medicines Agency; 2002.

(20) European Medicines Agency. Points to consider on the evaluation of medicinal products for the treatment of irritable bowel syndrome. European Medicines Agency; 2003.

(21) European Medicines Agency. Concept paper on the revision of the CHMP points to consider on the evaluation of medicinal products for the treatment of irritable bowel syndrome. European Medicines Agency; 2012.

(22) National Collaborating Centre for Nursing and Supportive Care. Clinical Practice Guideline. Irritable bowel syndrome in adults: diagnosis and management or irritable bowel syndrome in primary care. National Institute for Health and Clinical Excellence (NICE); 2008.

(23) European Medicines Agency. Guideline on the development of new medicinal products for the treatment of Crohn's disease. European Medicines Agency; 2007.

(24) Horizon Scanning Centre. Tofacitinib for moderate to severe active ulcerative colitis. www hsc nihr ac uk 2014 [cited 14 A.D. May 3];5752Available from: URL: www.hsc.nihr.ac.uk

(25) Khanna R, D'haens G, Feagan BG, et al. Evaluation of the operating properties of candidate patient reported outcomes for use in randomized controlled trials in Crohn's disease. Inflammatory Bowel Dis. In press 2013.

(26) Achleitner U, Coenen M, Colombel JF, Peyrin-Biroulet L, Sahakyan N, Cieza A. Identification of areas of functioning and disability addressed in inflammatory bowel disease-specific patient reported outcome measures. J Crohns Colitis 2012 Jun;6(5):507-17.

(27) The IBD Standards Group. Standards for the Healthcare of People who have Inflammatory Bowel Disease. 2013.

(28) Fazio VW, O'Riordain MG, Lavery IC, Church JM, Lau P, Strong SA, et al. Long-term functional outcome and quality of life after stapled restorative proctocolectomy. Ann Surg 1999 Oct;230(4):575-84.

(29) Love JR, Irvine EJ, Fedorak RN. Quality of life in inflammatory bowel disease. J Clin Gastroenterol 1992 Jan;14(1):15-9.

(30) Guyatt G, Mitchell A, Irvine EJ, Singer J, Williams N, Goodacre R, et al. A new measure of health status for clinical trials in inflammatory bowel disease. Gastroenterology 1989;96:804-10.

(31) Joachin G, Milne B. Inflammatory bowel disease: effects on lifestyle. Journal of Advanced Nursing 1987;12(483-487).

(32) Alcala M, Casellas F, Prieto L, Malagelada JR. Development of a short questionnaire for quality of life specific for inflammatory bowel disease. Gastroenterology 2001;120(Suppl 1):A450.

(33) Alcala MJ, Casellas F, Fontanet G, Prieto L, Malagelada JR. Shortened questionnaire on quality of life for inflammatory bowel disease. Inflamm Bowel Dis 2004 Jul;10(4):383-91.

(34) Drossman DA, Leserman J, Li ZM, Mitchell CM, Zagami EA, Patrick DL. The rating form of IBD patient concerns: a new measure of health status. Psychosom Med 1991 Nov;53(6):701-12.

(35) Irvine EJ, Zhou Q, Thompson AK. The Short Inflammatory Bowel Disease Questionnaire: a quality of life instrument for community physicians managing inflammatory bowel disease. CCRPT Investigators. Canadian Crohn's Relapse Prevention Trial. Am J Gastroenterol 1996 Aug;91(8):1571-8.

(36) Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. Pharmacoeconomics 1993 Nov;4(5):353-65.

(37) IBD Registry. IBD Registry 2014Available from: URL: http://www/ibdregistry.org.uk

(38) Bodger K, Ormerod C, Shackcloth D, Harrison M. Development and validation of a rapid, generic measure of disease control from the patient's perspective: the IBD-Control questionnaire. Gut 2013 Oct 9.

(39) Wilburn J, McKenna SP, Twiss J, L'amri M, Kemp K, Campbell S. Development and Validation of the Crohn's Life Impact Questionnaire (CLIQ). 2013.

(40) Ogden CA, Akobeng AK, Abbott J, Aggett P, Sood MR, Thomas AG. Validation of an instrument to measure quality of life in British children with inflammatory bowel disease. J Pediatr Gastroenterol Nutr 2011 Sep;53(3):280-6.

(41) Abdullah B, Murphy T, Langer SL, Romano J, Levy R. Physician measures of disease severity significantly predict health-related quality of life in children with IBD. Gastroenterology 2014146(5 SUPPL. 1):S-206.

(42) Hutchings HA, Upton P, Cheung WY, Maddocks A, Eiser C, Williams JG, et al. Adaptation of the Manchester-Minneapolis Quality of Life instrument for use in the UK population. Arch Dis Child 2007 Oct;92(10):855-60.

(43) Hutchings HA, Upton P, Cheung W-Y, Maddocks A, Eiser C, Williams JG, et al. Development of a parent version of the Manchester-Minneapolis quality of life survey for use by parents and carers of UK children: MMQL-UK (PF). Health Qual Life Outcomes 20086

(44) Abdovic S, Mocic PA, Milosevic M, Persic M, Senecic-Cala I, Kolacek S. The IMPACT-III (HR) questionnaire: a valid measure of health-related quality of life in Croatian children with inflammatory bowel disease. J Crohns Colitis 2013 Dec 1;7(11):908-15.

(45) Duffy CM, Wells GA, Russell AS, Haraoui B. Quality of life issues in pediatric immune-mediated inflammatory disease. J Rheumatol 201138(SUPPL. 88):20-25. Available from: URL: http://jrheum.org/content/supplements/88/20.full.pdf+html

(46) Marcus SB, Strople JA, Neighbors K, Weissberg-Benchell J, Nelson SP, Limbers C, et al. Fatigue and health-related quality of life in pediatric inflammatory bowel disease. Clin Gastroenterol Hepatol 2009 May;7(5):554-61.

(47) Perrin JM, Kuhlthau K, Chughtai A, Romm D, Kirschner BS, Ferry GD, et al. Measuring quality of life in pediatric patients with inflammatory bowel disease: psychometric and clinical characteristics. J Pediatr Gastroenterol Nutr 2008 Feb;46(2):164-71.

(48) Upton P, Eiser C, Cheung I, Hutchings HA, Jenney M, Maddocks A, et al. Measurement properties of the UK-English version of the Pediatric Quality of Life Inventory 4.0 (PedsQL) generic core scales. Health Qual Life Outcomes 2005;3:22.

(49) Varni JW, Bendo CB, Denham J, Shulman RJ, Self MM, Neigut DA, et al. PedsQL Gastrointestinal Symptoms Module: Feasibility, Reliability, and Validity. J Pediatr Gastroenterol Nutr 2014 May 5.

(50) Lane M, Varni JW, Czyzewski DI, Shulman RJ. Health-related quality of life in pediatric functional vs. Organic gastrointestinal (GI) disorders: The pedsql gastrointestinal symptoms module. Gastroenterology 2012142(5 SUPPL. 1):S382.

(51) Griffiths AM, Nicholas D, Smith C, Munk M, Stephens D, Durno C, et al. Development of a quality-of-life index for pediatric inflammatory bowel disease: dealing with differences related to age and IBD type. J Pediatr Gastroenterol Nutr 1999 Apr;28(4):S46-S52.

(52) Otley A, Smith C, Nicholas D, Munk M, Avolio J, Sherman PM, et al. The IMPACT questionnaire: a valid measure of health-related quality of life in pediatric inflammatory bowel disease. J Pediatr Gastroenterol Nutr 2002 Oct;35(4):557-63.

(53) Otley A. Measurement of quality of life in pediatric inflammatory bowel disease. In: Mamula P, Baldassano RN, ., Markowitz JE, editors. Pediatric Inflammatory Bowel Disease. 1st ed. New York: Springer; 2008. p. 565-80.

(54) Hyams JS, Ferry GD, Mandel FS, Gryboski JD, Kibort PM, Kirschner BS, et al. Development and validation of a pediatric Crohn's disease activity index. J Pediatr Gastroenterol Nutr 1991 May;12(4):439-47.

(55) Kundhal PS, Critch JN, Zachos M, Otley AR, Stephens D, Griffiths AM. Pediatric Crohn Disease Activity Index: responsive to short-term change. J Pediatr Gastroenterol Nutr 2003 Jan;36(1):83-9.

(56) Griffiths AM, Otley AR, Hyams J, Quiros AR, Grand RJ, Bousvaros A, et al. A review of activity indices and end points for clinical trials in children with Crohn's disease. Inflamm Bowel Dis 2005 Feb;11(2):185-96.

(57) Kovacs M. Children's Depression Inventory (CDI): technical manual update. Toronto: Multi-Health Systems; 2003.

(58) Landgraf JM, Abetz L. Functional status and well-being of children representing three cultural groups: initial self reports using the CHQ-CF87. Psychology and Health 1997;12:839-54.

(59) Waters E, Davis E, Ronen GM, Rosenbaum P, Livingston M, Saigal S. Quality of life instruments for children and adolescents with neurodisabilities: How to choose the appropriate instrument. Dev Med Child Neurol 200951(8):660-669.

(60) Eiser C, Morse R. A review of measures of quality of life for children with chronic illness. Arch Dis Child 2001 Mar;84(3):205-11.

(61) National., National Institute for Health and Care Excellence. Infliximab for subacute manifestations of ulcerative colitis. 2008. Report No.: TA140.

(62) Wailoo A, Tosh J, Hemingway P. Use of Tumour Necrosis Factor Alpha (Tnf A) Inhibitors (Adalimumab and Infliximab) For Chrohns Disease. Decision Support Unit; 2009.

(63) Bryan S, Andronis L, Hyde C, Connock M, Fry-Smith A, Wang D. Infliximab for the treatment of acute exacerbations of ulcerative colitis. National Institute for Health and Clinical Excellence; 2008.

(64) Dretzke J, Edlin R, Round J, Connock M, Hulme C, Czeczot J, et al. A systematic review and economic evaluation of the use of tumour necrosis factor-alpha (TNF-alpha) inhibitors, adalimumab and infliximab, for Crohn's disease. Health Technol Assess 2011 Feb;15(6):1-244.

(65) Gregor JC, McDonald JW, Klar N, Wall R, Atkinson K, Lamba B, et al. An evaluation of utility measurement in Crohn's disease. Inflamm Bowel Dis 1997;3(4):265-76.

(66) Buxton MJ, Lacey LA, Feagan BG, Niecko T, Miller DW, Townsend RJ. Mapping from disease-specific measures to utility: an analysis of the relationships between the Inflammatory Bowel Disease Questionnaire and Crohn's Disease Activity Index in Crohn's disease and measures of utility. Value Health 2007 May;10(3):214-20.

(67) Armstrong N, Joore M, van AT, Misso K, Manning N, Tomini F, et al. Golimumab for the treatment of ankylosing spondylitis: a NICE single technology appraisal. Pharmacoeconomics 2013 May;31(5):415-25.

(68) Rodgers M, Epstein D, Bojke L, Yang H, Craig D, Fonseca T, et al. Etanercept, infliximab and adalimumab for the treatment of psoriatic arthritis: a systematic review and economic evaluation. Health Technol Assess 2011 Feb;15(10):i-329.

(69) Khanna R, D'haens G, Feagan BG, Sandborn WJ., Vandervoort MK, Zou G, et al. Patient reported outcome measures derived from the Crohn's disease activity index: correlation between PRO2 and PRO3 scores and CDAI-defined clinical thresholds. 2014.

(70) van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health 2012;15(5):708-15.

(71) Meeske K, Katz ER, Palmer SN, Burwinkle T, Varni JW. Parent proxy-reported health-related quality of life and fatigue in pediatric patients diagnosed with brain tumors and acute lymphoblastic leukemia. Cancer 2004 Nov 1;101(9):2116-25.

(72) Varni JW, Seid M, Kurtin PS. PedsQL 4.0: reliability and validity of the Pediatric Quality of Life Inventory version 4.0 generic core scales in healthy and patient populations. Med Care 2001 Aug;39(8):800-12.

(73) Varni JW, Burwinkle TM, Seid M, Skarr D. The PedsQL 4.0 as a pediatric population health measure: feasibility, reliability, and validity. Ambul Pediatr 2003 Nov;3(6):329-41.

(74) Varni JW, Sherman SA, Burwinkle TM, Dickinson PE, Dixon P. The PedsQL Family Impact Module: preliminary reliability and validity. Health Qual Life Outcomes 2004;2:55.

(75) Varni JW, Limbers CA, Burwinkle TM. Impaired health-related quality of life in children and adolescents with chronic conditions: a comparative analysis of 10 disease clusters and 33 disease categories/severities utilizing the PedsQL 4.0 Generic Core Scales. Health Qual Life Outcomes 2007;5:43.