Development and initial psychometric validation of a patient reported outcome measure for Crohn's perianal fistula –

the Crohn's Anal Fistula Quality of Life (CAF-QoL) Scale

Samuel O. Adegbola[1,2], Lesley Dibley[3], Kapil Sahnan[1,2], Tiffany Wade[5], Azmina Verjee[4], Rachel Sawyer[4], Sameer Mannick[4], Damian McCluskey[4], Paul Bassett[1], Nuha Yassin[1], Janindra Warusavitarne[1,2], Omar D Faiz[1,2] Robin KS Phillips[1,2], Philip J Tozer[1,2], Christine Norton[5], Ailsa L Hart[1,2]

Short title:	Crohn's perianal fistula PROM – the CAF-QoL scale
Category:	Quality of Life, socio-economical and psychological end points
Word count:	main text (4590), abstract (201)
Keywords:	Crohn's anal fistula; patient reported outcomes; quality of life

Correspondence to:

Samuel O. Adegbola, St. Mark's Hospital, Watford Road, Harrow HA1 3UJ.
Email: samadeg@doctors.net.uk Tel: (+44) 0208 869 5328 mob: (+44) 07949 900 488

Author Institutions:

1. Robin Phillips Fistula Research Unit, St Mark's Hospital & Academic Institute, Harrow, Middlesex, HA1 3UJ, United Kingdom
2. Department of Surgery, Imperial College, St Mary's Hospital, Praed Street, London W21NY, United Kingdom.
3. Faculty of Education, Health and Human Sciences, University of Greenwich, London UK
4. CAF-QoL Patient and Public Involvement (PPI) Team, UK
5. Florence Nightingale Faculty of Nursing, Midwifery & Palliative Care, King's College London, UK.

Affiliations

Mr Samuel O Adegbola BSc (Hons), MBBS, MRCS – Clinical Research Fellow

Dr Lesley Dibley PhD, RN, RSCN – Reader in Nursing Research and Education

Mr Kapil Sahnan BSc (Hons), MBBS, MRCS – Clinical Research Fellow

Ms Tiffany Wade MSc – Qualitative Researcher

Ms Azmina Verjee – Patient Representative, CAF-QoL PPI Team

Ms Rachel Sawyer – Patient Representative, CAF-QoL PPI Team

Mr Sameer Mannick – Patient Representative, CAF-QoL PPI Team

Mr Damian McCluskey – Patient Representative, CAF-QoL PPI Team

Mr Paul Bassett Msc – Statistician

Ms Nuha Yassin PhD FRCS – Consultant Colorectal Surgeon

Mr Janindra Warusavitarne PhD FRCS – Consultant Colorectal Surgeon

Prof Omar D Faiz MD FRCS – Consultant Colorectal Surgeon, Professor of Practice

Prof Robin K.S Phillips FRCS, MS - Professor of Colorectal Surgery

Mr Philip J Tozer MD FRCS – Consultant Colorectal Surgeon

Prof Christine Norton PhD MA RN – Professor of Clinical Nursing Research

Prof Ailsa Hart BA(Hons), BMBCh, FRCP, PhD – Consultant Gastroenterologist, Professor of Practice

Author Contributions: All authors have contributed substantially to this manuscript design/delivery. AH, CN, JW, ODF, RKSP, were involved in the design of the study. AV, RS, DM, SM served as patient and public involvement and gave critical input and refinements to design. SOA, LD carried out patient interviews. SOA, KS, TW, PT, CN, AH, NY, AV, RS, SM, DM, performed independent qualitative analysis of audio transcripts. SOA, PB performed statistical analysis. SOA, KS, PT, LD wrote the article. CN, AH, JW, ODF, NY reviewed the article drafts providing refinements and steer. All authors reviewed and approved final draft.

Abbreviations

Core Outcome Set (COS)

Crohn's disease (CD)

Factor analysis (FA)

Harvey-Bradshaw Index (HBI)

Abbreviations (continued)

Hospital Anxiety and Depression Scale (HADS)

Inflammatory bowel disease (IBD)

Intraclass correlation (ICC) coefficient

Magnetic resonance imaging (MRI)

Principal Axis Factoring (PAF)

Patient reported outcomes measure (PROM)

Patient and Public Involvement (PPI)

Perianal Disease Activity Index (PDAI)

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)

Quality of life (QoL)

United Kingdom Inflammatory Bowel Disease Questionnaire (UK-IBDQ),

Abstract

Introduction

Crohn's perianal fistulas are challenging for patients and clinicians. Many do not respond to available treatments and despite recommendations by a global consensus, there are currently no specific patient-derived quality of life (QoL) tools to measure response to treatment. We present a new validated patient reported outcomes measure (PROM) for this complicated disease phenotype.

Methods

A draft questionnaire was generated using unstructured qualitative patient interviews on the experience of living with Crohn's perianal fistula, a nationwide multidisciplinary consensus exercise, a systematic review of outcomes assessing medical/surgical/combined treatment and a patient and public involvement day. Psychometric properties were assessed including construct validity (by comparison with the Hospital Anxiety and Depression Scale (HADS) and the United Kingdom Inflammatory Bowel Disease Questionnaire (UK-IBDQ), and reliability and responsiveness was assessed by test-retest analysis.

Results

Data from 211 patients contributed to development of a final 28-item questionnaire. The CAF-QoL demonstrated good internal consistency (Cronbach's alpha 0.88), excellent stability (intra-class correlation 0.98) and good responsiveness and construct validity, with positive correlation with the UK IBDQ and HADS.

Conclusion

The CAF-QoL scale is ready for use as a PROM in research and clinical practice. It complements objective clinical evaluation of fistula by capturing impact on the patient.

What is already known about this subject?

There is heterogeneity in outcome measurement in Crohn's perianal fistula, limiting comparison of treatment options, with no widely accepted gold standard. There is currently no patient reported outcome measure (PROM) for Crohn's perianal fistula, representing an unmet need, which was highlighted in a published core outcome set for perianal Crohn's disease.

What are the new findings?

The development and <u>initial </u>validation of a new PROM for Crohn's perianal fistula, the Crohn's anal fistula quality of life (CAF-QoL) scale.

How might it impact on clinical practice in the foreseeable future?

This new PROM lends itself to utility in clinical trial design for Crohn's perianal fistulas, enabling consistency in datasets collected in this condition. It also has the potential utility of assessment of disease impact in a clinic setting, to guide stratification of patients according to severity of disease impact and tailor appropriate management.

Introduction

Perianal fistulas occur in a third of all Crohn's disease (CD) patients[1]. They represent a distinct and aggressive phenotype of CD[1,2] and often follow a chronic course with symptoms including anal pain and purulent discharge, commonly leading to a severely impaired quality of life (QoL). Crohn's perianal fistulas are challenging to treat; often complex in nature, they can be refractory to conventional medical treatment strategies such as antibiotics, immunomodulators[3,4] and biologic drugs, such as anti-TNF agents[5–9]. Surgical treatments fare little better and despite a multidisciplinary approach (i.e. concomitant surgical and medical therapies), most patients experience recurrence or persistence. The associated morbidity of the disease and its treatment can have profound effects on patients' physical and psychosocial wellbeing.

In clinical trials, success is usually measured by clinical assessment of closure of fistula tracks, sometimes accompanied by radiological assessment of 'healing'. This is appropriate in trials of treatment with 'curative intent', but as most patients do not achieve sustained fistula closure, there is a need to measure any benefit produced by treatments in situations where fistulas do not 'heal'. Further, it is important to assess additional impacts of any intervention on QoL even if successful 'healing' occurs. For example, continence impairment might occur after a 'curative' fistulotomy or advancement flap repair. There are also interventions, such as setons, which are not designed to heal fistulas and are performed with intent to ameliorate symptoms rather than cure. QoL assessment may represent the primary outcome in such situations.

The only current measure designed to assess Crohn's perianal fistula activity and its impact on QoL is the Perianal Disease Activity Index (PDAI) [10]. This clinical assessment tool aims to measure disease activity in patients with Crohn's perianal fistulas. It assesses pain, restriction of physical and sexual activities, and perianal disease severity (discharge, disease type and induration). Items are scored from 0 (no problem) to 4 (severe problem) on a Likert scale[10–12]. Although the PDAI was developed to measure clinical disease activity in patients with "perianal disease", it is not specific to perianal fistula and was not developed using methodology that conformed to accepted principles of evaluative index development[13]. A significant limitation to its use as a patient reported outcome measure is the lack of patient involvement in its development, relying instead on QoL issues which physicians believed to be important to patients. Thus, it does not assess the global QoL impact of fistulas on patients, and its relevance to what patients themselves consider as important is unknown. Recent guidelines recommend that improved instruments are needed to score perianal CD fistula activity and impact on QoL[8]. In particular, a tool which assesses the impact of fistula(s) on a patient's QoL and which detects a meaningful change in their QoL after treatment is required. Early exploratory qualitative work demonstrates that the impact of CD perianal fistulas extends far beyond restricting daily and sexual activities[14,15] and a patient-centred, patient-derived tool is needed. We aimed to develop and undertake initial psychometric validation of a new QoL scale for CD patients with perianal fistulas ensuring patients were involved in all phases of development.

Methods

Overview

The study used a three-phase mixed methods design utilising an exploratory instrument development model[16,17] to support a QoL patient-reported outcome measure (PROM) development process [18]. Phase 1 used an exploratory qualitative approach[19] to create a longlist of items. Phase 2 involved cognitive interviews to refine items. Phase 3 subjected the draft PROM to psychometric testing via completion by participants with Crohn's perianal fistula and completed final item reduction.

Study Steering group and Patient and Public Involvement (PPI)

Members of the study steering group included all relevant stakeholders: a colorectal surgeon with expertise in fistulas, a gastroenterologist with expertise in inflammatory bowel disease (IBD), IBD specialist nurses (2), researchers with experience in PROM design[20] (3) and patient representatives (4). The latter constituted our patient and public involvement team and consisted of four members of Crohn's and Colitis UK (CCUK) charity. All had Crohn's disease and previous or current experience of living with a perianal fistula. The PPI team helped with analysis and contributed to discussion and development of the preliminary items for the draft questionnaire and final item reduction. The study design was also presented at a dedicated PPI day for critical feedback and to gather suggestions to optimise recruitment and response rates.

Recruitment and Sampling

Participants were recruited using purposive sampling from the membership (via advertisement) of collaborating registered patient charities including those specific to IBD / bowel diseases (ForCrohns, Crohn's and Colitis UK, Bowel Disease Research Foundation) and UK-wide patient repositories (UK IBD Bioresource). Recruitment also occurred via IBD outpatient clinics, aiming to recruit equal numbers of men and women, with a broad age range.

Selection of participants

Inclusion criteria: over 16 years; living in the UK; diagnosis of Crohn's disease with perianal fistula; ability to read, speak and understand English; ability to give informed consent.

Exclusion criteria: Patients with resolved fistulas were excluded from Phase 3 and participants contributing to any phase of the study were excluded from contribution to another.

Ethical considerations

The study was approved by King's College London (ref: HR-16/17-4421) and the UK Health Research Authority (ref: 17/LO/1563). Informed consent was collected prior to data collection. Participants chose which phase of the study they wanted to be involved in.

Data Collection / Generation of a draft PROM

Phase 1 – Experience interviews / data from outcomes review and consensus exercise

a) Individual unstructured interviews were conducted face to face or by telephone/video call according to participant preference, to explore the experience of living with CD-related perianal fistula(s) and the impact of the disease and surgical and/or medical treatments on the individual. This facilitated an in-depth understanding of the complex experience of living with Crohn's perianal fistulas. Interviews were digitally recorded and transcribed verbatim by a professional transcriber. A thematic analysis[15,21] was undertaken to generate a long-list of items for inclusion in the draft PROM. Individual followed by steering group analysis and agreement ensured that findings represented a consensus of the steering group[15].

b) To enhance the face and content validity of the questionnaire, we included data from a systematic review of fistula outcome measures and a consensus exercise undertaken to develop a Core Outcome Set (COS) for Crohn's perianal fistulas[22]. The systematic review followed PRISMA guidelines and involved a search limited to studies conducted in adults aged ≥18 years old and papers published between 1 January 2010 and 12 July 2016. Reference lists were searched to ensure complete capture. We assessed the results to identify patient reported outcomes and included these in the long list of items for the draft PROM. The COS involved stakeholder recruitment from across the UK[22], including clinicians (colorectal surgeons, gastroenterologists, radiologists and IBD nurses) and patients, the latter being the primary stakeholders. Qualitative data from patient focus groups (discussing QoL and outcome measurements) in the COS exercise were discussed by the consensus steering group and included in the long list of items for the draft PROM.

At the end of phase 1, items for inclusion in the PROM (*the Crohn's Anal Fistula Quality of Life, CAF-QoL Scale*) were collated using a mixture of questions and statements. Scoring systems (mixture Likert scale / closed ended ordinal answers) were added following steering group consensus. A background section (16 questions) collected demographic data and disease details (including duration, number of fistulas, medication and surgical history). This section also had an anchor question, using a 10-point Likert scale to measure self-reported global rating of fistula status. This enabled transitional analysis on the basis of the participants' subjective perceptions of their fistula.

Phase 2 – Cognitive interviews / pretesting of draft PROM

Participants in this phase provided further content validation via four cognitive interview rounds (involving three to four individual interviews per round), to enable participants to refine the questions using think-aloud techniques and verbal probing as they completed the draft PROM. Interviews were digitally recorded and transcribed by a professional transcriber. Following each round of cognitive interviews, items were reviewed by the steering group to ensure any suggested changes were agreed by consensus. Subsequent rounds of interviews were used to confirm acceptability until data saturation was achieved and no new suggestions were received.

Phase 3 - Testing of the new Crohn's Anal Fistula-Quality of Life (CAF-QoL) PROM

Participants with CD and active (i.e. presence of symptoms related to) perianal fistulas were sent the draft CAF-QoL PROM by post or online and asked to complete it again two weeks later for test-retest reliability. At the initial administration of the CAF-QoL PROM, participants were also asked to complete a demographic details form, the validated UK-

IBDQ[23], and the Hospital Anxiety and Depression Scale (HADS)[24]. Overall and perianal fistula disease activity information were collected on both occasions, using a modified version of the Harvey-Bradshaw Index (HBI)[25,26] and self-reported global rating of fistula status. This allowed stratification according to change in overall disease activity level between the test and retest analysis and allowed a sub-analysis for sensitivity to change in fistula status.

Testing of psychometric properties

Validity

Content Validity assessed the applicability, relevance, and clarity of question items in order to maximise the accurate completion of the questionnaire. Further to verification of content validity in Phase 1 and 2, it was explored by evaluating the levels of missing data per item in phase 3. The overall response rate was analysed to indicate the feasibility of the questionnaire for self-completion.

Construct Validity assessed the consistency of the scale with other instruments known to assess similar attributes with good validity and reliability. Two validated outcome measures (UK-IBDQ [23]and HADS[24]) were used as comparators to evaluate the relationships between patients' reports as these measures cover some similar concepts. Correlation was assessed using Pearson's correlation coefficient (r).

Reliability

Test-retest reliability (reproducibility or stability) assessed consistency between the two completions of the draft CAF-QoL[27,28]. Respondents completing the phase 3 test-retest in whom the IBDQ, HADS and anchor questions indicated stable disease were used to assess

reliability[23,29,30]. Relevant anchor questions were a general rating of fistula status (on a Likert scale of 0-10) and a question ascertaining whether there had been any flare up of their fistula (requiring medication, operation or seton) since the previous questionnaire completion. Those who reported no changes (absence of flare up, ≤ 1 point difference on Likert scale and stable overall IBD disease activity as defined by stable HBI scores) were included in the reproducibility analysis. The difference between test and retest responses was appraised using weighted statistics (the weighted kappa method[31,32]) due to the categorical nature of response options. Kappa is measured on a scale ranging up to a maximum agreement of 1, with values signifying a good (>0.6) and very good (>0.8) strength of agreement[33]. Reproducibility of the total CAF-QoL scores for stable patients was also assessed using the intraclass correlation (ICC) coefficient. The ICC is the proportion of total variation in CAF-QoL scores between test and retest completions that is due to variation between respondents (inter-respondent variation), as opposed to variation within respondents (intra-respondent variation). A high agreement between test and retest scores implies little intra-participant variation and thus an ICC value close to 1 with values exceeding 0.75 generally accepted as a marker of good reproducibility[27].

Sensitivity to Change

Sensitivity to change (or responsiveness) was assessed in retested patients who reported significant change (i.e. three or more points on the Likert scale) in their general rating of fistula status. Comparisons between baseline and retest overall CAF-QoL scores were assessed using mean change between successive completion scores, with a corresponding confidence interval. Changes with p-values of <0.05 were deemed significant.

Item reduction – questionnaire refinement

Factor analysis (FA) with Principal Axis Factoring (PAF) extraction and varimax rotation were used to determine individual dimensions or subscales of the CAF-QoL and to refine the number of items[34]. The internal reliability (i.e. correlation between different items) was determined using Cronbach's alpha (values exceeding 0.7 indicate good consistency)[35]. The strength of association between individual questions was assessed using Pearson's correlation coefficient. Those with strong associations (correlation coefficient > 0.75) were reviewed by the steering group to achieve consensus on excluding questions deemed to duplicate information and those unlikely to change with intervention. Contentious items were discussed, and consensus agreed on exclusion of questions following review of those with missing data from a majority of participants (regardless of fistula status), questions with poor test-retest agreement (suggesting unreliability), questions with strong associations with others (i.e. suggesting duplicated themes) and questions with poor spread of response.

Results

A total of 211 participants with Crohn's perianal fistula were involved in the three study phases (Table 1). Most participants (~60%) reported having IBD for longer than 10 years and over 80% had experienced Crohn's perianal fistulas for at least one year. Fewer than 15% had a stoma (temporary or permanent).

Table 1: Participant characteristics

Characteristic		Number (n)	Proportion of study group
Gender	Female	122	57.8%
	Male	88	41.7%
	Transgender	1	0.5%
Age (*)	42.9 ± 12.7(*)		
Duration of Crohn's disease	< 1 year	7	3.3%
	1 – 4 years	31	14.7%
	5 – 9 years	47	22.2%
	10 – 14 years	34	16.1%
	15+ years	92	43.6%
Duration of perianal fistula	< 1 year	18	8.5%
	1 – 4 years	67	31.8%
	5 – 9 years	61	28.9%
	10 – 14 years	25	11.8%
	15+ years	40	18.9%
Current / previous medication for perianal fistula	No	182	86.3%
	Yes	29	13.7%
Previous surgery for perianal fistula	No	61	28.9%
	Yes	150	71.1%
Previous seton	Yes	131	62.1%
	No	59	27.9%
	Unsure	21	10.0%
Stoma due to perianal fistula	Yes	31	14.7%
	No	180	85.3%
Self-reported global fistula rating**	Remission (0)	21/184***	11.4%
	Mild (1-3)	42/184	22.8%
	Moderate (4-6)	43/184	23.3%
	Severe ($\geq$7)	78/184	42.4%

(*) Mean ± standard deviation reported, (**) rating on 10 point Likert scale (***) denominator is 184, i.e. number of phase 3 respondents.

Devising and pretesting the draft CAF-QoL Scale: Phases 1 and 2

Twelve interviews were conducted (median of 43 minutes, range 18 – 145 minutes) achieving apparent data saturation. Three broad themes emerged: *burden of symptoms*, *burden of treatment* and *impact on emotional, physical and social well-being* (Figure 1). Each included several sub-themes, with considerable interplay between these[15]. The impact of perianal fistula(s) on patients with CD was intense and wide reaching, negatively affecting intimate, close and social relationships and causing losses in life and work-related opportunities[36]. Data from the qualitative interviews were combined with that from the core outcome set exercise which involved more than 230 stakeholders from across the UK[22], including 80 patients with Crohn's perianal fistula. Data from the COS included the systematic review of patient reported outcomes, qualitative data from patient focus groups, and the final list of core outcomes. An original long list of 45 items was collated for the draft PROM and these were reviewed by the steering group, reducing these to 35 items following elimination of repetition and overlap. Those retained were grouped under three domains (A-C) mirroring the broad themes described above and converted into questions (Domain A – symptoms) and statements (Domain B – treatment and Domain C – QoL impact). Likert scales were adopted measuring frequency (0-4) in Domain A, and degree of agreement (Likert scales 0-4) in domains B and C. The questions and statements were phrased to ensure that positive answers (indicating good QoL) were always at the lower end of the Likert scale, and higher scores indicated worse QoL. Pretesting occurred during cognitive interviews (n = 15; average questionnaire completion time was 9 minutes) and no additional items were suggested for inclusion.

Testing the CAF-QoL: Phase 3

Patient sample

The test data consisted of responses from 184 patients. Three (1.6%) were excluded due to excessive missing data leaving 181 CAF-QoL questionnaires for analysis. The retest data had a response rate of 79% (143/181) completing the questionnaire.

Data spread and completeness

Results of the test data indicated that the majority of responses were spread across the available categories, demonstrating that the questionnaire can discriminate between patients with a range of symptoms and concerns. There was little missing data with fewer than 1% of respondents missing sporadic responses. Most questions were applicable to the majority of respondents. Some questions had 'not applicable' responses. These were the questions relating to presence of seton or side effects of medication and was highest for the question relating to presence of a stoma (not applicable to 83% of respondents). Analysis of the association between questions demonstrated that there were no 'perfect' correlations, with no correlation coefficient over 0.9. Therefore, no questions were perfectly duplicating information. However, there were some strong associations between questions (correlation coefficient >0.8), and these were considered for item reduction.

Item reduction

Nine questions were removed from the final CAF-QoL (see Appendix 1 for item reduction overview). Three questions pertaining to symptoms (Domain A) were removed. Two of these had very little spread in response categories, with low standard deviation, and responses were grouped toward the lower end of the Likert scale (indicating minimal relevance to most respondents). These questions related to the use of painkillers and the association of urinary

tract infections to the fistula. One question relating to the use of pads or gauze for the discharge from the fistula was removed due to strong association with the preceding question on fistula discharge, with consensus decision that there was probable overlap in the item being assessed. Four questions relating to fistula treatment (Domain B) were removed due to poor spread and poor test/retest agreement (in those reporting no change in global fistula status). Two questions relating to QoL impact (Domain C) were removed. These questions involving fistula effect on intimacy and career progression were found to be strongly associated with other questions (correlation coefficient > 0.83), replicating themes.

The final version of the questionnaire included 28 items, 27 of which were scored from 0 to 4, plus one free-text question (Figure 2), giving a score range of 0 – 108. Data from 181 respondents demonstrated scores of 0 – 95 (possible range 0-108), as depicted in Figure 3. Scores were approximately normally distributed with a mean score of 42.0, and a standard deviation of 26.0.

Internal consistency and factor analysis

Cronbach's alpha was 0.88, demonstrating a good degree of internal consistency amongst the individual items. A factor analysis was performed to determine if the score naturally broke down into different subscales. Only one factor was found to be significant, encompassing the majority of questions, indicating no sub-scales.

Reliability (agreement between test and retest scores)

Sixty-nine patients of the 143 retest respondents (48%) met the criteria for disease stability (as defined by ≤ 1 point difference between test/retest scores on global self-rating of fistula status, and stable disease on HBI and no fistula flare between test/retest completion).

Weighted Kappa was >0.7 for all items, demonstrating very good agreement. The mean difference between the retest and test measurements (calculated as retest minus test) was -0.6 with a standard deviation of 6.0. The agreement between test and retest scores was also examined using intra-class correlation (ICC) method. This method gave an ICC value of 0.98 (95% confidence interval: 0.96 to 0.99). This high value demonstrated very good agreement between the test and retest scores.

Sensitivity to change

In order to examine the change in CAF-QoL total score between test and retest in specific groups of patients, three groups were examined relating to patients with 'stable', 'improving' and 'worsening' global fistula rating (Table 2). In patients with stable disease there was no significant change in CAF-QoL scores between the two timepoints. CAF-QoL scores decreased significantly at the retest timepoint in those with improved disease, by an average of four units. Conversely in those with worsening global fistula rating scores increased significantly between timepoints, by a mean of 7 units.

Table 2: Changes in CAF-QoL score from test to retest

Patient group	n	Test Mean ± SD	Retest Mean ± SD	Change Mean (95% CI)	P-value
Stable [*]	69	41.5 ± 28.4	40.9 ± 27.8	-0.6 (-2.0, 0.9)	0.46
Improved [**]	12	52.2 ± 20.7	48.2 ± 22.3	-4.0 (-7.0, -1.0)	**0.01**
Worsening [***]	9	30.1 ± 17.3	37.7 ± 21.5	7.6 (1.6, 13.5)	**0.02**

[*] Defined as self-reported global fistula rating test/retest scores differing by no more than 1 point, no fistula flare up, and same test/retest Harvey-Bradshaw criteria
[**] Defined as self-reported global fistula rating retest score ≥ 3 units lower than test score
[***] Defined as self-reported global fistula rating retest score ≥ 3 units higher than test score
n - number of patients fulfilling criteria for analysis

Association with other health related measures

Assessment of the agreement between the total CAF-QoL scores and other health related measures (UK-IBDQ / HADS) used Pearson correlation (Table 3). The results demonstrated significant positive correlations between the total CAF-QoL score and each of the other measures. Higher CAF-QoL scores were associated with higher levels of anxiety and depression, and also worse IBD-Q scores.

Table 3: Association between CAF-QoL scores and other health scores

Variable	n	Correlation coefficient	P-value
HADS anxiety	174	0.52	<0.001
HADS depression	174	0.53	<0.001
UK IBD-Q	176	0.58	<0.001

n – number of respondents with completed questionnaires

Discussion

The Crohn's Anal Fistula Quality of Life Questionnaire (CAF-QoL) Scale has demonstrated good reliability, internal consistency and face/content validity. The robustness of the questionnaire is confirmed through the rigorous developmental process, with participants with Crohn's perianal fistula contributing throughout, not only as participants but also on the steering group. This has ensured that the language is straightforward, using lay terms that increase readability and enable accurate interpretation, and that the content is relevant to patients.

There is limited evidence in the literature on QoL in patients with anal fistula. Reported patient assessment of interventions highlight that patients have concerns with regards to anal fistulas and the treatment options, however, data collection is often rudimentary, involving patient satisfaction or incontinence[37,38]. The absence of a validated PROM means patients' concerns are often expressed based on clinicians' perspectives and this was highlighted in an Australian study by Wong et al.[39] comparing surgeon and patient preferences for surgical operations for idiopathic anal fistulas. The different cure and incontinence rates for each procedure were quoted for the patients to consider and their preferences explored. A clear mismatch was reported between what the surgeons and patients felt were important QoL issues. For example 91% of surgeons versus 25% of patients nominated continence as an important QoL issue[39]. The study demonstrated the potential assumptions made by clinicians with regards to the impact of fistula surgery on patients' QoL. More recently, Ferrer-Marquez et al.[40] developed a QoL questionnaire in patients with (non-Crohn's) anal fistula, however, major limitations of this study were the testing on a small sample (n=54) of Spanish-speaking patients, and crucially patients were not included in the initial development of the questionnaire. There is no objective patient-centred disease scoring tool for Crohn's perianal fistula and this results in an inability to quantify adequately the effect of Crohn's fistulas on psychosocial wellbeing and has led to calls to address the unmet need for a disease specific PROM[41].

Study strengths and limitations

The strength of the study lies in its design, incorporating a literature review, consensus exercise, patient focus groups and individual patient interviews[19] (unstructured and

cognitive) to inform the item generation. Recruitment of participants occurred via nationwide charities, hospitals, and social media, ensuring a broad sample of those with experience of Crohn's perianal fistulas. The involvement of patients both as participants and in the steering group ensured their voice was central to every stage of the process, facilitating creation of a true PROM. This is often lacking in PROM development[42]. Another strength was the diversity of stakeholders in the steering group (patient representatives, gastroenterologist, colorectal surgeon, specialist nurses, qualitative researchers) while ensuring that patient input was always prioritised.

Limitations include the absence of data collected on other manifestations of perianal Crohn's disease which often coexist with fistulas. This was a deliberate decision based on a consensus view to create a fistula specific PROM. Another potential limitation was the lack of access to exact data e.g. magnetic resonance imaging (MRI) and other anatomic fistula data for participants recruited via charities. However, the large number of patients contributing to the CAF-QoL, as well as the recruitment strategy (i.e. recruited from various locations and not just tertiary centres) mean that a spread of fistula/inflammatory characteristics is to be expected. Furthermore, patients reported different levels of symptoms at different points in the process (the basis of the test retest phase) and were asked to describe the longitudinal history of their fistula during the interview stage so we would expect variation in fistula 'phenotype'. Regarding analysis, respondents with temporary defunctioning stomas as opposed to following proctectomy, did not have separate analyses assessing the influence of either intervention on QoL. The group consensus was that such patients might well have different priorities, goals and symptoms, and separate analysis might have contributed to a score which was useful for neither group. During the data collection, those with stomas had no CD-specific marker of disease activity pertaining to looser stool, and the HBDAI question

on diarrhoea may have been ambiguous for these patients. However, only 14.6% (31/211) of the entire study cohort and 25/184 (13.6%) of the test-retest respondents had stomas and we made the assumption that disease activity might affect the fistula symptoms less in defunctioned patients, and thus not compromise the integrity of the score. In the analysis process, there were fewer numbers in the sensitivity to change / responsiveness analysis and a self -reported global fistula rating was used as a marker to assess transition in fistula status due to absence of a widely accepted/reliable clinical measure. Indeed, responsiveness and stability calculations were limited by the fact that there is no true gold standard to define a change in disease state. However, use of subjective transition questions is an accepted technique utilised in several outcome measure studies and can be advantageous in assessment of QoL because it directly addresses patients' perceptions of change over time and is short and simple[23,28–30].

Clinical usefulness and future direction

Assessing QoL in patients with IBD is an important component of medical and surgical management and clinical decision-making, and the last decade has seen a rapid increase in the number of measures to assess the QoL in patients with IBD[43]. There was previously no validated outcome measure that captured the patients' evaluation of the effect of interventions for Crohn's perianal fistulas on their wellbeing. This has importance in the face of elusive cure /non-sustained fistula closure, and is equally valuable as a secondary outcome measure for determining the impact of interventions in studies with curative intent, and as a primary outcome in studies of disease- or symptom-ameliorating interventions[44]. The CAF-QoL is the first disease specific PROM in Crohn's perianal fistula developed with a patient-centred methodology. Completion takes about 9minutes on average and can be done

with the clinician present or independently by the patient, with the goal of assessing the benefit of interventions in clinical trials, both medical and surgical, as well as defining disease impact and severity in the clinical setting to aid decision-making. It may also play a role in highlighting changing concerns and priorities that arise for patients between remission and relapse, which may guide individualised patient advice and support. Future research will include validation of its use in clinical trials, measuring the CAF-QoL scale scores and determining if scores correlate with fistula activity levels and what constitutes a clinically significant response.

The currently available PDAI was developed to measure clinical disease activity in patients with "perianal disease" and despite its use in clinical trials[41], it is not specific to perianal fistula. It was strongly considered for adjunctive use in this study, but we elected to go with a pragmatic recruitment strategy that obviated the need for clinical assessment of all patients (including those recruited via charities for questionnaire testing), which would have otherwise been required to ensure robust completion of the PDAI.

Future studies will utilise PDAI in addition to CAF-QOL (we hope) to enable comparison. There is also ongoing work to aiming to improve the objective measures available for assessing perianal fistula disease activity (e.g. using MRI[13]). In addition, a comprehensive classification for Crohn's perianal fistulas, which integrates all elements that are important for medical and surgical management is required and may need to incorporate a combination of diagnostics (endoscopy / MRI / endoanal ultrasound) and examination under anaesthesia to ensure robustness. Ongoing refinement and validation of CAF-QoL will ultimately aid the utility of this PROM alongside an objective clinical outcome measure. Both of which are necessary tools in the evaluation of perianal fistula. The CAF-QoL scale will also need

translation into different languages and cross- cultural validation to aid international dissemination.

References:

1   Ardizzone S, Porro GB. Perianal Crohn's disease: overview. *Dig Liver Dis* 2007;**39**:957–
     8. doi:10.1016/j.dld.2007.07.152

2   Economou M, Zambeli E, Michopoulos S. Incidence and prevalence of Crohn's disease
     and its etiological influences. *Ann Gastroenterol* 2009;**22**:158–67.

3   Thia KT, Mahadevan U, Feagan BG, *et al.* Ciprofloxacin or metronidazole for the
     treatment of perianal fistulas in patients with Crohn's disease: A randomized, double-
     blind, placebo-controlled pilot study. *Inflamm Bowel Dis* 2009;**15**:17–24.
     doi:10.1002/ibd.20608

4   Present DH, Korelitz BI, Wisch N, *et al.* Treatment of Crohn's disease with 6-
     mercaptopurine. A long-term, randomized, double-blind study. *N Engl J Med*
     1980;**302**:981–7. doi:10.1056/NEJM198005013021801

5   Present DH, Rutgeerts P, Targan S, *et al.* Infliximab for the treatment of fistulas in
     patients with Crohn's disease. *N Engl J Med* 1999;**340**:1398–405.
     doi:10.1056/NEJM199905063401804

6   Mak WY, Lung PFC, Hart A. In Patients With Perianal Crohn's Fistulas, What Are the
     Outcomes When 'Radiological Healing' Is Achieved? Does Radiological Healing of
     Perianal Crohn's Fistulas Herald the Time Point for Stopping a Biologic? *J Crohn's
     Colitis* 2017;**11**:1506–1506. doi:10.1093/ecco-jcc/jjx094

7   Mak WY, Adegbola S, Sahnan K, *et al.* P319 Long-term outcomes after anti-TNF
     therapy withdrawal in patients with radiologically healed perianal Crohn's fistulas. *J
     Crohn's Colitis* 2018;**12**:S262–S262. doi:10.1093/ecco-jcc/jjx180.446

8   Gecse KB, Bemelman W, Kamm MA, *et al.* A global consensus on the classification,
     diagnosis and multidisciplinary treatment of perianal fistulising Crohn's disease. *Gut*

2014;**63**:1381–92. doi:10.1136/gutjnl-2013-306709

9   Tozer P, Ng SC, Siddiqui MR, *et al.* Long-term MRI-guided combined anti-TNF-α and thiopurine therapy for Crohn's perianal fistulas. *Inflamm Bowel Dis* 2012;**18**:1825–34. doi:10.1002/ibd.21940

10   Irvine EJ. Usual therapy improves perianal Crohn's disease as measured by a new disease activity index. McMaster IBD Study Group. *J Clin Gastroenterol* 1995;**20**:27–32. doi:10.1007/BF02053756

11   Losco A, Viganò C, Conte D, *et al.* Assessing the activity of perianal Crohn's disease: comparison of clinical indices and computer-assisted anal ultrasound. *Inflamm Bowel Dis* 2009;**15**:742–9. doi:10.1002/ibd.20826

12   Horsthuis K, Ziech MLW, Bipat S, *et al.* Evaluation of an MRI-based score of disease activity in perianal fistulizing Crohn's disease. *Clin Imaging* 2011;**35**:360–5. doi:10.1016/j.clinimag.2010.09.003

13   Hindryckx P, Jairath V, Zou G, *et al.* Development and Validation of a Magnetic Resonance Index for Assessing Fistulas in Patients With Crohn's Disease. *Gastroenterology* 2019;**157**:1233-1244.e5. doi:10.1053/j.gastro.2019.07.027

14   Yassin N, Dibley L, Nachiappan S, *et al.* PTU-201 Living with anal fistulae. *Gut* 2015;**64**:A151.2-A151. doi:10.1136/gutjnl-2015-309861.316

15   Adegbola S, Dibley L, Sahnan K, *et al.* N002 A qualitative exploration into experiences of living with Crohn's anal fistula. *J Crohn's Colitis* 2018;**12**:S568–9. doi:10.1093/ecco-jcc/jjx180.1016

16   Ritchie J, Lewis J. *QUALITATIVE RESEARCH PRACTICE: A Guide for Social Science Students and Researchers*. London: : SAGE Publications 2003. https://mthoyibi.files.wordpress.com/2011/10/qualitative-research-practice_a-

guide-for-social-science-students-and-researchers_jane-ritchie-and-jane-lewis-eds_20031.pdf (accessed 23 Nov 2017).

17   Creswell J, Clark V, Plano Clark V. *Designing and Conducting Mixed Methods Research*. 3rd ed. Thousands Oaks, California: : SAGE Publications 2007. https://uk.sagepub.com/en-gb/eur/designing-and-conducting-mixed-methods-research/book258100

18   Rothrock NE, Kaiser KA, Cella D. Developing a Valid Patient-Reported Outcome Measure. *Clin Pharmacol Ther* 2011;**90**:737–42. doi:10.1038/clpt.2011.195

19   Brédart A, Marrel A, Abetz-Webb L, *et al.* Interviewing to develop Patient-Reported Outcome (PRO) measures for clinical research: eliciting patients' experience. *Health Qual Life Outcomes* 2014;**12**:1–10. doi:10.1186/1477-7525-12-15

20   Dibley L, Czuber-Dochan W, Woodward S, *et al.* Development and Psychometric Properties of the Inflammatory Bowel Disease Distress Scale (IBD-DS): A New Tool to Measure Disease-Specific Distress. *Inflamm Bowel Dis* 2018;**24**:2068–77. doi:10.1093/ibd/izy108

21   Ritchie J, Spencer L, O' Connor W. 'Carrying Out Qualitative Analysis' - Book Chapter. In: Ritchie J, Lewis J, eds. *Qualitative research practice: a guide for social science students and researchers*. London: : SAGE 2003. 219–62.http://www.library.mmu.ac.uk/secure/index.php?cat_file=&filename=y_21010010_ritchie_carrying.pdf

22   Sahnan K, Tozer PJ, Adegbola SO, *et al.* Developing a core outcome set for fistulising perianal Crohn's disease. *Gut* 2019;**68**:226–38. doi:10.1136/gutjnl-2017-315503

23   Cheung WY, Garratt AM, Russell IT, *et al.* The UK IBDQ-a British version of the inflammatory bowel disease questionnaire. development and validation. *J Clin*

Epidemiol 2000;**53**:297–306.http://www.ncbi.nlm.nih.gov/pubmed/10760641

(accessed 2 Mar 2019).

24   Régimbeau JM, Panis Y, Cazaban L, *et al.* Long-term results of faecal diversion for

refractory perianal Crohn's disease. *Color Dis* 2001;**3**:232–7. doi:10.1046/j.1463-

1318.2001.00250.x

25   Harvey RF, Bradshaw JM. A simple index of Crohn's-disease activity. *Lancet (London,*

England) 1980;**1**:514. doi:10.1016/S0140-6736(80)91372-0

26   Bennebroek Evertsz' F, Hoeks CCMQ, Nieuwkerk PT, *et al.* Development of the patient

Harvey Bradshaw index and a comparison with a clinician-based Harvey Bradshaw

index assessment of Crohn's disease activity. *J Clin Gastroenterol* 2013;**47**:850–6.

doi:10.1097/MCG.0b013e31828b2196

27   Marx RG, Bombardier C, Hogg-Johnson S, *et al.* Clinimetric and psychometric

strategies for development of a health measurement scale. *J Clin Epidemiol*

1999;**52**:105–11. doi:10.1016/s0895-4356(98)00148-6

28   Alrubaiy L, Cheung W-Y, Dodds P, *et al.* Development of a Short Questionnaire to

Assess the Quality of Life in Crohn's Disease and Ulcerative Colitis. *Orig Artic*

2015;**4**:66–76. doi:10.1093/ecco-jcc/jju005

29   Deyo RA, Inui TS, Wood R, *et al.* Articles Toward Clinical Applications of Health Status

Measures: Sensitivity of Scales to Clinically Important Changes.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1068815/pdf/hsresearch00516-

0006.pdf (accessed 7 Aug 2019).

30   Bindman AB, Keane D, Lurie N. Measuring health changes among severely ill patients.

The floor phenomenon. *Med Care* 1990;**28**:1142–

52.http://www.ncbi.nlm.nih.gov/pubmed/2250498 (accessed 7 Aug 2019).

31   Sim J, Wright CC. The Kappa Statistic in Reliability Studies: Use, Interpretation, and

        Sample Size Requirements. *Phys Ther* 2005;**85**:257–68. doi:10.1093/ptj/85.3.257

32   Vanbelle S. A New Interpretation of the Weighted Kappa Coefficients. *Psychometrika*

        2016;**81**:399–410. doi:10.1007/s11336-014-9439-4

33   Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic.

        Fam Med 2005;**37**:360–3.http://www.ncbi.nlm.nih.gov/pubmed/15883903

34   Dibley L, Norton C, Cotterill N, *et al.* Development and initial validation of a disease-

        specific bowel continence questionnaire for inflammatory bowel disease patients:

        The ICIQ-IBD. *Eur J Gastroenterol Hepatol* 2016;**28**:233–9.

        doi:10.1097/MEG.0000000000000513

35   Marx RG, Bombardier C, Hogg-Johnson S, *et al.* Clinimetric and psychometric

        strategies for development of a health measurement scale. *J Clin Epidemiol* Published

        Online First: 1999. doi:10.1016/S0895-4356(98)00148-6

36   Adegbola S, Dibley L, Sahnan K, *et al.* N001 Developing a quality of life score for

        Crohn's anal fistula (CAF-QoL). *J Crohn's Colitis* 2018;**12**:S568–S568.

        doi:10.1093/ecco-jcc/jjx180.1015

37   Sailer M, Bussen D, Debus ES, *et al.* Quality of life in patients with benign anorectal

        disorders. *Br J Surg* 1998;**85**:1716–9. doi:10.1046/j.1365-2168.1998.00958.x

38   García-Aguilar J, Davey CS, Le CT, *et al.* Patient satisfaction after surgical treatment

        for fistula-in-ano. *Dis Colon Rectum* 2000;**43**:1206–12. doi:10.1007/bf02237422

39   Wong S, Solomon M, Crowe P, *et al.* Cure, continence and quality of life after

        treatment for fistula-in-ano. *ANZ J Surg* 2008;**78**:675–82. doi:10.1111/j.1445-

        2197.2008.04616.x

40   Ferrer-Márquez M, Espínola-Cortés N, Reina-Duarte A, *et al.* Design and psychometric

evaluation of the quality of life in patients with anal fistula questionnaire. *Dis Colon Rectum* 2017;**60**:1083–91. doi:10.1097/DCR.0000000000000877

41   Gecse KB, Sebastian S, Hertogh G de, *et al.* Results of the Fifth Scientific Workshop of the ECCO [II]: Clinical Aspects of Perianal Fistulising Crohn's Disease—the Unmet Needs: Table 1. *J Crohn's Colitis* 2016;:jjw039. doi:10.1093/ecco-jcc/jjw039

42   Wiering B, de Boer D, Delnoij D. Patient involvement in the development of patient-reported outcome measures: a scoping review. *Heal Expect* 2017;**20**:11–23. doi:10.1111/hex.12442

43   Cheung W-Y, Garratt AM, Russell IT, *et al.* The UK IBDQ-A British version of the inflammatory bowel disease questionnaire: development and validation. 2000. https://ac.els-cdn.com/S0895435699001523/1-s2.0-S0895435699001523-main.pdf?_tid=e9bd3e43-d4cc-4202-83c2-04c451c37db4&acdnat=1551509339_da2d5d8571f3b9477179099025d57623 (accessed 2 Mar 2019).

44   Adegbola SO, Sahnan K, Tozer PJ, *et al.* Symptom amelioration in Crohn's perianal fistulas using video assisted anal fistula treatment (VAAFT). *J Crohn's Colitis* 2018;:1–6. doi:10.1093/ecco-jcc/jjy071