Health-related Quality of Life in Patients with Inflammatory Bowel Disease Measured with the Short Form-36: Psychometric Assessments and a Comparison with General Population Norms

Tomm Bernklev, BSc,† Jørgen Jahnsen, MD, PhD,‡ Idar Lygren, MD, PhD,§ Magne Henriksen, MD,‖ Morten Vatn, MD, PhD,*¶ and Bjørn Moum, MD, PhD†*

Background: We compared health-related quality of life (HRQOL) in a population-based cohort of Norwegian patients with inflammatory bowel disease (IBD) with a normal reference population by means of the short form-36 (SF-36) questionnaire, including the effect of age, sex, educational status, and symptom severity and the psychometric properties of the questionnaire.

Methods: The SF-36 was self-administered and was answered by the patients at the hospital at 2 occasions that were 6 months apart.

Results: Five hundred fourteen patients with IBD were eligible for analysis: 348 with ulcerative colitis (UC) and 166 with Crohn's disease (CD). The comparison group consisted of 2323 Norwegian people. The dimension scores for SF-36 were significantly lower in 6 of 8 dimensions for patients with UC and in 7 of 8 dimensions for patients with CD than for the reference population. In both patients with UC and patients with CD, we found lower scores in elderly patients, which also was found in the background population. Women scored lower than men in all dimension scores. In both patients with UC and patients with CD, there was a statistically significant reduction in HRQOL score with increasing symptoms. The SF-36 has satisfactory reliability and discriminant ability for scores for all dimensions in both patients with UC and patients with CD. However, when measuring responsiveness, the figures were generally low. This finding, together with the high ceiling effects, may indicate that the SF-36 has limitations regarding detecting deterioration or improvement over time.

Conclusion: We have shown that HRQOL in a Norwegian population-based cohort of patients with IBD, measured with the SF-36, is lower than that of a Norwegian reference population. In general, the SF-36 was found to have satisfactory psychometric properties in this IBD population.

Key Words: Crohn's disease, health-related quality of life, inflammatory bowel disease, short form-36, ulcerative colitis

(*Inflamm Bowel Dis* 2005;11:909–918)

Ulcerative colitis (UC) and Crohn's disease (CD) are chronic inflammatory bowel diseases (IBDs) that affect patient physical, mental, and social well-being, often referred to as the patient's health-related quality of life (HRQOL).[1–5]

HRQOL is now frequently used as a primary or secondary outcome measure in epidemiological studies and clinical trials. From a clinical point of view, it is important that these results can be compared across studies and also between different countries. This means that the questionnaires must be translated and validated in the local language.

Instruments that measure HRQOL can be divided into generic or disease-specific questionnaires. Generic instruments are nonspecific to diseases or populations, and this makes it easier to compare different diseases and conditions and to relate the findings to a background population.[6,7] Several generic instruments are available, such as the Sickness Impact Profile[8] and the Nottingham Health Profile.[9]

The Short Form-36 (SF-36) survey is the most widely used health survey and has been used in many patient groups and normal populations. It has been used by several investigators to measure HRQOL in patients with IBD, either alone or in combination with the Inflammatory Bowel Disease Questionnaire (IBDQ). However, most of these studies have been performed with selected patient groups, either as part of a drug comparison trial or in patients with IBD scheduled for surgical intervention.[10,11] Few studies have been performed with unselected patients with IBD followed prospectively over

Received for publication June 7, 2005; accepted July 15, 2005.

From the *Rikshospitalet University Hospital, Oslo, Norway; †Østfold Central Hospital, Fredrikstad, Norway; ‡Aker University Hospital, Oslo, Norway; §Ullevål University Hospital, Oslo Norway; ‖Østfold Central Hospital, Moss, Norway; ¶EPIGEN, A-HUS, University of Oslo and the Inflammatory Bowel South-Eastern Norway (IBSEN) Group of Gastroenterologists, Oslo, Norway.

Supported by a research fellowship grant from Helse Øst, Sykehuset Østfold, Fredrikstad.

Reprints: Tomm Bernklev, Solløkkasletta 10, 3233 Sandefjord, Norway (e-mail: tomm.bernklev@bluezone.no)

time, and at present, no published data are available regarding the use of the SF-36 in Norwegian patients with IBD. The SF-36 has been translated, validated, and tested cross-culturally among Norwegians.[12]

The purpose of this study was to compare, for the first time, HRQOL in a population-based cohort of Norwegian patients with IBD with a normal reference population by means of the SF-36 questionnaire, including the effect of age, sex, and educational status. We also studied whether there were differences in HRQOL between symptom-free patients with IBD and those with moderate to severe symptoms. Finally, we evaluated the psychometric properties of the SF-36 in this population-based cohort of patients with IBD.

MATERIALS AND METHODS

Study Population

All newly diagnosed cases of IBD or possible IBD between 1 January 1990 and 31 December 1993 in 4 well-defined areas in southeastern Norway (the counties of Oslo, Østfold, Telemark, and Aust-Agder) were registered at their local departments of gastroenterology. To ensure complete ascertainment, all 1236 general practitioners and clinicians at the 14 hospitals in the 4 participating counties received information about symptoms consistent with IBD and were invited to refer all potential cases of UC and CD to the local gastroenterological outpatient clinic. The information was given before the start of the study in the form of 3 written reminders. In addition, information about the study was presented at meetings with the general practitioners, and all the practicing gastroenterologists, internists, surgeons, and pediatricians in the relevant area were informed of the study. The total population of the geographical area was 966,427 on 1 January 1992.

Endoscopy was chosen as the main instrument for diagnosis and for determining the distribution of the disease in the colon. A total of 618 cases of UC and indeterminate colitis and 225 cases of CD were diagnosed.[13,14] At follow-up between 1 and 2 years later, the diagnosis was re-evaluated. Ninety-eight percent of the patients were available for follow-up. Four percent were excluded at this point because the diagnosis of IBD could not be definitely confirmed.[15] Five years after the initial diagnosis, the patients were invited to make a clinical follow-up visit to their local hospitals. Each case was reviewed by a gastroenterologist, and patients without a confirmed diagnosis of UC or CD or who had died were excluded from the analysis in this study. Separate records were made for nonparticipants and deaths.[16]

At the 5-year follow-up, 654 patients fulfilled the criteria for maintaining the diagnosis of IBD (454 with UC and 200 with CD). Of these, 533 gave written informed consent and completed the Norwegian version of the SF-36. Nine patients were excluded because they were less than 18 years of age, 3 were above 79 years of age, and an additional 7 patients were excluded because of incomplete answers to the questionnaires. Thus, the SF-36 data from 514 patients (348 UC and 166 CD) were available for analysis. The demographic data are shown in Table 1.

Information on the 121 patients who did not attend the follow-up visit (29 had moved away from the area, 21 were suffering from concomitant serious disease or old age, and 71 did not respond to the invitation) was based on hospital records and telephone interviews.

General Population Sample

SF-36 data for the general population were collected from the study by Loge et al.[17] Their material was based on a sample of 3500 Norwegian citizens 18 to 80 years of age randomly drawn from the Norwegian National Population Register by the Norwegian Government Data Center (SDS), which was representative of the entire Norwegian population. Of the 3500 questionnaires that were sent out, 2323 were returned, and it is the data from this population, 18 to 80 years of age, that were used for comparison with our patients with IBD.

Diagnostic Criteria and Classification of Intestinal Disease

Uniform methods and diagnostic criteria were used. The initial diagnosis of UC and CD was based on symptoms

TABLE 1. Demographic Data Compared with Background Population

	UC	CD	IBSEN	Normal Population
Number	348	166	514	2323
Age, mean (SD)	45.0 (14.7)	38.6 (15.2)	42.9 (15.1)	44.9 (16.5)
Age range (years)	18–80	18–78	18–80	19–80
Age groups (years)				
≤29	16%	34%	22%	22%
30–39	27%	31%	28%	21%
40–49	25%	12%	21%	19%
50–59	13%	9%	12%	16%
60–69	12%	8%	10%	12%
≥70	7%	6%	7%	10%
Sex (%)				
Female	48%	51%	49%	51%
Male	52%	49%	51%	49%
Education				
Second level, first stage	33%	32%	33%	27%
Second level, second stage	28%	31%	29%	45%
Third level (university)	39%	36%	38%	28%

consistent with IBD for more than 4 weeks, excluding infections and other acute or chronic non-IBDs.

The diagnosis of UC[13] was based on the presence of at least 3 of the following criteria: (1) a history of diarrhea and/or blood/pus in the stools; (2) macroscopic appearance at endoscopy of continuous mucosal inflammation affecting the rectum in continuity with some or all of the colon; (3) microscopic features at biopsy compatible with UC; and (4) no suspicion of CD at small bowel X-ray, ileoscopy, or biopsy.

The diagnosis of CD was based on the presence of 2 or more of the following established criteria[18]: (1) typical clinical features including abdominal pain, diarrhea, and weight loss; (2) macroscopic appearance at surgery or endoscopy: segmental, discontinuous, and/or patchy lesions with or without rectal involvement, discrete or aphthous ulcerations, fissuring and penetrating lesions, cobblestone or strictures; (3) radiologic evidence of stenosis in the small bowel, segmental colitis, or findings of fistulae; and (4) histologic evidence of transmural inflammation or epithelial granulomas with giant cells.

The extent and localization of colonic disease were based on endoscopic findings and characteristic histologic signs of inflammation. When the extent of the disease changed during follow-up, the maximum extent of bowel involvement was recorded.

Data Collection

All patient data were collected between January 1995 and December 1999. The SF-36 was self-administered and was answered by the patients at the hospital before they were interviewed and clinically examined by the gastroenterologist (visit 1). A standardized procedure was followed at all the centers, which allowed the patients to fill in the questionnaire alone in peace and quiet. Before the patients left the clinic, the investigator or the nurse checked the questionnaire to ensure that all the questions had been answered. The SF-36 was readministered to all patients 6 months later (visit 2) in connection with a rheumatological examination by a rheumatologist.[19]

Assessment of HRQOL

Short form-36

The SF-36 is a generic HRQOL questionnaire designed to assess functional status, well being, and general perception of health. The questionnaire consists of 8 subscales: physical functioning (10 items), bodily pain (2 items), vitality or energy level (4 items), social functioning (2 items), mental health (5 items), general health perceptions (5 items), role limitation because of physical problems (4 items), and role limitation because of personal or emotional problems (3 items). An additional item (HT) measures health transition over the past year. A physical health and a mental health summary score can be computed as well.[20] The questionnaire has been translated into Norwegian and validated.[21,22]

Internal consistency reliability

Internal consistency reliability was calculated by means of the Cronbach α coefficient. This is a measure of how strongly the items within a scale correlate with each other and therefore is indicative of the extent to which the items reflect the same underlying phenomenon. A value of 0.70 was set as the cut-off point.

Discriminant ability

The discriminant ability of an instrument is the ability to distinguish between groups that are of interest. In this study, this ability was tested by dividing the patients into groups according to their answers regarding their overall IBD-related clinical condition at the 5-year follow-up visit. The possible answers were none, mild, moderate, or severe symptoms.

Reliability

Reliability was tested by measuring differences between patients at a particular point in time, expressed statistically as an intraclass correlation coefficient that relates the between-subject variance to the total variance.[23] The SF-36 was administered on 2 occasions: visit 1 and visit 2. In this study, reliability was only tested for those patients who reported unchanged symptoms between the 2 visits.

Sensitivity to change

The instrument must be able to detect changes in the health state of an individual patient over time.[24] In this study, sensitivity to change was tested by a paired *t* test in those patients who reported deterioration or improvement in IBD symptoms at visit 2 compared with visit 1. The calculation was made for both the UC and CD groups together by reversing the score for those who reported deterioration at visit 2.

We also evaluated responsiveness by calculating the Guyatt statistic.[25] This is the ratio of the change score in those with improvement or deterioration to the SD of the change score in the stable group. The Guyatt statistic detects changes beyond the underlying variability in change scores as represented by those patients with no clinically defined change in health status. We calculated the Guyatt statistic according to patient response to a single item and patient overall evaluation of their IBD at visit 2 as better, unchanged, or worse compared with visit 1.

Clinical Condition Questionnaire

The patients were asked about their clinical condition at visit 1. The recall period was the previous 2 weeks. The 4 possible answers were: no symptoms, mild symptoms (do not interfere with everyday activities), moderate symptoms (do interfere with everyday activities, may result in sick leave), and severe symptoms (unable to carry out everyday activities, on sick leave or hospitalized).

IBD Status Questionnaire

At visit 2, all patients were asked to answer a transition question regarding their IBD status: "How would you generally classify your IBD this time compared to the last time you completed this questionnaire—improved, unchanged or worse?" The recall period was the previous 6 months.

Demographic and Clinical Data

Demographic data and a detailed symptom and treatment history, including all disease-relevant aspects since the initial diagnosis 5 years previously, were obtained at the interview with the gastroenterologist.

Statistical Methods

All descriptive data are given as means with SD; range is given where appropriate. The η^2 test (adjusted) and the Wilks λ test were used to evaluate the discriminatory power of the SF-36 questionnaire. A χ^2 test was used to compare HRQOL scores between the different symptom groups. Comparisons of quality of life between patients with UC, patients with CD, and the reference population were performed by means of a 2-sided *t* test for independent samples. Means were calculated as marginal mean scores adjusted for age, sex, and educational level.

The scale scores were also standardized by computing mean scores (analysis of covariance) adjusted for age, sex, and educational level. A deviation score was calculated by subtracting the scores for patients with UC and patients with CD from the scores of controls; finally, the deviations were divided by the SD of each scale in the total sample.

The significance level was set at 5%; all tests were 2-sided. To adjust for multiple comparisons, we used the Bonferroni method. All statistical analyses were performed with SPSS Version 12.01 (SPSS, Chicago, Ill.) for Windows.

Ethical Considerations

The study was performed in accordance with the principles set out in the Helsinki Declaration and was approved by the regional ethics committee. Permission was also obtained from the Norwegian Data Inspectorate.

RESULTS

Demographic and Clinical Characteristics

There were no significant differences in sex between the patients with UC and patients with CD. However, the mean age in the CD group was significantly lower than in the UC group: 38.6 years (range, 36.3–40.9 yr) versus 45.0 years (range, 43.9–46.1 yr). The mean age in the patient group as a whole (42.9 yr) was somewhat lower than in the reference population (44.9 yr), where there were more subjects in the elderly age cohorts. Sex distribution was similar in the 2 groups, but more people in the patient group had tertiary education than in the reference population[26] (Table 1). The mean time since diagnosis was similar in the 2 diagnostic groups: 5.37 ± 0.74 years for patients with UC and 5.34 ± 0.71 years for patients with CD.

Scale Analysis

Seven patients (1.3%) were withdrawn because of incomplete responses to the questionnaire. The overall percentage of missing data was 1.4%, and no particular question had a higher noncompletion rate than any other.

Mean dimensional scores (SD), ceiling and floor effects, and the internal consistencies for all scales, presented separately for the 2 diagnostic groups, are shown in Table 2. Using the 15% recommended critical value for the largest proportion of the study population that should score at the highest or lowest possible scale level,[27] we observed substantial ceiling effects for the dimensions physical functioning, role-physical, bodily pain, social functioning, and role-emotional in both diagnostic groups. Relevant floor effects were observed for role-physical (both UC and CD) and role-emotional (CD).

Internal Consistency Reliability

Internal consistency estimates for the scales (Cronbach α) ranged from 0.72 to 0.90 (CD) and 0.74 to 0.91 (UC), and exceeded the 0.70 standard for all the scales (Table 2). The recommended minimum of 0.90 for individual comparisons

TABLE 2. Mean Dimensional Scores (SD), Distribution and Internal Reliability

	Mean Score ± SD		Floor (%)		Ceiling (%)		Cronbach α	
SF-36 Scales	UC	CD	UC	CD	UC	CD	UC	CD
Physical functioning	88.0 ± 18.4	85.9 ± 16.7	0	0	40.4	32.3	0.91	0.85
Role-physical	71.3 ± 39.0	62.9 ± 41.1	16.3	22.8	58.2	46.1	0.89	0.88
Bodily pain	70.6 ± 24.4	64.7 ± 28.1	0.3	1.8	28.4	24.0	0.88	0.90
General health	66.5 ± 22.9	61.1 ± 25.3	0.3	0.6	2.9	4.2	0.81	0.83
Vitality	56.1 ± 22.0	52.9 ± 23.6	0.9	0.6	2.3	1.8	0.86	0.86
Social functioning	82.7 ± 22.5	76.6 ± 23.3	1.1	0	49.0	34.1	0.81	0.80
Role-emotional	73.6 ± 35.5	67.7 ± 37.4	11.5	15.0	57.9	49.1	0.74	0.72
Mental health	77.2 ± 16.7	74.9 ± 17.2	0.3	0.6	7.7	7.8	0.81	0.84

was met by the physical functioning scale (in UC) and the bodily pain scale (in CD).

Discriminant Ability

In answer to the question regarding clinical condition at visit 1, 194 patients with UC reported no symptoms, 125 patients reported mild symptoms, and 29 patients reported moderate to severe symptoms. Similar figures for patients with CD were 76, 67, and 23, respectively. The discriminatory power of the SF-36 was evaluated by the η^2 test and the Wilks λ test, and the pairwise comparisons between symptom groups were found to be significant by the *F* test (data not shown). Thus, in this setting, the SF-36 showed good discriminant ability.

Test–Retest Reliability

A total of 281 patients who reported no change in their disease (UC = 193 and CD = 88) completed the IBDQ for both visit 1 and visit 2 and could be used in the reliability test. Table 3 shows the reliability outcomes for each diagnosis separately. Intraclass correlation ranged from 0.56 to 0.89 (CD) and from 0.63 to 0.92 (UC). None of the differences between the baseline and follow-up values in patients with CD were statistically significant. For patients with UC, the differences for physical functioning and bodily pain were statistically significant, whereas those for the other 6 dimensions were not significantly different.

Sensitivity to Change

The sensitivity to change of the SF-36 was examined by means of the transition question at visit 2. One hundred fourteen patients with UC and 84 patients with CD reported worsening or improvement of their disease at visit 2, 6 months after visit 1. The paired *t* test showed a statistically significant difference between visit 1 and visit 2 in all dimension scores for both diagnostic groups, with the exception of the physical functioning dimension in patients with UC (Table 4).

Responsiveness

A Guyatt statistic greater than 1.00 (or –1.00) for changed groups was considered indicative of a measure that was highly responsive to change, whereas a value greater than 0.20 (or –0.20) was considered acceptable. The stable group was expected to have values close to 0. The results are presented in Table 5. Most of the figures for stable patients with UC and patients with CD were close to 0, but for physical functioning (in UC) and role-physical (in CD), the figures were higher than expected. In those reporting improvement or deterioration, the figures were lower than 0.20 for role-physical (in UC) and for physical functioning and role-physical (in CD).

TABLE 3. ICC Between the SF-36 Dimensional Scores in Patients Who Reported No Change in Disease Activity Between 2 Consultations with an Interval of 6 Months

	Mean Score at Time 1	Mean Score at Time 2	Mean Difference (95% CI)	ICC
UC (n = 193)				
PF	88.3	85.5	2.7 (1.2–4.3)	0.92
RP	73.6	73.5	0.1 (−6.7–7.0)	0.64
BP	73.6	70.5	3.1 (0.1–6.1)	0.75
GH	67.1	65.6	1.5 (−0.1–3.9)	0.83
VT	57.1	57.2	0.1 (−2.3–2.1)	0.86
SF	84.7	84.8	0.1 (−2.9–2.7)	0.74
RE	78.4	78.2	0.2 (−4.6–5.1)	0.63
MH	78.6	79.2	0.6 (−2.3–1.0)	0.82
CD (n = 88)				
PF	84.8	82.8	2.0 (−1.3–5.1)	0.84
RP	64.8	71.9	7.1 (−14.8–0.6)	0.72
BP	66.8	67.0	0.2 (−5.3–4.9)	0.74
GH	62.3	59.7	2.4 (−0.8–5.9)	0.89
VT	56.1	53.6	2.5 (−1.3–6.3)	0.80
SF	79.7	82.4	2.7 (−6.7–1.3)	0.78
RE	70.8	76.5	5.7 (−14.0–2.6)	0.56
MH	76.7	78.5	1.8 (−4.4–0.9)	0.81

PF, physical functioning; VT, vitality; RP, role-physical; SF, social functioning; BP, bodily pain; RE, role-emotional; GH, general health; MH, mental health.

TABLE 4. Comparison of Baseline Scores and Scores After 6 Months in Unstable Patients Reporting Deterioration or Improvement in Their IBD by Diagnostic Group

	Baseline Means	Follow-up Means	Mean Difference (95% CI)	*P*
UC (n = 114)				
PF	82.5	85.6	3.1 (−1.1–6.3)	0.059
RP	52.6	60.5	7.9 (0.25–15.5)	0.043
BP	56.4	64.1	7.7 (3.8–11.8)	<0.001
GH	55.9	60.6	4.7 (1.2–8.1)	0.008
VT	46.7	54.1	7.4 (4.1–10.7)	<0.001
SF	70.4	80.2	9.8 (5.7–13.8)	<0.001
RE	57.6	70.2	12.6 (5.6–19.6)	0.001
MH	70.9	76.4	5.4 (2.2–8.6)	0.001
CD (n = 84)				
PF	80.7	86.1	5.4 (2.1–87)	0.001
RP	46.1	55.7	9.5 (0.1–18.9)	0.047
BP	51.6	60.1	8.5 (2.6–14.4)	0.005
GH	51.7	58.6	6.9 (2.7–11.2)	0.002
VT	41.8	48.9	7.0 (3.2–10.9)	<0.001
SF	68.0	77.5	9.5 (5.0–14.0)	<0.001
RE	52.4	66.7	14.3 (5.6–22.9)	0.001
MH	68.3	74.3	6.0 (2.3–9.7)	0.002

PF, physical functioning; VT, vitality; RP, role-physical; SF, social functioning; BP, bodily pain; RE, role-emotional; GH, general health; MH, mental health.

None of the figures reached a score of 1 or above 1; in general, the scores were low (>0.50).

Symptoms and SF-36 Score

In both patients with UC and patients with CD, there was a statistically significant reduction in HRQOL score, with increasing symptoms in all 8 dimensions, calculated by χ^2. The score differences in those reporting moderate to severe symptoms compared with those without symptoms were most pronounced for role-physical, role-emotional, bodily pain, and general health. Physical functioning and mental health seemed to be the least affected by increasing IBD symptoms (Table 6). When the patients with IBD were stratified according to the presence or absence of symptoms, those who were symptom-free had SF-36 scores close to those of the normal reference population. However, patients with IBD with moderate and severe symptoms had significantly lower scores than the reference population, irrespective of sex or diagnosis.

Age and Sex

In patients with UC, we found lower scores in the elderly patients (60 years and older, data not shown) for the physical dimensions, physical functioning, role-physical, bodily pain, and general health, and for the mental dimension, role-emotional. In patients with CD 50 years and older, the findings were similar for physical functioning, role-physical, bodily pain, and role-emotional. Data from the Norwegian reference population showed a similar pattern, so this finding is not

TABLE 5. The Guyatt Statistic for Responsiveness for Patients Who Had Improved or Deteriorated and Those Whose IBD was Unchanged According to Their Own Evaluation at Visit 2

		PF	RP	BP	GH	VT	SF	RE	MH
UC	Improved/deteriorated	−0.25	−0.18	−0.38	−0.32	−0.50	−0.57	−0.41	−0.46
	Unchanged	0.26	0.003	0.15	0.009	−0.01	0.003	0.001	−0.05
CD	Improved/deteriorated	−0.18	−0.14	−0.35	−0.27	−0.37	−0.37	−0.35	−0.63
	Unchanged	0.13	−0.20	−0.01	0.16	0.14	−0.14	−0.14	−0.14

To increase the power of the analysis, patients reporting improvement or deterioration were grouped together by reversing the sign of the difference in score of patients reporting improvement.

PF, physical functioning; VT, vitality; RP, role-physical; SF, social functioning; BP, bodily pain; RE, role-emotional; GH, general health; MH, mental health.

TABLE 6. Patients' Dimensional Scores* with 95% CI in Each IBD Symptom Group,† Presented Separately for Patients with UC and Patients with CD

	Group 1 (n = 194)	Group 2 (n = 125)	Group 3 (n = 29)	*P*‡
UC (n = 348)				
PF	89.94 (87.57–92.30)	87.78 (84.87–90.69)	75.43 (68.84–82.02)	<0.001
RP	82.17 (77.26–87.08)	64.92 (58.87–70.97)	25.66 (11.96–39.35)	<0.001
BP	77.53 (74.35–80.71)	64.12 (60.20–68.04)	49.63 (40.75–58.50)	<0.001
GH	71.87 (68.74–74.99)	61.55 (57.70–65.40)	49.92 (41.20–58.63)	<0.001
VT	60.13 (57.16–63.11)	52.84 (49.18–56.51)	42.59 (34.29–50.88)	<0.001
SF	87.48 (84.46–90.49)	79.62 (75.91–83.33)	64.73 (56.33–73.14)	<0.001
RE	79.49 (74.64–84.33)	70.16 (64.19–76.12)	50.87 (37.36–64.38)	<0.001
MH	79.48 (77.16–81.81)	76.02 (73.16–78.88)	67.27 (60.79–73.75)	0.001
	Group 1 (n = 76)	**Group 2 (n = 67)**	**Group 3 (n = 23)**	***P*‡**
CD (n = 166)				
PF	86.48 (83.01–89.88)	88.32 (84.70–91.95)	76.68 (70.49–82.86)	0.006
RP	70.16 (61.48–78.84)	66.86 (57.61–76.11)	27.71 (11.93–43.49)	<0.001
BP	73.30 (67.67–78.93)	64.19 (58.19–70.19)	34.98 (24.75–45.22)	<0.001
GH	69.49 (64.19–74.78)	59.18 (53.54–64.83)	39.60 (29.97–49.22)	<0.001
VT	58.00 (53.04–60.34)	55.05 (49.75–60.34)	31.70 (22.67–40.73)	<0.001
SF	82.31 (77.64–86.98)	80.16 (75.18–85.14)	48.84 (40.35–57.33)	<0.001
RE	73.96 (66.02–81.91)	69.74 (61.28–78.21)	41.94 (27.50–56.38)	0.001
MH	78.00 (74.44–81.56)	77.61 (73.81–81.40)	58.17 (51.69–64.64)	<0.001

PF, physical functioning; VT, vitality; RP, role-physical; SF, social functioning; BP, bodily pain; RE, role-emotional; GH, general health; MH, mental health.
*Based on estimated marginal means corrected for age and sex.
†Symptom score: 1, no symptoms; 2, mild symptoms; 3, moderate/severe symptoms.
‡χ^2, adjustment for multiple comparisons: Bonferroni.

specific to the IBD study population. Women scored lower than men in all dimension scores. In patients with UC, this difference was significant in 5 of 8 dimensions, whereas in patients with CD, the differences were statistically significant in all dimensions (Fig. 1).

Comparison with the Background Population

The dimension scores for SF-36 were significantly lower in 6 of 8 dimensions for patients with UC and in 7 of 8 dimensions for patients with CD than for the reference population. These differences were most pronounced in the dimensions role-physical, role-emotional, and general health for both diagnostic groups (Table 7). With regard to the standardized scores, we found that the most seriously affected dimension was general health, followed by the 2 role dimensions, physical and emotional (Fig. 2).

DISCUSSION

In this study, we found lower SF-36 scores for almost all dimensions in an unselected, population-based cohort of Norwegian patients with IBD compared with general population norms. For both patients with UC and patients with CD, the most pronounced differences were found in the dimensions general health, role-physical, and role-emotional.

The general health dimension assesses whether a person considers themselves as being more prone to illness than

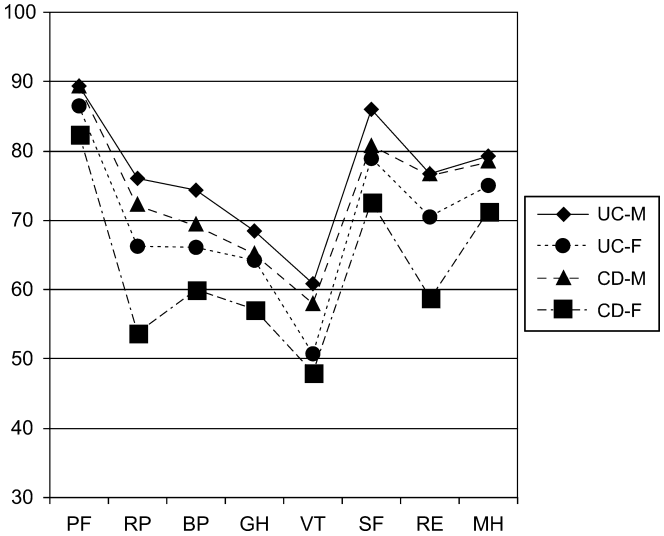

FIGURE 1. SF-36 dimensional scores according to sex and diagnostic group.

TABLE 7. SF-36 Scores for Patients with UC, Patients with CD, and the Reference Population, Presented as Crude Mean Scores (SD) and Scores Adjusted for Age, Sex, and Educational Status

	UC		CD		Reference Population			
Scales	Crude (SD)	Adjusted	Crude (SD)	Adjusted	Crude (SD)	Adjusted	UC Versus Reference (*P*)	CD Versus Reference (*P*)
Physical functioning	88.0 (18.4)	88.6	85.9 (16.7)	86.5	88.4 (17.4)	88.3	NS	NS
Role-physical	71.3 (39.0)	72.6	62.9 (41.1)	62.6	79.7 (34.4)	79.5	<0.001	<0.001
Bodily pain	70.6 (24.4)	71.3	64.7 (28.1)	64.0	75.7 (25.6)	75.6	<0.001	<0.001
General health	66.5 (22.9)	67.1	61.1 (25.3)	59.7	77.4 (21.8)	77.2	<0.001	<0.001
Vitality	56.1 (22.0)	56.3	52.9 (23.6)	52.7	60.3 (20.7)	60.3	<0.001	<0.001
Social functioning	82.7 (22.5)	82.6	76.6 (23.3)	76.4	86.1 (21.8)	86.1	0.007	<0.001
Role-emotional	73.6 (35.5)	73.8	67.7 (37.4)	68.2	82.5 (31.8)	82.5	<0.001	<0.001
Mental health	77.2 (16.7)	77.2	74.9 (17.2)	75.1	78.8 (16.4)	78.7	NS	0.003

NS, not significant.

others and whether they expect their health to get worse. A low score on this item often reflects a patient's knowledge that they have a chronic disease, which could relapse at any time.

The role-physical dimension measures problems with work or other daily activities as a result of reduced physical health, and the role-emotional dimension measures the same type of problems as a result of emotional impairment. It is obvious that patients with a chronic disease such as IBD have more problems with work or daily activities than the reference population. Correspondingly, they may have more serious emotional problems because of anxiety and depression[5] and problems with social or everyday activities. Lower scores for physical and mental role dimensions have also been observed in patients with other chronic conditions.[28,29]

In patients with CD, the scores for the bodily pain and social function dimensions were significantly lower than those for patients with UC, indicating that patients with CD have more pain than patients with UC and that it interferes with their ability to work and their social life.

Standardization makes the scores comparable, and this allows us to assess which aspect of HRQOL is most affected. The largest numerical difference between patients with UC and patients with CD on the one hand and the reference population on the other was found in the dimension general health, and this was confirmed by the standardized scores. Because the general health dimension measures a patient's own evaluation of their general health, it is perhaps 1 of the most important indicators of the influence of IBD on a patient's HRQOL. A significantly low score for this dimension is thus a reflection of the fact that a chronic condition such as IBD significantly affects a patient's daily life.

HRQOL scores decreased with increasing age in both patients with UC and patients with CD, and this trend was most pronounced in the physical dimension scores. The reference population showed the same trend. These scores reflect decreasing physical activity in elderly persons and are not in themselves specific to the IBD population.

We also found a sex difference, with women scoring lower than men. This sex effect has been reported by several others,[30–32] and a number of explanations have been put forward to explain it. Psychologic factors may play a greater role in some women than in men[2,33]; women often have more serious disease-related worries and concerns than men.[3,33–35] Another possibility is that more extensive disease is more common in women.[36] However, the sex effect is also present in the general population,[37,38] making it likely that there are other explanations.

We found a relation between the severity of IBD symptoms and HRQOL scores. When the mean scores were transformed to standardized scores (data not shown), the general health, bodily pain, and role-physical dimensions were the most affected in both patients with UC and patients with CD

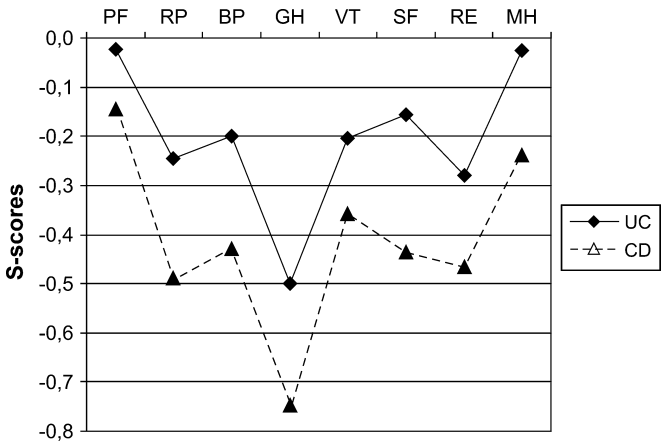

FIGURE 2. Standardized scores for patients with UC and patients with CD, adjusted for age, sex, and educational level (0 = reference population).

with moderate and severe symptoms, and in these patients with CD, social function and role-emotional were also strongly affected. In a previous publication,[39] we found a similar relation between symptom score and HRQOL using the disease-specific questionnaire N-IBDQ. This relation therefore seems to be independent of the type of questionnaire used, whether it is generic, like the SF-36, or disease-specific, like the IBDQ.

McColl et al[40] reported that patients with UC had lower scores for vitality, social functioning, role-physical, and general health than the population norms for the United Kingdom. We did not find a similar reduction in vitality and social functioning in our patients, which may be because of differences in the selection of patient populations in the 2 studies.

Like us, Hjortswang et al[31] found a relation between SF-36 scores and severity of symptoms in Swedish patients with UC. They reported that patients in remission had scores similar to a Swedish reference population, whereas patients with a relapsing disease had lower scores for almost all dimensions, and most markedly in general health and social functioning. This supports our finding that patients in clinical remission have the same HRQOL scores as the background population.

Analysis of the distribution of the scores revealed substantial floor effects for the dimensions role-physical and role-emotional. These 2 dimensions also had the lowest reliability estimates as measured by the Intraclass coefficient (ICC) and Cronbach α, which may indicate that these scales are not sensitive enough to detect small but important changes in specific patient groups.[41,42] This is in line with findings from other patient groups with chronic conditions, and the results for these 2 dimensions should therefore be interpreted with caution.

We found a marked ceiling effect in 5 of 8 dimensions (for both patients with UC and patients with CD). This finding has important implications for the use of SF-36 in longitudinal studies and indicates that the instrument may have limitations regarding detecting improvement or deterioration over time in patients with chronic conditions. Other groups have reported similar distributional properties when using the SF-36 in patients with chronic diseases.[43–45]

Test–retest reliability showed rather low ICC values for the role dimensions, and we found that the differences in mean score for physical functioning and bodily pain were significantly different in patients with UC who reported that they were stable, even though the numerical differences were small (2.7 and 3.1, respectively). Despite these findings, our interpretation is that the SF-36 has satisfactory reliability and a discriminant ability for scores for all dimensions in both patients with UC and patients with CD.

When testing for sensitivity to change, we found that the SF-36 was capable of detecting changes in mean score over time. However, when measuring responsiveness by means of the Guyatt statistic, we found that the figures for all dimensional scores were generally low. This finding, together with the high ceiling effects, may indicate that the SF-36 has limitations regarding detecting deterioration or improvement over time. This observation is in contrast to our previous findings for the disease-specific HRQOL questionnaire N-IBDQ,[46] which had high scores as measured by the Guyatt statistic for all 5 dimensions and very good responsiveness.

We would expect a disease-specific questionnaire to be more capable of detecting changes over time, because it is more sensitive to changes in disease condition than a generic questionnaire. The findings of this study support this hypothesis, but they also emphasize the importance of using a generic as well as a disease-specific questionnaire when investigating HRQOL in patients with chronic conditions.

In summary, we have shown for the first time that HRQOL in a Norwegian population-based cohort of patients with IBD, as measured with the generic questionnaire SF-36, is lower than that of a Norwegian reference population.

Patients who were symptom-free at the time of the study had a scoring profile similar to that of the reference population, but increasing symptoms led to a significantly lower HRQOL in both diagnostic groups. When using standardized scores, we found that general health, bodily pain, and role-physical were the most strongly affected dimensions in both patients with UC and patients with CD with moderate to severe symptoms. Social function and role-emotional were also strongly affected in patients with CD.

In general, the SF-36 was found to have satisfactory psychometric properties in this IBD population. However, we found marked floor and ceiling effects, which may have important implications for the use of the SF-36 in longitudinal studies and may indicate that the instrument has limitations regarding detecting improvement or deterioration over time in patients with chronic conditions.

ACKNOWLEDGMENTS

The following members of the Inflammatory Bowel South-Eastern Norway (IBSEN) Study Group of Gastroenterologists are thanked for participating in the study: Erling Aadland, Aker University Hospital; Jostein Sauar, TSSH, Skien; Tom Schulz, Sørlandet Hospital, Arendal; Njål Stray, Diakonhjemmet Hospital, Oslo; Øystein Kjellevold, TSSH, Kragerø; Erik Aubert, SØ-Sarpsborg; Taran Søberg and the late Magnus Melsom, SØ-Moss; Per Tolås, SØ-Halden; Peter Utzon, Lovisenberg Hospital; and Borgar Flaaten, TSSH-Notodden. Permission to use the Norwegian version of SF-36 was kindly granted by Professor Stein Kaasa, St. Olav University Hospital, Trondheim, Norway.

REFERENCES

1. Mitchell A, Guyatt G, Singer J, et al. Quality of life in patients with inflammatory bowel disease. *J Clin Gastroenterol*. 1988;10:306–310.
2. Drossman DA, Patrick DL, Mitchell CM, et al. Health-related quality of life in inflammatory bowel disease. Functional status and patient worries and concerns. *Dig Dis Sci*. 1989;34:1379–1386.

3. Drossman DA, Leserman J, Li Z, et al. The rating form of IBD patients' concerns: A new measure of health status. *Psychosom Med*. 1991;53: 701–712.
4. Irvine EJ. Quality of life issues in patients with inflammatory bowel disease. *Am J Gastroenterol*. 1997;92:18–24.
5. Hjortswang H, Strøm M, Almer S. Health-related quality of life in Swedish patients with ulcerative colitis. *Am J Gastroenterol*. 1998;93:2203–2211.
6. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med*. 1993;118:622–629.
7. Anderson RT, Aaronsen NK, Wilkin D. Critical review of the international assessments of health-related quality of life. *Qual Life Res*. 1993;2: 369–395.
8. Bergner M, Bobitt RA, Kressel S, et al. The sickness impact profile: conceptual formulation and methodology for the development of a health status measure. *Int J Health Serv*. 1976;6:393–415.
9. Hunt SM, McEwen J, McKenna SP. Measuring health status: a new tool for clinicians and epidemiologists. *J R Coll Gen Pract*. 1985;35:185–188.
10. Carmon E, Keidar A, Ravid A, et al. The correlation between quality of life and functional outcome in ulcerative colitis patients after proctocolectomy ileal pouch anal anastomosis. *Colorectal Dis*. 2003;5:228–232.
11. Robb B, Pritts T, Gang G, et al. Quality of life in patients undergoing ileal pouch-anal anastomosis at the University of Cincinnati. *Am J Surg*. 2002;183:353–360.
12. Loge JH, Kaasa S. Short form 36 (SF-36) health survey: normative data from the general Norwegian population. *Scand J Soc Med*. 1998;26:250–258.
13. Moum B, Vatn MH, Ekbom A, et al. Incidence of ulcerative colitis and indeterminate colitis in four counties of southeastern Norway, 1990–93. A prospective population-based study. *Scand J Gastroenterol*. 1996;31: 362–366.
14. Moum B, Vatn MH, Ekbom A, et al. Incidence of Crohn's disease in four counties of southeastern Norway, 1990–93. A prospective population-based study. *Scand J Gastroenterol*. 1996;31:355–361.
15. Moum B, Ekbom A, Vatn MH, et al. Inflammatory bowel disease: re-evaluation of the diagnosis in a prospective population based study in south eastern Norway. *Gut*. 1997;40:328–332.
16. Jahnsen J, Moum B, Schulz T, et al. Inflammatory bowel disease, disease course and status 5 years after diagnosis (The IBSEN study). *Gastroenterology*. 2000;118:1005.
17. Loge JH, Foss Abrahamsen A, Ekeberg Ø, et al. Reduced health-related quality of life among Hodgin's survivors: a comparative study with general population norms. *Ann Oncol*. 1999;10:71–79.
18. Binder V, Both H, Hansen PK, et al. Incidence and prevalence of ulcerative colitis and Crohns' disease in the county of Copenhagen, 1962 to 1978. *Gastroenterology*. 1982;83:563–568.
19. Palm Ø, Moum B, Jahnsen J, et al. The prevalence and incidence of peripheral arthritis in patients with inflammatory bowel disease, a prospective population-based study (the IBSEN study). *Rheumatol*. 2001;40:1256–1261.
20. Ware JE, Kosinski M, Keller SD. *SF-36 Physical and Mental Health Summary Scales: A User Manual*. Boston, MA: New England Medical Center, The Health Institute; 1994.
21. Loge JH, Kaasa S, Hjermstad MJ, et al. Translation and performance of the Norwegian SF-36 health survey in patients with rheumatoid arthritis. I. Data quality, scaling assumptions, reliability and construct validity. *J Clin Epidemiol*. 1998;51:1069–1076.
22. Kvien T, Kaasa S, Smedstad LM. Performance of the Norwegian SF-36 health survey in patients with rheumatoid arthritis. II. A comparison of the SF-36 with disease-specific measures. *J Clin Epidemiol*. 1998;51: 1077–1086.
23. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measurements: statistics and strategies for evaluation. *Control Clin Trials*. 1991;12:142S–158S.
24. Bulpitt CJ, Fletcher AE. The measurement of quality of life in hypertensive patients: a practical approach. *Br J Clin Pharmacol*. 1990;30: 353–364.
25. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chron Dis*. 1987;40:171–178.
26. Bernklev T, Jahnsen J, Henriksen M, et al. Sick leave, disability and educational level in patients with inflammatory bowel disease. *Scand J Gastroenterol*. 2004;240:33.
27. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res*. 1995;4:293–307.
28. Stavem K, Loge JH, Kaasa S. Health status of people with epilepsy compared with a general reference population. *Epilepsia*. 2000;41:85–90.
29. Wahl A, Loge JH, Wiklund I, et al. The burden of psoriasis: a study concerning health-related quality of life among Norwegian adult patients with psoriasis compared with general population norms. *J Am Acad Dermatol*. 2000;43:803–808.
30. Irvine EJ. Quality of life in inflammatory bowel disease: bias and other factors affecting scores. *Scand J Gastroenterol*. 1995;30:136–140.
31. Hjortswang H, Jarnerot G, Curman B, et al. The influence of demographic and disease-related factors on health-related quality of life in patients with ulcerative colitis. *Eur J Gastroenterol Hepatol*. 2003;15:1011–1020.
32. Casellas F, Lopez-Vivancos J, Casado A, et al. Factors affecting health-related quality of life of patients with inflammatory bowel disease. *Qual Life Res*. 2002;11:775–781.
33. Blondel-Kucharski F, Chircop C, Marquis P, et al. Health-related quality of life in Crohn's disease: a prospective longitudinal study in 231 patients. *Am J Gastroenterol*. 2001;96:2915–2920.
34. Maunder R, Toner B, De Rooy E, et al. Influence of sex and disease on illness-related concerns in inflammatory bowel disease. *Can J Gastroenterol*. 1999;13:782–732.
35. Moser G, Tillinger W, Sachs G, et al. Disease-related worries and concerns: a study on out-patients with inflammatory bowel disease. *Eur J Gastroenterol Hepatol*. 1995;7:853–858.
36. Zimmerman J, Gavish D, Rachmilewitz D. Early and late onset ulcerative colitis: distinct clinical features. *J Clin Gastroenterol*. 1985;7:492–498.
37. Stewart AL, Hays RD, Ware JE. The MOS short-form general health survey. Reliability and validity in a patient population. *Med Care*. 1988;26:724–735.
38. Dimenas E, Carlsson G, Glise H, et al. Reference of norm values as part of the documentation of quality of life instruments for use in upper gastrointestinal disease. *Scand J Gastroenterol Suppl*. 1966;31:8–13.
39. Bernklev T, Jahnsen J, Aadland, et al. Health-related quality of life in patients with inflammatory bowel disease five years after the initial diagnosis. *Scand J Gastroenterol*. 2004;39:365–373.
40. McColl E, Han SW, Barton JR, et al. A comparison of the discriminatory power of the inflammatory bowel disease questionnaire and the SF-36 in people with ulcerative colitis. *Qual Life Res*. 2004;13:805–811.
41. McHorney CA, Ware JE Jr, Lu JF, et al. The MOS 36-item short-form health survey (SF-36): III. tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Med Care*. 1994;32:40–66.
42. Bindman AB, Keane D, Lurie N. Measuring health changes among severely ill patients. The floor phenomenon. *Med Care*. 1990;28:1142–1152.
43. Wolinsky FD, Wyrwich KW, Nienaber NA, et al. Generic versus disease-specific health status measures. An example using coronary artery disease and congestive heart failure patients. *Eval Health Prof*. 1998;21: 216–243.
44. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol*. 1999;52:861–873.
45. Jacoby A, Baker GA, Steen N, et al. The SF-36 as a health status measure for epilepsy: a psychometric assessment. *Qual Life Res*. 1999;8: 351–364.
46. Bernklev T, Moum B, Moum T, et al. Quality of life in patients with inflammatory bowel disease: translation, data quality, scaling assumptions, validity, reliability and sensitivity to change of the Norwegian version of IBDQ. *Scand J Gastroenterol*. 2002;37:1164–1174.