The impact of paediatric inflammatory bowel disease. Epidemiology, disease activity and quality of life

Loonen, H.J.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 3.3

Measuring Quality of Life in Children with Inflammatory Bowel Disease: The Impact-II (NL)

Hester J Loonen, Martha A Grootenhuis, Bob F Last, Rob J de Haan, Jan Bouquet, and Bert HF Derkx
Abstract

Background
Inflammatory bowel disease (IBD) is a chronic debilitating disorder. Measures of quality of life are only available for adult patient populations. We developed a new disease specific health-related quality of life instrument in Dutch for paediatric patients with IBD, called Impact-II (NL). We translated and strongly modified the original (Canadian) Impact questionnaire. It comprises 35 items in six domains.

Methods
Eighty-three children (66%) completed the questionnaire, 39 children were assessed twice. Disease symptoms were recorded and disease course severity assessed through chart review. Summated disease activity scores and disease course severity scores were dichotomised into two categories.

Results
Reliability coefficients were good for five out of six domains (Cronbach’s $\alpha$ ranged from 0.57 to 0.86) and measures of test-retest stability in clinically stable patients were good for all domains (intra-class correlation coefficients ranged from 0.67 to 0.91). The instrument showed good discriminant validity between symptom groups and disease course severity on all domains. Convergent validity with a validated generic instrument [TNO-AZL Children’s Quality of life questionnaire (Taqcol)] showed satisfactory coefficients.

Conclusions
In conclusion, the developed questionnaire shows good psychometric properties. Test-retest stability and responsiveness to change should be further assessed in larger patient samples. Cross-cultural translation and validation procedures into other languages are being conducted to enable international use of Impact-II.
Introduction

Many indices to assess health status, or disease activity, in patients with inflammatory bowel disease (IBD) have been developed over the years in an attempt to objectively compare patients.\(^1\)\(^{-}6\) These measures do not assess the impact of a disease on the quality of a patients’ life. The expanding interest in the last decades in health-related quality of life (HRQoL) assessments in both research and therapeutic settings has resulted in the development of questionnaires to measure this aspect of disease. Most questionnaires have been developed in English speaking countries.\(^7\) Translation into various languages has proven to be more difficult than initially assumed.\(^8\) This is not just because of language and idiom differences, but more importantly because of cultural differences in perception of health problems and societal values placed upon these problems.\(^9\) In multinational clinical trials where HRQoL assessment is more and more included as an outcome parameter, reliable and (cross-culturally) validated questionnaires are essential.

Disease specific instruments are only available for a small number of patient populations. Generic measures have the advantage that they are not developed for a specific target population and are therefore suitable for comparison across various patient populations or healthy reference groups.\(^10\),\(^11\) They may, however, not be very sensitive to specific problems of the studied patient population. In contrast, disease specific instruments are more sensitive and therefore more suitable to detect small but clinically important differences in HRQoL between and within patients. It is often recommended when studying HRQoL of patients to use both a generic as well as a disease specific instrument. In paediatrics few validated generic questionnaires are available. Disease specific questionnaires are even more scarce and are mostly developed in asthma, diabetes, epilepsy and cancer.\(^12\),\(^13\),\(^20\)

Inflammatory bowel disease is a chronic condition with symptoms that can cause significant emotional and social impairments.\(^21\)\(^{-}26\) It has an intermittent course, with unexpected recurrences. Patients often require life long medication use with numerable side effects. Many patients at some time during the course of their disease will require surgical removal of parts of the bowel because of intractable disease. In order to optimise chronic care for children with this disorder, and to monitor emotional and social development, measurements of HRQoL can provide new insights and targets for intervention. For adults a disease specific HRQoL instrument has been developed and validated.\(^27\)\(^{-}30\) Recently, a group in Canada developed a disease specific questionnaire for paediatric IBD, called Impact.\(^31\)

The purpose of our study was to develop a Dutch version of the Impact. Based on pilot studies both in the Netherlands and the United Kingdom, the original Impact was found by children and physicians to be difficult and multi-interpretatable. We describe the translation, according to proposed guidelines,\(^7\) and the subsequent thorough modifications made to the
Part III Quality of Life: Methodology

Impact, resulting in the development of the Impact-II. Furthermore, we assessed its psychometric properties.

Material and methods

Patients

Inclusion criteria were definite IBD (Crohn’s disease, ulcerative colitis or IC) for more than 6 months, and age between 8 and 18 years old inclusive. Exclusion criteria were: isolated proctitis, mentally unable to fill out a questionnaire, and not able to read and understand Dutch. Patients from two hospitals were contacted: the Emma Children’s Hospital, AMC, Amsterdam, and the Sophia Children’s Hospital, Rotterdam, The Netherlands. Both hospitals serve as secondary care centres for the western part of the Netherlands, and they are tertiary referral centres for the central and western part of the Netherlands. This makes the population in these hospitals a good representation of the paediatric IBD population in the Netherlands.

Measurements

The Impact-II (NL)

The original Impact questionnaire was developed in Canada in a thorough and structured way. It consists of 33 questions with a 10 cm visual analogue scale (VAS) as the answering mode with anchoring statements expressing extreme answers. This format was preferred over categorical response options because of ease of completion for children. Scores are multiplied by 0.7 to mirror the score range in the adult instrument. Scores on the Impact could therefore more easily be interpreted by people familiar with the widely used adult instrument. Potentially relevant issues and concerns to children were gathered through interviews with patients and caregivers, experts and literature. Item-reduction procedures based on item importance and frequency, were performed for all raised issues and concerns on 117 paediatric patients. Items included in the final questionnaire aim to assess both concerns of children with Crohn’s disease and ulcerative colitis, as well as concerns of younger patients and older ones. The questionnaire is organized in six hypothesized subject scales or domains: IBD related symptoms, emotional functioning, social functioning, systemic symptoms, treatment/intervention related concerns and body image.

Translation into Dutch was done following the guidelines proposed by Guillemí et al. in 3 steps: (1) translation was done by an independent professional translator. Medical idioms were translated by Dutch paediatric gastroenterologists fluent in English. (2) Back-translations were performed by two professional translators. The back-translation version was approved by the developers of the original questionnaire to having retained its original meaning. (3) Committee review of translations and back-translations. The Dutch translation was pilot tested on 10 healthy and 15 affected subjects to assess its feasibility. Based upon the
children’s comments and review by an expert committee consisting of two psychologists
expert in the field of questionnaires, one paediatric gastroenterologist and one paediatric
researcher, a number of difficulties were encountered. 1) The questionnaire was found to
consist of many long, double-barrelled and multi- interpretable questions and answers. 2) The
absence of a specified time frame in the questions was considered sub-optimal. 3) The format
with the answer representing most impairment of HRQoL on the left hand side of the VAS
was found to be leading and upsetting to children. Therefore, thorough modification was
performed (see Figure 1). Answering phrases were significantly simplified in that answering
statements were changed into ‘very much’ and ‘not at all’ or ‘very often’ and ‘never’, where
applicable.

Two double-barrelled questions were separated into four questions, one question which
was considered inappropriate to Dutch children was omitted (‘feeling stressed out’), and one
question was added based upon clinical experience of the researchers (‘able to talk to anyone
about bowel condition’). A time frame of ‘past 2 weeks’ was included into those questions
asking about frequency of problems. The answering anchors were switched with the answer
representing least impairment on HRQoL on the left hand side of the VAS. This resulted in a
rigorously modified 35-item questionnaire, which was called the Impact-II (NL) (Appendix I).
The original six domains were retained.

Disease activity

A five item symptom card asking about presence and severity of stomach aches, presence
and frequency of diarrhoea, presence of fever, presence and amount of blood in stool and
presence and amount of weight loss over the last two weeks was constructed to assess disease
activity (Table 1). Summated scores range from 5-17. Children were concluded to have
inactive or mild disease activity if their score was 5-8 points, and moderate to severe if they
scored 9-17 points.

Disease course severity

Disease course severity since the onset of disease was assessed by the researcher (HJL)
through chart review. For seven characteristics, points were scored and final scores were
summated (Table 2). Total scores could range between 4 and 21. Children were said to have
mild disease course (total score 4-7) or moderate to severe disease course (total score 8-21).

Procedures

Children were sent the Impact-II (NL), the Tacqol and the five-item symptom card. Ten
children completed the questionnaire in the hospital at the time of an outpatient visit or
clinical admission. Non-responders were contacted once by telephone. For test-retest analysis,
the first 52 responders were sent the Impact-II (NL) questionnaire a second time, 3 weeks
after the child had filled out the questionnaire the first time, accompanied by the symptom score card. Also, they were asked whether they thought the activity of their disease had changed compared to the first assessment. Answering options were: improved, the same, or deteriorated.

The study was approved by the medical ethics committee in both centres and all patients and parents provided written informed consent to participate.

Figure 1.
Example of modification of Impact into Impact-II.

(Original) Q: Put a mark on the line to tell how much your stomach has been hurting recently.

A
My stomach has been hurting

B
I have not had any stomach

so much, the pain is as bad as

cramps at all recently.

it has ever been, and comes all

the time; it’s really hard to deal with.

(Original) Q: Do you feel you have to give up or miss out on things you would like to do such as hobbies, just playing, going to parties or other special things?

A
I always have to stay out of

B
I can do all the things I like as much

things; I feel like I’m missing

as I want. I don’t feel my bowel

out on everything.

condition gets in the way at all.

(Original) Q: How often did you have to miss out on things (like hobbies, just playing or going to parties) in the last two weeks?

Never

Always

Statistical analysis

Differences between responders and non-responders were analysed on the following characteristics: age (student t-test), disease type ($\chi^2$-test), disease course severity (Mann-Whitney U-test) and disease duration shorter or longer than one year ($\chi^2$-test).

Item scores were obtained by measuring in centimetres with one decimal the middle of the cross children placed on the VAS. Per domain summated scale scores were obtained by adding the items on the scale, and transforming scale scores linearly to a 0-100 scale, with higher scores representing better HRQoL.

Reliability in terms of homogeneity was evaluated by calculating the standardized Cronbach’s $\alpha$ for each scale.$^{33}$ Homogeneity is considered to be good if alpha > 0.60 and excellent in the case of values > 0.90. Reliability also refers to the question whether an
instrument gives the same score if measured repeatedly assuming the studied concept has not changed. Test-retest reliability was analysed calculating the intra class correlation coefficient (ICC) on scores of stable patients. Two ways of determining clinical stability were performed: patients who rated their disease as 'the same' on the one hand, and patients who showed a change in symptoms of −1 to +1 on the symptom card on the other hand. These cut-off borders to classify children as changed or unchanged were also previously used by Juniper et al. in a study on asthma. ICCs should exceed 0.60 for all domains, since it has been shown that measures of relatively 'unstable' states, such as anxiety, show such an ICC.

Validity concerns the question whether an index actually measures what it is intended to measure. Validation testing was based upon content validity (do the components of the scale cover all the aspects of the attribute being measured), convergent validity (does the scale correlate with another scale that is supposed to measure somewhat the same construct) and discriminant validity (does the scale discriminate between subjects with an expected difference in quality of life, for example does it discriminate between children with varying disease activity or disease course severity).

Content validity was assured by asking children at the end of the questionnaire whether any issues or concerns were missing.

Convergent validity was measured using the validated generic HRQoL instrument Tacqol. Equivalent items in both questionnaires were used to correlate the Impact-II (NL) to the validated generic one, expressed as Spearman rank order coefficients. Equivalent scales were compared calculating the Pearson product-moment correlation coefficient (PPMCC).

Discriminant validity was assessed performing the student t-test on scale scores of patients in the two disease activity and the two severity categories. This difference was expected to be significant for all domains on both disease activity and disease course severity, expressed as a $p$ value $< 0.05$.

Statistical analyses were done using SPSS version 9.0.

Results

Patients

One hundred and twenty-six children fitted the inclusion criteria, of which 83 (66%) returned the questionnaires. Demographic characteristics of responders and non-responders in the total study group are presented in Table 3.

The only significant difference between responders and non-responders was age (non-responders were older, $p$ value for the difference in mean age= 0.043). For all other variables, no differences were found.

For test-retest analyses, 39 of 52 children (75%) completed the questionnaire again within a time frame of 4-10 weeks (average 7 weeks). Twenty-two children were stable with respect
to the difference in self-reported disease activity (symptom card), six had improved and 10 had deteriorated. One child did not fill out the symptom card the second time. With respect to a comparison of symptoms, 15 reported to be the same, seven had deteriorated and nine said they had improved. Eight did not fill out this question. It was therefore decided that this way of determining clinical stability was not used.

**Table 1.**

Symptom card.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diarrhoea</td>
<td>No</td>
<td>Yes, 1-6 times/day</td>
<td>Yes, 6-10/day</td>
<td>Yes, &gt;10/day</td>
</tr>
<tr>
<td>Blood with your bowel movement</td>
<td>No</td>
<td>Yes, sometimes/a little</td>
<td>Yes, moderate</td>
<td>Yes, often/a lot</td>
</tr>
<tr>
<td>Stomach aches</td>
<td>No</td>
<td>Yes, little</td>
<td>Yes, moderate</td>
<td>Yes, severe</td>
</tr>
<tr>
<td>Fever</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight loss</td>
<td>No</td>
<td>Yes, up to 1 kg</td>
<td>Yes, more than 1 kg</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2.**

Disease course severity scoring card.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio of hospital admissions and diseased years</td>
<td>NA</td>
<td>0 - 0.25</td>
<td>0.25 - 1.0</td>
<td>&gt; 1.0</td>
</tr>
<tr>
<td>Number of exacerbations</td>
<td>NA</td>
<td>1</td>
<td>2-3</td>
<td>&gt;3</td>
</tr>
<tr>
<td>Time of corticosteroid use per exacerbation</td>
<td>NA</td>
<td>0-3 months</td>
<td>3-6 months</td>
<td>&gt; 6 months</td>
</tr>
<tr>
<td>Azathioprine use</td>
<td>No</td>
<td>As starting therapy</td>
<td>For corticosteroid-refractory disease</td>
<td></td>
</tr>
<tr>
<td>Cyclosporine</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surgery</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Height by age</td>
<td>NA</td>
<td>No growth delay</td>
<td>&lt;2 SD under normal curve</td>
<td>&gt;2 SD under normal curve</td>
</tr>
</tbody>
</table>

NA= not applicable; SD= standard deviation.

**Measurements**

**Reliability**

Internal consistency was satisfactory for five out of the six hypothesized domains (α ranged from 0.57 to 0.85). Only the treatment/intervention domain did not reach the required 0.60 value. ICCs were good for all domains (range 0.67 to 0.91) (Table 4).
Chapter 3.3 Validation of Impact-II

Validity

Only one child reported that an issue important to her was not included in the questionnaire (the role of nutrition with IBD). Student t-test showed all domains of the Impact-II (NL) to significantly discriminate between patients with mild and moderate symptoms, and between mild and moderate disease course severity (Table 5). Convergent validity showed correlation coefficients in the mid-range between overlapping questions and scales of the Tacqol (Table 6).

Table 3.
Patient characteristics: responders and non-responders.

<table>
<thead>
<tr>
<th></th>
<th>Responders N= 83 (66%)</th>
<th>Non-responders N= 43 (34%)</th>
<th>Value of test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male: n (%)</td>
<td>45 (54%)</td>
<td>22 (50%)</td>
<td>( \chi^2 = 0.11 ) (p= 0.74)</td>
</tr>
<tr>
<td>Mean age: years (SD)</td>
<td>14.3 (2.7)</td>
<td>15.3 (2.3)</td>
<td>t= 2.05 (p = 0.043)</td>
</tr>
<tr>
<td>Disease type n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crohn’s disease</td>
<td>41 (50%)</td>
<td>19 (44%)</td>
<td>Fisher’s exact= 0.32 (p= 0.85)</td>
</tr>
<tr>
<td>Ulcerative colitis</td>
<td>40 (48%)</td>
<td>23 (54%)</td>
<td></td>
</tr>
<tr>
<td>Indeterminate colitis</td>
<td>2 (2%)</td>
<td>1 (2%)</td>
<td></td>
</tr>
<tr>
<td>Disease activity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inactive – mild</td>
<td>56 (67%)</td>
<td>NA*</td>
<td></td>
</tr>
<tr>
<td>Moderate – severe</td>
<td>25 (31%)</td>
<td>NA*</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>2 (2%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disease course severity</td>
<td></td>
<td></td>
<td>( \chi^2 = 0.87 ) (p= 0.35)</td>
</tr>
<tr>
<td>Inactive – mild</td>
<td>41 (50%)</td>
<td>25 (58%)</td>
<td></td>
</tr>
<tr>
<td>Moderate – severe</td>
<td>42 (50%)</td>
<td>18 (42%)</td>
<td></td>
</tr>
<tr>
<td>Disease duration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean years (SD)</td>
<td>3.0 (2.8)</td>
<td>3.9 (3.0)</td>
<td>( \chi^2 = 1.59 ) (p=0.21)</td>
</tr>
<tr>
<td>( \leq 1 ) year</td>
<td>24</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>( &gt;1 ) year</td>
<td>59</td>
<td>35</td>
<td></td>
</tr>
</tbody>
</table>

NA = non-applicable since disease activity was assessed by the symptom card.

Table 4.
Internal consistency and intra-class correlation coefficients of subscales.

<table>
<thead>
<tr>
<th>Item consistency for hypothesized domains</th>
<th>Number of items</th>
<th>Cronbach’s ( \alpha ) (standardised( a ))</th>
<th>ICC( b )</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBD symptoms</td>
<td>7</td>
<td>0.73</td>
<td>0.71</td>
</tr>
<tr>
<td>Systemic symptoms</td>
<td>3</td>
<td>0.84</td>
<td>0.80</td>
</tr>
<tr>
<td>Emotional functioning</td>
<td>7</td>
<td>0.86</td>
<td>0.91</td>
</tr>
<tr>
<td>Social functioning</td>
<td>12</td>
<td>0.85</td>
<td>0.88</td>
</tr>
<tr>
<td>Body image</td>
<td>3</td>
<td>0.65</td>
<td>0.83</td>
</tr>
<tr>
<td>Treatment/interventions</td>
<td>3</td>
<td>0.57</td>
<td>0.67</td>
</tr>
</tbody>
</table>

\( a \) Standardised item-\( \alpha \) means the \( \alpha \) corrected for the number of items within the scale.

\( b \) ICC in stable patients as assessed by the five-item symptom card.
Discussion

We describe the psychometric properties of a new health-related quality of life questionnaire to be used in children with IBD. The questionnaire has been developed in a thorough and structured way, with input of a large number of patients with IBD, ensuring the content validity and sensitivity of the instrument. The questionnaire’s format has been reviewed by experts in the field, ensuring its feasibility for use in children.

A response rate of 66% can be considered satisfactory for a mailing survey. Responders did not differ from non-responders with respect to type of disease, disease course severity or disease duration. The only difference was that non-responders were on average one year older (p = 0.043). A possible explanation for this difference could be lying in rebellious adolescent behaviour, not being very motivated to report on emotional and social well being.

All six scales have good internal consistency coefficients, although the scale assessing concerns and worries over treatments or interventions did not completely fulfil the criteria. In the test-retest analysis, this scale also showed the lowest ICC, although this did fulfil the criteria. This scale is a very important scale however because it covers three items that assess important emotional concerns of children. We therefore decided to retain the scale in its present format.

We encountered a couple of problems in the test-retest study. First of all, as the study design was a mailing survey, we were dependent upon the respondents’ timing of completion and returning of the questionnaire. Despite explanations to the children that it should be done directly after receiving it, we received the questionnaires back with a mean interval of seven weeks between initial completion and second completion. This is in the literature considered to be a too long interval to assume the measured construct, HRQoL, has not changed. However, we did have two measures of disease stability included in our study design: 1) children were asked to rate their disease as improved, the same or deteriorated compared to the first assessment; and 2) the symptom diary was filled out the first and the second time allowing for a computation of a change in symptoms. Of the 39 respondents, eight did not fill out the question comparing their disease, suggesting that it might be too difficult for some children to remember. Therefore, the validity of the question for the remaining thirty-one respondents can be questioned. We therefore chose to use the more objective measure of symptoms. Based on the hypothesis that HRQoL is influenced by disease activity, this can be considered a good “proxy” rating for stability of disease.
Table 5.
Mean scale scores on the Impact-II (NL) in relation to disease activity and disease severity.

<table>
<thead>
<tr>
<th>Scale score mean (SD)</th>
<th>Disease activity</th>
<th>Disease course severity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mild (n=56)</td>
<td>Moderate (n=25)</td>
</tr>
<tr>
<td>IBD symptoms</td>
<td>85 (11)</td>
<td>66 (18) ***</td>
</tr>
<tr>
<td>Systemic symptoms</td>
<td>77 (22)</td>
<td>55 (27) ***</td>
</tr>
<tr>
<td>Emotional functioning</td>
<td>79 (16)</td>
<td>63 (20) ***</td>
</tr>
<tr>
<td>Social functioning</td>
<td>86 (9)</td>
<td>72 (18) ***</td>
</tr>
<tr>
<td>Body image</td>
<td>75 (19)</td>
<td>63 (25) *</td>
</tr>
<tr>
<td>Treatment/investigations related</td>
<td>74 (18)</td>
<td>60 (25) **</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mild (n=41)</th>
<th>Moderate (n=42)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBD symptoms</td>
<td>83 (11)</td>
<td>75 (19) **</td>
</tr>
<tr>
<td>Systemic symptoms</td>
<td>80 (19)</td>
<td>60 (28) ***</td>
</tr>
<tr>
<td>Emotional functioning</td>
<td>79 (16)</td>
<td>69 (21) **</td>
</tr>
<tr>
<td>Social functioning</td>
<td>88 (8)</td>
<td>76 (16) ***</td>
</tr>
<tr>
<td>Body image</td>
<td>76 (20)</td>
<td>66 (22) *</td>
</tr>
<tr>
<td>Treatment/investigations related</td>
<td>75 (18)</td>
<td>63 (23) **</td>
</tr>
</tbody>
</table>

Higher scale scores represent better quality of life. Differences in mean scale scores between the groups were tested with the Student t-test.
* p < 0.05, ** p < 0.01, *** p < 0.001.

Table 6.
Convergent validity of the Impact-II (NL): correlations with overlapping items and domains on the Tacqol.

<table>
<thead>
<tr>
<th>Impact-II (NL) item</th>
<th>Tacqol item</th>
<th>Spearman rank order coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stomach aches</td>
<td>Stomach/tummy aches</td>
<td>0.63</td>
</tr>
<tr>
<td>Sick to stomach</td>
<td>Sick to stomach</td>
<td>0.46</td>
</tr>
<tr>
<td>Feeling tired</td>
<td>Tired</td>
<td>0.58</td>
</tr>
<tr>
<td>Being happy</td>
<td>Happy</td>
<td>0.44</td>
</tr>
<tr>
<td>Having had fun</td>
<td>Feeling gay</td>
<td>0.62</td>
</tr>
<tr>
<td>Having had fun</td>
<td>Feeling cheerful</td>
<td>0.58</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Impact-II (NL) scale</th>
<th>Tacqol scale</th>
<th>PPMCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBD symptoms</td>
<td>Body complaints</td>
<td>0.58</td>
</tr>
<tr>
<td>Systemic symptoms</td>
<td>Body complaints</td>
<td>0.72</td>
</tr>
<tr>
<td>Emotional functioning</td>
<td>Positive emotions</td>
<td>0.46</td>
</tr>
<tr>
<td>Social functioning</td>
<td>Social</td>
<td>0.59</td>
</tr>
</tbody>
</table>

PPMCC = Pearson product moment correlation coefficient.

Consensus has not been reached, but the common opinion is that when analysing test-retest stability of a measure of a subjective trait, stability of the trait or derivatives thereof should be assessed by the patient himself. Also, disease activity is considered one important determinant of HRQoL, but it is certainly not the only one. Other (unmeasurable) events can cause the patient to respond to the questionnaire in a different fashion than the first time, for example due to response shift bias.\(^{35}\) This bias refers to the fact that internal standards of patients can change, influencing the reporting of perceived HRQoL. So, it could very well be that patients rate their disease activity as unchanged, but that their HRQoL score changes dramatically. This could be concluded as a failure of the questionnaire to measure the trait, but the
aforementioned response shift could be another explanation of the change in score. Very few data have been published on test-retest accuracy in children and response-shift bias has never been described in children. Despite all these encountered problems, the reported ICCs are fairly good for the domains. Larger samples and more structured studies should provide more details on this important psychometric property of the instrument in the future.

Another weakness of our test-retest results is that we used an unvalidated measure of disease activity, the symptom card. Although Crohn’s disease, ulcerative colitis and indeterminate colitis are widely accepted to belong to the group of inflammatory bowel diseases, no single validated disease activity index exists for IBD. There is one disease activity index developed for children with Crohn’s disease, however this index requires blood sampling for calculations which is not feasible for survey research. Furthermore, no index exists for children with ulcerative colitis or indeterminate colitis, and no index covers all three disease entities. Therefore, a global symptom activity card was the best possible disease activity assessment we could use in our study. However, due to this weakness, test-retest results should be interpreted with caution, and responsiveness analyses on scores of clinically changed versus stable patients were therefore not performed. In the future, clinical studies using both the Impact questionnaire and validated disease activity indices for Crohn’s disease and ulcerative colitis should reassess test-retest reliability and discriminant validity of the Impact questionnaire.

Convergent validity, comparing the disease specific questionnaire to a generic one, has some practical problems. First of all, one considers the validated measure as the “gold standard” and therefore correlates the newly developed one to see how well it fits this gold standard. It is obvious, however, that a generic measure and a disease specific instrument aim to measure somewhat the same construct, but will not exactly do so. One wants to find coefficients that are not too low, suggesting poor validity, and not too high, undermining the additional value of the new questionnaire to the old one. When comparing individual items on the generic and disease specific instrument, moderate coefficients were found. This can be explained by the fact that the items cover somewhat the same issues, however they are asked in a different fashion, using different answering options. Also, the generic measure asks about health-related problems in the past few weeks, where the disease specific instrument asks about a more specific time frame (past two weeks). Taking all these aspects of validity testing into consideration, the authors feel that the described mid range coefficients between the two instruments satisfactorily support the validity of the newly developed instrument.

Generic instruments and disease specific questionnaires both have their strengths and weaknesses. Generic measures allow for comparisons between populations, but they have been criticised for their relative insensitivity to small differences. Disease specific instruments are particularly valuable when assessing changes related to a particular disease. Although high correlations between the two instruments found in our study might suggest the new index to be simply a mirror of the generic instrument, the value of both types of instruments for this
population should be further assessed in a cross sectional as well as in a longitudinal study design.

The Impact-II (NL) significantly discriminates between patients with varying disease activity and disease course severity. This gives a first impression of the potential responsiveness of the questionnaire, which is an important characteristic if it were to be used as an outcome parameter in clinical trials. However, between-patient discrimination is important, but within-patient responsiveness to change is not the same concept. Especially in the paediatric IBD population, where sample sizes of trials are small and therapeutic differences hard to find, responsiveness to change is an important feature of a HRQoL instrument, since that kind of study design makes it hard to find cross sectional differences in any parameter. This characteristic should be thoroughly tested in a longitudinal study design.

Psychometric analysis is an ongoing process that can never be fully described in a single study. Therefore, we recommend that the Impact-II should be incorporated into clinical trials in order to fully appreciate its usefulness as an outcome measure both in clinical care and in research. Cross-cultural translation and validation studies are currently being planned to enable international use of the questionnaire. Furthermore, on preliminary meetings with members of the European Society for Paediatric Gastroenterology and Nutrition, the Impact-II has been accepted to be the instrument to be used to assess HRQoL in clinical trials in children with IBD after its validity has been proven.

References


### Appendix 1.

**Items and scales of Impact-II (NL).**

<table>
<thead>
<tr>
<th>Scale</th>
<th>Items</th>
</tr>
</thead>
</table>
| IBD symptoms      | Stomach aches  
Not being able to eat what you want because of disease  
Diarrhoea  
Worried about blood with bowel movement  
Being sick  
Afraid to soil paints  
Having to pass gas |
| Systemic symptoms | How much energy  
How did you feel  
How tired did you feel |
| Emotional functioning | Worried about having a flare-up  
Worried about having a chronic condition  
Worried about health in future  
Thinking it is unfair to have this disease  
Being angry to have this disease  
Being ashamed  
Being happy |
| Social functioning  | The influence of the disease upon the family  
Having to miss out on hobbies  
Having rules imposed because of the disease  
Having fun  
Is it harder to make friends  
Worried not to be able to go out on dates  
Teased or bullied because of the disease or treatment  
Does the disease make it harder to travel or go on holiday  
Try and keep your disease a secret  
Able to talk to anyone about worries  
Able to play sports as much as you would like  
Able to go to school |
| Body image         | How do you feel about height  
How do you feel about weight  
How do you feel about the way you look |
| Treatment/interventions | How do you feel about taking medicines  
How do you feel about investigations  
Worried about ever having an operation |