Clinimetric properties of 3 instruments measuring postoperative recovery in a gynecologic surgical population

Kirsten B. Kluivers, MD,a Jan C. M. Hendriks, PhD,b Ben W. J. Mol, MD, PhD,c,d
Marlies Y. Bongers, MD, PhD,e Mark E. Vierhout, MD, PhD,f Hans A. M. Brößmann, MD, PhD,f and
Henrica C. W. de Vet, PhD,g Veldhoven, Nijmegen, and Amsterdam, the Netherlands

Background. General, health-related quality-of-life questionnaires and recovery-specific questionnaires have been used to measure recovery in surgical patients. The aim of this study was to evaluate the clinimetric properties of 3 recovery instruments and to examine whether recovery-specific instruments are useful.

Methods. The Quality of Recovery-40 (QoR-40), Recovery Index-10 (RI-10), and RAND-36 health survey were used to measure recovery in women undergoing different types of hysterectomy in the first 12 weeks after operation. Construct validity was assessed by testing predefined hypotheses. The changes observed during the postoperative period were used as indicators for responsiveness.

Results. One hundred and sixty-one women were included. Response rate and internal consistency were found satisfactory. The highest number of hypotheses used for assessment of construct validity was confirmed in the RI-10. The RI-10 was more responsive compared with the QoR-40 and the RAND-36.

Conclusions. Because construct validity and responsiveness were greatest in the RI-10, we conclude that this short recovery-specific instrument is useful in studies evaluating postoperative recovery. We recommend the use of the RI-10, unless the immediate postoperative days are of interest in which the QoR-40 was valid.

Accepted for publication March 30, 2008.
Reprint requests: Kirsten B. Kluivers, MD, Department of Obstetrics and Gynecology, Radboud University Nijmegen Medical Centre, P.O. Box 9101, 6500 HB Nijmegen, the Netherlands. E-mail: K.Kluivers@obgyn.umcn.nl.

0039-6060/$ - see front matter
© 2008 Mosby, Inc. All rights reserved.
doi:10.1016/j.surg.2008.03.027
instrument or other recovery-specific instruments. The SF-36 is identical to the RAND-36, which was used in this study. Myles et al have developed recently an instrument to measure quality of recovery (QoR-40) and have tested the validity, reliability, and responsiveness of the instrument in 160 patients after a wide range of operative interventions. They concluded that the QoR-40 was a good objective measure of quality of recovery after anesthesia and operation.

We translated the QoR-40 in Dutch and developed a recovery-specific instrument, the Recovery Index-10 (RI-10). The objective of the study was to compare these instruments with the RAND-36 health survey and to study their performance for the evaluation of recovery after gynecologic open and laparoscopic interventions.

MATERIAL AND METHODS

Population. This clinimetric study was performed using the data from a randomized and parallel, nonrandomized study comparing the outcomes of different types of hysterectomy. Patients were recruited from August 2002 until January 2005 in the Máxima Medical Centre, a large teaching hospital with 865 beds on 2 locations in the southern part of the Netherlands. Women who underwent a hysterectomy with or without salpingo-oophorectomy for benign and malignant disease were asked to participate in the study. Patients who needed simultaneous interventions like vaginal repair or who were not fluently speaking Dutch were not eligible for the study. A vaginal hysterectomy was preferred unless there was suspicion of malignancy, the size of the uterus exceeded 12 weeks’ gestation, or the descent of the uterine cervix under traction did not reach halfway the length of the vagina. In patients who did not meet these criteria, either a laparoscopic or abdominal hysterectomy was performed for a mobile uterus not exceeding 18 weeks’ gestation and in case there was no suspicion of endometrial carcinoma other than FIGO stage I (ie, limited to the uterine corpus). After informed consent was given, those patients in whom a vaginal hysterectomy was not possible but a laparoscopic hysterectomy was feasible were randomized to either laparoscopic or abdominal hysterectomy. When the uterus exceeded 18 weeks’ gestation and if there was suspicion of advanced cancer, an abdominal hysterectomy was always preferred.

The patients were asked to complete the baseline measurement of the QoR-40 and RAND-36 after randomization but before the operation (median 4 weeks before hysterectomy, range 1–13 weeks). Postoperatively, the RI-10 and RAND-36 were completed at 5 time points, and the QoR-40 was completed at 8 time points as shown in Table I. The researchers did not assist the patients during completion of the questions. The questionnaires that were completed after discharge were returned by mail.

Instruments. The QoR-40 is a questionnaire containing 40 items on recovery, which consists of 5 subscales: emotional state, physical comfort, psychological support, physical independence, and pain. A Dutch translation was used. Each item was answered on a 5-point Likert scale, ranging from none of the time to all the time. The QoR-40 score was defined as the sum of the scores of all items. The QoR-40 score ranged from 40 to 200, in which 200 indicates a perfect recovery. The items referred to the past week. Because most items in the RI-10 referred to the postoperative situation, no baseline measurement was available. The items were best suited for recently discharged patients but could be filled out in the hospital as well.

Furthermore, we developed the Recovery Index-10 (RI-10), a 10-item questionnaire measuring postoperative recovery on 5-point Likert scales ranging from full disagreement to full agreement. Two gynecologists developed the RI-10 in 1996. At that time, no other recovery instruments like the QoR-40 were available. The instrument was considere to be 1 scale, and the score ranged from 40 to 200, where 200 indicates a perfect recovery. A translation of the RI-10 is shown in Appendix I. The items referred to the past week. Because most items in the RI-10 referred to the postoperative situation, no baseline measurement was available. The items were best suited for recently discharged patients but could be filled out in the hospital as well.

The RAND-36 health survey is a widely known, general, health-related quality-of-life instrument. The RAND-36 is identical to the SF-36 but is generally available and free of charge. We used

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Baseline</th>
<th>1 day</th>
<th>2 day</th>
<th>3 day</th>
<th>1 week</th>
<th>2 weeks</th>
<th>4 weeks</th>
<th>6 weeks</th>
<th>12 weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td>QoR-40</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>RI-10</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>RAND-36</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

QoR-40, Quality of Recovery-40; RI-10, Recovery Index-10; RAND-36, RAND-36 health survey; X, measurement has been performed.
the validated Dutch version. This instrument measured subjective health status in 8 subscales. The range per subscale was 0–100, and thus, the total score ranged from 0 to 800, where 0 was the poorest quality of life and 800 was the best imaginable. The version referring to the past 4 weeks was used in this study.

**Statistical analysis.** A power analysis was performed for a randomized, controlled trial on the difference in quality of life between laparoscopic and abdominal hysterectomy patients. A difference of 15 per scale was considered as clinically relevant. With a standard deviation of 20, a type I error of 0.05 and 80% power, 28 patients were needed per arm. During the intake period of a randomized, controlled trial between August 2002 and January 2005, patients undergoing hysterectomy who did not give consent or were not eligible for randomization filled out the recovery questionnaires as well. This trend has resulted in a total of 161 patients who were included in a cohort used for the present validation of the questionnaires.

The response rate and number of missing items were identified. When items were missing and more than half of the items of the particular subscale were missing, the score of that subscale as well as the total scale were regarded as missing. When half or less of the items of the subscale were missing, the average of the available items was used to complete the missing items and to determine the total score. Only the total scores, ie, not the scores on subscales, were used in the study.

The Cronbach α was calculated at 1 week postoperatively for each questionnaire to assess the internal consistency. The point of measurement at 1 week was used, because this was the first where data on all 3 instruments were available.

To assess the construct validity of the instruments in their use of postoperative recovery, hypothesis testing was used. The hypotheses were based on literature and defined before the study.

The 6 hypotheses tested on the 3 instruments were as follows:

Patients younger than 45 years have a better recovery compared with patients older than 60 years.

Patients with ASA score 1 do recover better compared with patients with ASA 2 or more.

Patients with a total hospital stay of 7 days or more have a worse recovery compared with patients admitted for 4 days or less.

Patients with complications affecting recovery (ie, all complications leading to readmission or reoperation, visceral lesions without readmission or reoperation, neurological damage, exacerbation of preexisting colitis, severe wound infection, and preexisting micturition problem) have a worse recovery compared with patients without any such complications.

Patients undergoing vaginal hysterectomy have a better recovery compared with patients undergoing open abdominal hysterectomy. Patients with peroperative conversions from laparoscopic hysterectomy to abdominal hysterectomy, laparotomy caused by complications in the days after laparoscopic hysterectomy and peroperative conversions from vaginal hysterectomy to abdominal hysterectomy were excluded from this hypothesis.

Patients with a hemoglobin decrease of less than 0.8 g/dL (0.5 mmol/L) have a better recovery compared with patients with a hemoglobin decrease of more than 2.4 g/dL (1.5 mmol/L). Patients who received a blood transfusion were excluded from this hypothesis, because their blood transfusions were expected to influence the effects on recovery of the hemoglobin decrease (preoperative hemoglobin level versus the level on the first postoperative day).

Differences between the groups, as defined in each of the 6 hypotheses of construct validity, were tested for statistical significance using the test of Mann-Whitney. The instrument for which most hypotheses were holding true was considered as having the greatest construct validity. Again the data at 1 week postoperatively were used. As the instruments referred to different time periods, ie, the past 24 hours, past week, and past 4 weeks, the testing of the aspects of validity as mentioned above was repeated making use of the 4-week data.

Correlations among the 3 instruments at 1 week after operation were calculated using the Spearman rank correlation coefficient.

Each instrument was assessed at several points in time. We assumed that the postoperative patients recovered continuously during at least 6 weeks. The responsive period of the instrument can be defined as the time period in which the score shows a change over time. The range of the mean scores over time as a percentage of the instruments total range was used as an indicator of responsiveness. Floor and ceiling effects were rated as present when more than 15% of respondents to an instrument achieved the least or greatest possible score, respectively. A linear, mixed model was used to study the mean levels in the quality-of-life score over time. The dependent variable was the quality-of-life score. The independent class variables were patient and time in weeks after operation. The intercept of each patient was treated as a random variable in the model. This way, random
differences between patients are allowed. The estimated mean levels per week with 95% confidence bands were visualized in figures.

**Recovery after hysterectomy as measured by the 3 instruments.** Difference in recovery over time on the 3 instruments was studied in the RCT comparing the data of patients randomized to laparoscopic versus abdominal hysterectomy, who were not threatened by malignancy. The recovery scores during the suggested responsive time period of each instrument were used. The treatment effect was analyzed on an intention-to-treat basis.

A linear, mixed model was used to study the increase in the quality-of-life score over time in the treatment groups. The dependent variable was the quality-of-life score. The independent class variables were patient, time in weeks after operation, and treatment group (laparoscopic and abdominal hysterectomy). In case of the RAND-36 and QoR-40, the baseline level was added to the model as an independent regression variable. The intercept of each patient was treated as a random variable in the model. Initially, the interaction term between time and group was added to the model; this term, however, never reached the level of statistical significance and was removed from the final model. The estimated regression parameters with standard errors of each score were used to calculate the average level per week in each treatment group. These levels with confidence bands are further presented in figures.

All data were analyzed using SAS 10.0 (SAS Institute, Inc., Chicago, IL) and SPSS 13.0 (SPSS, Inc., Inc., Chicago, IL). P-values below 0.05 were considered as statistically significant.

**RESULTS**

We included a total of 161 women undergoing hysterectomy in the clinimetric study. Twenty-one women underwent vaginal hysterectomy, 76 abdominal hysterectomy, and 64 laparoscopic hysterectomy. In the abdominal hysterectomy group, the operation started as a vaginal hysterectomy twice and as a laparoscopic hysterectomy in 7 patients. Patient characteristics and indications for the intervention are shown in Table II.

**Clinimetrics.** The overall response rate was 94%, ranging from 100% in the first days until 84% 12 weeks postoperatively. The percentage of missing items ranged per point of measurement from 0.2% to 1.9% in the QoR-40, from 0.1% to 0.7% in the RI-10, and 0.1% to 0.6% in the RAND-36. The Cronbach α at 1 week postoperatively was 0.93 for the QoR-40, 0.81 for the RI-10, and 0.86 for the RAND-36, showing good internal consistency.

---

**Table II.** Patient characteristics

<table>
<thead>
<tr>
<th>Total (n = 161)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (yr)</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
</tr>
<tr>
<td>Parity</td>
</tr>
<tr>
<td>ASA score</td>
</tr>
<tr>
<td>Employed women</td>
</tr>
<tr>
<td>Benign disease</td>
</tr>
<tr>
<td>Uterine weight (gr)</td>
</tr>
<tr>
<td>No complication</td>
</tr>
<tr>
<td>Hb decrease (g/dL)</td>
</tr>
<tr>
<td>Hospitalization (dy)</td>
</tr>
<tr>
<td>Readmissions</td>
</tr>
<tr>
<td>Return to theatre</td>
</tr>
</tbody>
</table>

ASA score, physical status classification of the American Society of Anesthesiologists; BMI, body mass index; Hb, hemoglobin; Hospitalization (dy), hospitalization including readmissions in days.

Data shown as mean ± standard deviation, median [range] or absolute numbers (percentage).

Table III shows the results of the assessment of construct validity as measured by testing 6, predefined hypotheses at 1 week after operation on all 3 instruments. Two hypotheses were confirmed in the QoR-40, 3 in the RI-10, and 1 in the RAND-36. In the hypothesis on difference in recovery among age groups, the QoR-40 and RI-10 showed unexpectedly a significantly better recovery in the older age group. A prolonged hospital stay and the occurrence of complications was significantly associated with a lesser mean score in all questionnaires, although this difference was only borderline significant (P = 0.07) in the hypothesis on complications in the RAND-36. The results of the construct validity testing at 4 weeks postoperatively were comparable with those at 1 week postoperatively, although a slight reduction in the performance of the QoR-40 and a slight improvement in the performance of the RAND-36 has been found (data not shown).

Correlation at 1 week after operation was high among the 2 recovery-specific instruments (r = 0.63), and moderate among each of the recovery-specific instruments and the RAND-36 (r = 0.37–0.38). Observed mean scores at baseline and observed and estimated mean scores in the postoperative period of the 161 patients are shown in Fig 1, A–C. No floor and ceiling effects in any of the 3 questionnaires were found. The range of the QoR-40 mean scores over time was 25% of the instruments total range, whereas the RI-10 scores ranged over 33% of the total range and the RAND-36 scores over 17% of the total range. Between 2 and 12 weeks after operation, the QoR-40 mean scores ranged over no more than 4% of
the total range of the instrument, which suggests that the QoR-40 was not responsive beyond 2 weeks postoperatively. The RI-10 showed an increase and thus responsiveness during the entire, 12-week postoperative period, whereas the mean score on the RAND-36 did not change in the first 2 weeks after operation and was not responsive in this time period.

The reliability, as a measure of precision, was somewhat greater in the QoR-40, compared with both the RI-10 and the RAND-36 (intraclass correlation of 56%, 42%, and 43% respectively).

Recovery after hysterectomy: results on the 3 instruments in the randomized subgroup. Fifty-nine patients with benign indications for operation had been randomized to either laparoscopic (n = 27) or abdominal hysterectomy (n = 32). There was a statistically significant treatment effect of 11.7 (3.9) units in the QoR-40 and 5.72 (1.45) units in the RI-10 both favoring laparoscopic hysterectomy, both P < .01 (Table IV). Figure 2, A–C, show the estimated mean scores with 95% confidence bands by treatment group.

**DISCUSSION**

To our knowledge, this is the first study comparing postoperative recovery measured by specific and generic instruments. The study was conducted to examine whether the application of a recovery-specific, quality-of-life instrument was more valid compared with a generic, health-related, quality-of-life instrument for the assessment of postoperative recovery. The objective of the study was to evaluate whether there is a need for recovery-specific questionnaires. The QoR-40, RI-10, and RAND-36 showed good internal consistency. The 3 instruments showed comparable construct validity with the best performance in the RI-10. Moreover, the RI-10 showed the best responsiveness with an increase during the entire 12-week postoperative period, as well as the greatest score range over time in relation to the instruments score range.

At the start of this research project in 1996, no valid, recovery-specific instrument was available to measure recovery from laparoscopy versus open surgery in gynecology. The number of items was an important issue in the development of the RI-10 in attempt to cause as little trouble as possible to patients in their recovery period. Consequently, this instrument may be advantageous because of its limited number of questions. The validation process has only been performed for the Dutch language, and for a single operative procedure (ie, hysterectomy), until now. Because the questions are of a general nature and not specific to gynecologic surgery patients, the questionnaire is likely to be valid for many other procedures as well.

A limitation of the questionnaire is that a baseline measurement is lacking, because most of the questions refer to the postoperative situation. Consequently, this instrument may be advantageous because of its limited number of questions. The validation process has only been performed for the Dutch language, and for a single operative procedure (ie, hysterectomy), until now. Because the questions are of a general nature and not specific to gynecologic surgery patients, the questionnaire is likely to be valid for many other procedures as well.

A limitation of the questionnaire is that a baseline measurement is lacking, because most of the questions refer to the postoperative situation. Furthermore, the target population has not been involved in the development, and no item reduction techniques were applied, as has been recommended. The QoR-40 has been shown to be a valid and reliable tool for measurement of recovery in a wide range of surgical patients. In a recent literature review on recovery-specific questionnaires, the QoR-40 appeared as one of the most valid instruments. In a cardiac surgical population, a poor
QoR-40 score on day 3 was correlated moderately to a poor quality of life as measured by SF-36 at 3 months after the operation. Although Leslie et al have shown that the questionnaire did not discriminate between the recovery of cranial and spinal surgical patients, the conclusion of their study was that the QoR-40 had been shown to be valid and reliable in a neurosurgical population as well. The mean score on day 1 in the adult cardiac surgical population as described by Myles et al and the neurosurgical population as described by Leslie et al was 163 and 160, respectively. Whereas baseline scores were comparable, the women undergoing hysterectomy in the present study had a lesser QoR-40 score, with a mean score of 143 on day 1. Unfortunately, the QoR-40 has not been translated in accordance with prevailing rules of back and forward translations, but the Cronbach α was comparable with those in the English version.

Although the RAND-36 is a more extensive instrument and was less responsive compared with the RI-10, the RAND-36 has the advantage of being a more widely known questionnaire with well-validated subscales. In this study, the total score of the instruments have been used for validation. Whether one of the 8 subscales of the RAND-36 has better clinimetric properties for the measurement of postoperative recovery has not been studied.

The Cronbach α has been calculated for the 3 instruments and was found satisfactory. In instruments focusing on evaluation of health status, however, responsiveness is more important than internal consistency. Because recovery in the first weeks after operation is expected to be a very dynamic process, we did not check for test-retest reliability in the present study. It has been stated that at least 75% of predefined hypotheses should hold true on statistical
testing to conclude that there is good validity.\textsuperscript{22}

The instrument coming closest at 1 week postoperatively was the RI-10, in which 3 out of 6 hypotheses (50\%) were confirmed. In the present study, we were looking for the instrument, in which most hypotheses would hold true compared with the others. This approach led to the formulation of challenging hypotheses, which were less likely to be confirmed in our population. This was probably the case for the difference between women with an ASA score of 1 and women with an ASA score of at least 2, where none of the questionnaires could identify a difference between the groups. Because there were only 6 patients with an ASA score of 3, we were testing for a difference between ASA scores of 1 and 2. From a clinical point of view, a difference in recovery among these groups is not very likely. Lack of contrast was probably also the case for the hypothesis on hemoglobin decrease, because these 2 groups do not seem to differ enough in relation to the sensitivity of the instruments used.

Although it has been previously shown that older patients have lesser quality-of-life scores and poorer recovery,\textsuperscript{12,14} we were not able to confirm this in our population. In the RI-10 and QoR-40, the patients over 60 years of age had a significantly better recovery compared with the women of less than 45 years of age. Moreover, compared with the groups as defined in other hypotheses, the patient group over 60 years of age had a remarkably high quality-of-life score. This trend was in the exact opposite direction as defined in the hypothesis. Possibly, this difference in recovery was caused by a difference in indications for operation, since 11 women (65\%) in the older age group were operated for endometrial carcinoma and another 5 women (29\%) for atypical hyperplasia of the endometrium, which is a potentially malignant condition. Women being treated for (potential) malignancy may expect a lesser quality of life and, as a result, may report a more optimistic assessment of their posttreatment quality of life. This phenomenon has been referred to as “reframing” and is thought to be part of the patients’ adaptation to malignancy.\textsuperscript{26} Another possible explanation is that the older patients have corrected their expectations for their age, and thus, they did not experience any mismatch between expectations and actual recovery.\textsuperscript{27}

Surprisingly, the hypothesis comparing vaginal and abdominal hysterectomy was only confirmed in the RI-10. Because it is accepted widely that vaginal hysterectomy is the first choice of approach to hysterectomy, we would have expected to find differences in the other instruments as well.\textsuperscript{19} The comparison of vaginal and laparoscopic hysterectomy had not been defined as a hypothesis prior to the study. In this comparison, however, we have found no differences in recovery on any of the 3 questionnaires between the women who had undergone vaginal and laparoscopic hysterectomy.

The items in the questionnaires used should correspond with the time period studied. Depending on the extensiveness of the operation and the time passed since the hysterectomy, there is a shift in the main topics regarding recovery. Concerns about, for instance, nausea are more prominent in the first days after operation, whereas return to work plays a role at the end of the recovery period. Regarding the 3 instruments that were used in this study, the application of the QoR-40 is more appropriate during hospitalization, whereas the RI-10 and the RAND-36 are more suitable in

### Table IV. Evaluation of a general, health-related and 2 recovery-specific questionnaires in randomized laparoscopic and abdominal hysterectomy patients, using a linear mixed model

<table>
<thead>
<tr>
<th>Score range</th>
<th>QoR-40</th>
<th>P</th>
<th>RI-10</th>
<th>P</th>
<th>RAND-36</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>40–200</td>
<td>0.01</td>
<td>0.21</td>
<td>0.01</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postoperative time period 1 day–2 weeks</td>
<td>1 week–12 weeks</td>
<td>2 weeks–12 weeks</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment effect\textsuperscript{a}</td>
<td>11.7 (3.9)</td>
<td>&lt;0.01</td>
<td>5.72 (1.45)</td>
<td>&lt;0.01</td>
<td>42.3 (26.1)</td>
<td>0.11</td>
</tr>
<tr>
<td>Baseline effect</td>
<td>0.1 (0.1)</td>
<td>0.21</td>
<td>—</td>
<td>0.2 (0.1)</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Time effect\textsuperscript{b}</td>
<td>10.8 (1.6)</td>
<td>&lt;0.01</td>
<td>1.35 (0.08)</td>
<td>&lt;0.01</td>
<td>13.8 (1.4)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>SDw</td>
<td>13.4</td>
<td>—</td>
<td>4.40</td>
<td>—</td>
<td>75.7</td>
<td>—</td>
</tr>
<tr>
<td>SDb</td>
<td>11.8</td>
<td>—</td>
<td>5.13</td>
<td>—</td>
<td>87.0</td>
<td>—</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Treatment effect favors LH.

\textsuperscript{b}In the model, different increments between the points of measurement were allowed, but for the ease of presentation, the average increase per week was presented.

\textsuperscript{b}Baseline effect, estimated increase in score (standard error) by unit increase in baseline score; P, P value for difference among groups using the test of Mann-Whitney; QoR-40, Quality of Recovery-40; RI-10, Recovery Index-10; RAND-36, RAND-36 health survey; SDb, standard deviation between patients; SDw, standard deviation within patients; Treatment effect, estimated increase in score (standard error) per week; Time effect, estimated increase in score (standard error) by treatment group; —, not applicable.
(recently) discharged patients. Furthermore, the 3 instruments refer to different time periods after operation, because the QoR-40 addresses the past 24 hours, the RI-10 refers to the past week, and the RAND-36 refers to the past 4 weeks. These differences may have had an influence on the comparability of the instruments and might also explain the absence of change in the first 2 weeks in the RAND-36 and the weak correlation between the recovery-specific instruments and the RAND-36 at 1 week postoperatively. Furthermore, we did not apply the RI-10 and RAND-36 in the first days after hysterectomy as we did for the QoR-40. The QoR-40 showed great changes in quality of life in these days. Whether these same changes would also have been the case for the RI-10 and RAND-36 is less likely, although not proven at this stage. In retrospect, it would have been more appropriate to distribute all 3 questionnaires in the first few days after hysterectomy and to use the “acute-version” of the RAND-36 referring to the past week. The pattern of changes in scores over time imply that it is appropriate to use the QoR-40 on a daily basis in the first few days after operation and stop its use 2 weeks postoperatively. Because the RAND-36 score does not change in the first 2 weeks, it is appropriate to use the RAND-36 beyond 2 weeks postoperatively. The RI-10 can be applied for at least 12 weeks postoperatively as a change in score was observed along this time period.

Use of these instruments in randomized patients showed a statistically significant treatment effect in favor of the laparoscopic hysterectomy in the 2 recovery-specific instruments. Because this analysis was an evaluation over time, our findings are another indication for better responsiveness of these instruments. The curve of the QoR-40 scores does not fit a linear or parabolic model, which

---

**Fig 2.** A–C, Estimated mean scores with 95% confidence bands by treatment group, using a linear mixed model. RAND-36 = RAND-36 health survey, QoR-40 = Quality of Recovery-40, RI-10 = Recovery Index-10, AH = abdominal hysterectomy, LH = laparoscopic hysterectomy. Thick broken line = line that joins the estimated mean scores in AH group, dotted line = 95% confidence bands. Thick solid line = line that joins the estimated mean scores in LH group, dotted line = 95% confidence bands.
makes interpretation of the results over time more difficult. The decrease in mean score at 1 week postoperatively might be caused by discharge from hospital, because the patients undergoing laparoscopic hysterectomy returned home at a mean of 4.2 days and the patients undergoing abdominal hysterectomy at a mean of 5.4 days postoperatively. Possibly, there are 2 separate QoR-40 curves, ie, 1 for hospitalized and 1 for discharged patients, which was not further evaluated in the present study.

Both generic, health-related, quality-of-life questionnaires and recovery-specific quality-of-life questionnaires can be used as instruments for assessment of postoperative recovery. The generic, health-related instruments, such as the RAND-36, have the advantage of being used more frequently compared with recovery-specific instruments. Thus, more data from other studies will be available for comparison. Recovery-specific instruments, however, are more valid for measurement of recovery compared with the RAND-36. The dilemma here is either more comparability or more validity, which can be solved easily by administering a combination of the 2 types of instruments in future studies.

CONCLUSION

In conclusion, the QoR-40, RI-10, and RAND-36 showed comparable internal consistency. With regard to construct validity, the RI-10 seemed to perform best and was more responsive compared with the QoR-40 and RAND-36. Therefore, we recommend the use of the short recovery-specific questionnaire RI-10 in studies on postoperative recovery. In studies where the main interest is in hospitalized patients in their first postoperative days, the QoR-40 has likewise shown to be valid.

REFERENCES


28. Keller SD, Bayliss MS, Ware JE Jr, Hsu MA, Damiano AM, Goss TF. Comparison of responses to SF-36 Health Survey questions with one-week and four-week recall periods. Health Serv Res 1997;32:367-84.

APPENDIX 1. RECOVERY INDEX-10 (RI-10) IN DUTCH AND ENGLISH VERSION

1. Ik ben snel moe
   I get tired rapidly
2. Overdag moet ik regelmatig rusten
   I need to rest regularly during daytime
3. Ook als ik niets doe heb ik regelmatig last van buikpijn
   Even when I don’t do anything, I regularly have abdominal pains
4. Zelfs lichte inspanning (bijvoorbeeld koffie zetten) kan ik nauwelijks doen
   I can hardly perform even the least efforts (such as making coffee).
5. Ik voel me volledig hersteld na de operatie
   I have fully recovered from the operation
6. Ik kan mijn normale dagelijkse bezigheden in huis helemaal doen
   I can fully perform my daily activities at home
7. Sinds de operatie heb ik moeite met slapen
   Since the operation I have difficulty sleeping
8. De operatie en het herstel daarna zijn minder goed verlopen dan ik mij had voorgesteld
   The operation and the recovery afterwards went less smoothly than I had expected
9. Ik heb in het algemeen veel pijn gehad na de operatie
   Generally speaking I suffered a lot of pain after the operation
10. De klachten waarvoor ik ben geopereerd zijn volledig verdwenen
    The complaints, leading to surgery, have completely disappeared