

Research Article

Open Access

Development and Validation of the Post-Operative Recovery Index for Measuring Quality of Recovery after Surgery

Stephen F. Butler^{1*}, Ryan A. Black¹, Lee Techner², Kathrine C. Fernandez¹, David Brooks³, Mollie Wood¹ and Nathaniel Katz^{4,5}

¹Inflexxion, Inc., Newton, USA

²Cubist Pharmaceuticals, Lexington, USA

³Brigham and Women's Hospital, Boston, USA

⁴Analgesic Solutions, Natick, USA

⁵Tufts University School of Medicine, Boston, USA

Abstract

Purpose: Current methods used to quantify aspects of recovery after surgery and anesthesia tend to be narrowly focused, not patient-rated, or have not been appropriately validated. We set out to develop a quality of recovery score system that is self-report and multi-dimensional, with applicability across various surgeries and surgical settings, from immediately post-surgery through discharge and covering the first 30 days of recovery.

Methods: A Post-operative Recovery Index (PoRI) was validated on 225 patients ($N_{\text{Validation}}=96$; $N_{\text{Cross Validation}}=129$) who had undergone a surgical procedure within the last 30 days. Domain level internal consistency on the validation and cross validation samples yielded coefficients ranging from $\alpha=0.813$ to $\alpha=0.932$, while test-retest reliability yielded stability coefficients ranging from $r=0.660$ to $r=0.881$. Confirmatory factor analyses demonstrated validity of the factorial structure of the 37-item PoRI on the validation patient sample and confirmed on the cross validation patient sample. Exploratory psychometric analyses provided evidence of an overarching (second-order) "Recovery" factor.

Results: We developed, tested, validated, and cross validated the Post-operative Recovery Index (PoRI) consisting of 37 items assessing symptomatology a patient may experience after surgery.

Conclusion: The PoRI is offered as a valid, multidimensional measure of recovery after surgery and anesthesia with broad applicability in post-surgical settings.

Keywords: Symptom measurement; Post-operative recovery; Quality of life

Introduction

Optimizing the care and assisting in the recovery of patients after surgery as well as reduction of hospital stay time require the measurement of appropriate outcomes and symptoms through evaluation systems that quantify patients' post-operative health status. While a number of evaluation systems have been developed to assess a diversity of outcomes [1-5] there is an unmet need for a practical, patient self-report, comprehensive measure of recovery following surgery. Evaluation of this type of scale would need to include the following criteria [6]: appropriateness, reliability, validity, sensitivity to change, precision, interpretability, acceptability and feasibility. A review [6] of existing post-operative recovery scales shows that only one instrument, the 40-item Quality of Recovery (QoR-40) [7] meets the eight criteria; However, this scale is validated only for the immediate (within one day) postoperative period. Several studies have attempted to measure specific symptoms and signs after specific types of surgery [8-12]; however, few have attempted to develop a brief, reliable, multidimensional self-report measure for assessing post-operative recovery applicable to a broad range of surgeries and that can accommodate the entire post-operative course, from a few hours following surgery to complete recovery. One existing instrument is the 27-item Convalescence And Recovery Evaluation (CARE) tool, which assesses four domains identified through factor analysis (pain, gastrointestinal, cognitive, and physical activity) [13]. Although this instrument was suitable for immediate post-operative assessment with high test/re-test reliability and moderate to high internal consistency for all domains, it was developed on a small sample of patients ($n=96$) and it assessed recovery after only three different types of surgery (general, urology-related, and gynecology-related).

In the work presented here, we aimed to develop and validate a

practical and psychometrically sound assessment, the Post-operative Recovery Index (PoRI), to measure the quality of immediate post-surgical recovery and during the early recovery period (i.e., approximately 30 days following surgery) in patients undergoing a broad range of surgical procedures including open laparotomy, spinal fusion, total-knee replacement, full thoracotomy, and laparoscopic-colon resection. The goals in this work were to create a PoRI that: (1) uses a self-report format, (2) provides an assessment of the quality of recovery in relevant dimensions, (3) is brief enough for use in typical clinical and research settings, (4) is easy to administer and score, and (5) is psychometrically sound. These goals were addressed following a systematic approach consisting of five sequential studies. Study 1 established an initial pool of post-operative outcome items and content validity of the scale. Study 2 completed a conceptual evaluation of the initial item pool and initial item reduction, resulting in the creation of an *alpha* version of the PoRI. The *alpha* version was empirically evaluated in Study 3 producing a final version which was cross validated in Study 4. Finally in Study 5, an exploratory analysis examined the factor structure of the scale to determine the possible presence of a second-order, overall "Recovery" factor (total scale), and a scoring

***Corresponding author:** Stephen F. Butler, Inflexxion, Inc., 320 Needham St., Suite 100, Newton, MA 02464, Tel: 617-332-6028; Fax: 603-672-3162; E-mail: sfbutler@inflexxion.com

Received November 28, 2012; **Accepted** December 27, 2012; **Published** January 02, 2012

Citation: Butler SF, Black RA, Techner L, Fernandez KC, Brooks D, et al. (2012) Development and Validation of the Post-Operative Recovery Index for Measuring Quality of Recovery after Surgery. J Anesth Clin Res 3:267. doi:10.4172/2155-6148.1000267

Copyright: © 2012 Butler SF, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

system for the PoRI was proposed. The methods and results for each study are presented.

Study 1: Conceptual Development, Content Validation and Item Pool Construction

An initial step in developing a clinical scale is to establish content validity [14,15], which generally requires establishing a consensus among various content-knowledgeable sources such as literature reviews and subject-matter experts. The FDA's guidance on patient reported outcomes [16] emphasizes the necessity, of including patient perspectives in this process.

Methods

Participants: Two groups of stakeholders participated in this stage: health care professionals and patients. Professionals included physicians (surgeons and anesthesiologists), surgical nurses and pharmaceutical industry researchers in the area of gastroenterology. Health care professionals were identified by recommendations from colleagues and through Internet postings on professional listservs and message boards. Pharmaceutical industry representatives were identified through pharmaceutical company contacts. Patient participants were recruited via postings in hospitals, surgical clinics, patient listservs, and electronic classified advertisements. Inclusion criteria for patients were having undergone any of the following surgery types within the past 30 days: Open laparotomy, spinal fusion, total knee replacement, full thoracotomy, or laparoscopic colon resection. Exclusion criteria for patient recruitment included having active systemic disease that could exhibit symptoms similar to opioid treatment side effects, current major psychiatric disorder, pregnancy, or an inability to complete study questionnaires accurately. All participants provided informed consent [1].

Design: In order to establish the content validity of the scale, we applied concept mapping [17] which is a systematic method for collecting and establishing stakeholder input that has been used to establish content validity for other purposes [18]. The concept mapping consisted of three steps: (1) brainstorming a list of items proposed by stakeholders, (2) ratings of each item by individual participants as to the item's importance for post-operative recovery, and (3) grouping of items into conceptual themes.

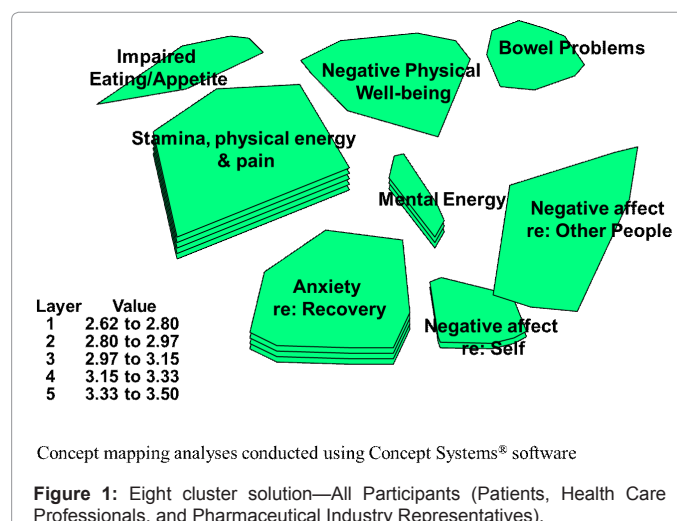
After consent, participants were emailed or mailed materials and instructions to complete concept mapping activities. Participants brainstormed statements in response to several focus prompts. Patient participants responded to prompts asking about their recovery experience after surgery, while professional participants responded to prompts capturing common post-operative symptoms and patient concerns. A list of unique statements was re-presented to participants for rating of importance of items and sorting into conceptually consistent clusters.

Average ratings for each item, statistical summaries of the groupings, and pictorial representations showing relationships between and among individual items and groups were produced. Statistical processes used in summarizing the group data included multidimensional scaling and hierarchical cluster analysis. Analyses and concept maps were performed by concept mapping software (Concept System® software: Copyright 2004-2012; Concept Systems Inc.) to derive "clusters."

Results

Participants: The professional group included seven physicians, five nurses and eight pharmaceutical industry researchers. Of 162 patients that responded to postings and were screened, 97 met the selection criteria. Of these, 18 were inpatients and 79 were outpatients. Average age was 41.6 years (SD=13.9), 67% of respondents were Caucasian, 17% were African American, 10% were Hispanic, 4% were Asian, and 2% identified themselves as "Other". All patient participants reported at least a high school education.

Content validity: The brainstorming stage resulted in an initial list with 1,710 statements. Duplicate, nonsensical, and illegible statements were removed. Other statements were deleted because they were general mood statements that made no reference to recovery (i.e., "distracted", "shock"). Statements that reflected specific attributes of the hospital stay (i.e., factors external to the patient) were also deleted (i.e., "didn't feel safe", "noise from all the alarms"). From this process, a list of 191 unique statements was developed. These statements were presented to stakeholders for sorting and rating, and responses were reviewed again by the research team. Statements that either (a) had low importance ratings based on stakeholder input or (b) were too general to convert into a scale item were deleted. This resulted in 134 unique potential items that were clustered using concept mapping software into the following eight groups: 1) Impaired Eating/Appetite; 2) Negative Physical Well-being; 3) Bowel Problems; 4) Stamina, Physical Energy, and Pain; 5) Mental Energy; 6) Negative Affect Regarding Other People; 7) Anxiety Regarding Recovery; and 8) Negative Affect Regarding Self. Figure 1 presents the concept mapping cluster map. The map presents each cluster as having from one to five layers that represent the average rating of the statements included in the cluster. The legend presents the value range included in each layer. Thus, single-layered clusters contain statements that were rated, on average, as least important with averages from 2.62 to 2.80. Conversely, clusters with five layers contain statements rated on average as most important, with averages from 3.33 to the maximum average rating of 3.50 (out of a possible 5). Note that the size of the cluster is a visual representation of the extent to which the items in a given cluster were sorted together. This means that the smaller the area of a cluster, the more often participants sorted these statements together. Conceptually, a small area suggests that the statements reflect a similar, probably well-defined underlying concept, whereas a larger area suggests a broader, perhaps less well-defined concept.



Study 2: Conceptual Evaluation and Item Reduction Via Sorting and Rating

The purpose of Stage 2 was to finalize items for the alpha scale. The item-pool underwent a second round of sorting and rating with a new group of stakeholders. This helped determine the wording of items and then determine if reworded statements fall into the same clusters as the original statements as in Stage 1. The quality of the items and answer options for the scale was established in this stage of the scale development.

Methods

Participants: New groups of professional and patient stakeholders were recruited using the same inclusion/exclusion criteria as Study1.

Design: The 134 statements selected in Study 1 were reworded into question form. Question-items were presented to participants for sorting and rating. Participants completed the concept mapping procedures described above.

Content analysis: Concept mapping software generated conceptual groupings, and participant ratings of item importance were reviewed. Items with low value ratings were deleted. Some items were retained with low ratings from patients when ratings from professional stakeholders and literature review suggested these items tapped critical aspects of postoperative treatment. Answer options for the scale were also established at this stage. This process resulted in an *alpha* version of the scale.

Results

Participants: Seventy patients responded to the recruitment advertisements, and 53 met the inclusion criteria. Of these, 48 completed the rating and sorting tasks. Among the 48 patients, four were inpatients and 44 were outpatients; 20 participants (42%) were males. The average age was 41.6 years (SD=13.4), 60% were Caucasian, 19% were African American, 4% were Hispanic, 2% were Asian, and 4% were “Other”. Most patient participants had some college education (range 11th grade through four years of graduate school). The professional participant group included five physicians, three nurses and three physician assistants.

Concept mapping confirmation and item reduction: A second concept map was created which yielded clusters and ratings similar to those obtained during Study 1. Again, items rated low were deleted except when input from professional stakeholders suggested

importance from their perspective. Review of the conceptual evaluation of items resulted in a 118-item *alpha* version of the scale.

Study 3: Empirical Evaluation of the PoRI alpha version

In this stage, an empirical evaluation of items was used to create a final version of the PoRI and to provide an initial test of the psychometrics of the scale.

Methods

Participants: Patient participants were recruited who fulfilled the following selection inclusion criteria: (1) ≥18 years old, (2) underwent open laparotomy, spinal fusion, total knee replacement, full thoracotomy, or laparoscopic colon resection within the past 30 days, (3) were prescribed opioid medication for postoperative pain, and (4) were willing to give informed consent. Exclusion criteria were: (1) active systemic disease that exhibits symptoms similar to side effects associated with opioid treatment, (2) current major psychiatric disorder, (3) pregnancy, and (4) inability to complete study queries accurately.

Design: Empirical evaluation of the *alpha* version of the PoRI consisted of examining: (1) item distributions, (2) the hierarchical structure of the scale (items of similar content clustering within sub-domains and sub-domains of similar content clustering within domains), (3) test-retest reliability, and (4) construct validity (domains should be more associated with comparison measures of the same construct than comparison measures of dissimilar constructs).

Inpatient participants completed the battery within three days post-surgery depending on when they could complete the task as determined by clinical staff. Two groups of outpatients completed the battery sent to them in the mail either 15 days after surgery or between 15 and 30 days after surgery. To assess item test-retest reliability, fifty subjects were randomly selected to retake the *alpha* version of PoRI three days after the first administration.

Analysis: Statistical analysis involved examining (1) item level variability via standard deviations, (2) internal consistency and test-retest reliability of domains and subdomains via Cronbach’s α and Pearson correlations, (3) construct validity, both convergent and discriminant, by estimating Pearson correlations between PoRI domains and comparison measures, and (4) factorial validity by employing Confirmatory Factor Analytic (CFA) techniques. All analyses were carried out using SPSS 19 and AMOS 19.

Comparison Scales	Number of items in scale	Outcome Assessed	Corresponding Items in PoRI
Oswestry Disability Scale [22]	10	physical functioning and ability to manage everyday life	physical functioning items
Patient Assessment of Constipation-Quality of Life (PAC-QOL) [23]	28	the impact of constipation symptoms	bowel functioning items
Patient Assessment of Constipation-Symptoms (PAC-SYM) [24]	12	assessing rectal symptoms, stool symptoms, and abdominal discomfort	bowel functioning items
Brief Pain Inventory (BPI) [25]	32	chronic non-malignant pain	pain items
Opioid Side Effects Checklist [26]	20	symptoms associated with opioid therapy	opioid side effect items
Symptom Distress Scale [27]		emotional/psychological symptoms	emotional subscale items
Rotterdam Symptom Checklist [28]	23	physical symptom distress, psychological distress, activity level, and overall global life quality	activity levels, miscellaneous symptoms, and psychological distress items
Brief Fatigue Inventory [29]	10	fatigue severity	physical functioning and general symptom items
Health Status Questionnaire [30]	36	comprehensive health questionnaire	overall health state items

Table 1: Comparison measures used to assess domains in the PoRI.

Measures: In addition to a demographics questionnaire, a battery of comparison measures was created. Comparison measures were selected on the basis of evidence that they provide a measure of the domain in question, and that they have the best-documented psychometrics among alternative measures of the domains. The comparison measures, outcomes targeted and corresponding candidate PoRI items are presented in Table 1.

Results

Participants: One-hundred-eighty-two (182) patients volunteered to participate and, of these, 106 were enrolled in the study and 96 participants completed all measures. Of these 96 participants, 62% were female; 84% were Caucasian, 12% were Black, 3% were Hispanic, and 1% were Asian. Most patients (83%) reported some college. Average age was 46.1 years (SD=14.4). Seventeen participants (N=17, 18%) were inpatients and 79 (82%) were outpatients; 24% had undergone laparoscopic surgery, 20% knee replacement, 21% spine fusion, 19% abdominal surgery, 14% chest surgery, and 3% other.

Item distribution: Although all items produced ratings that covered the entire five-point scale, several items had standard deviations less than half a scale point. This was used as a cut off to ensure that retained items evidenced a sufficient distribution of responses across the five-point scale.

Initial item reduction analyses: Due to the relatively small patient sample size (N=106), it was not feasible to evaluate the factorial validity of the scale using 118 items. The research team examined the zero-order correlations between individual items and the comparison measures. Items that correlated significantly and positively with their target comparison measure and more highly with the target comparison scale than the other scales were retained. A reduced set of 37 final items were retained for the beta version of the PoRI.

Domain and sub-domain creation: As part of the item reduction process, items were grouped into one of the following five domains: Psychological Symptoms, Physical Activities, General Symptoms, Bowel Symptoms, and Appetite Symptoms. Within each domain, similar items were further grouped into sub-domains. Table 2 presents the items placed within their respective domains and sub-domains. These groupings were evaluated empirically next.

Internal consistency: Internal consistency was obtained for each domain, yielding α coefficients ranging from 0.873 to 0.932 (Table 3). Cronbach α was also computed for similar items grouped within each domain (sub-domains). These coefficients were also within the acceptable range (Range: 0.634 to 0.933, Table 3).

Test-retest reliability: Domain level test-retest reliability over a 3-day window was obtained for a 43 of 50 recruited participants (86%), yielding stability coefficients that ranged from 0.700 to 0.881 (Table 3).

Construct validity: Construct validity is composed of both convergent and discriminant validity [19]. Convergent and discriminant validity were established by correlating each of the domains with the comparison measures. As seen in the correlation matrix in Table 4, correlations between the domains and the other comparison measures (i.e., average on-diagonal correlation=0.74, SD=0.11) was significantly greater than the correlations with dissimilar constructs (off-diagonal correlation=0.45, SD=15; $t=5.8$, $df=48$, $p<0.001$).

Factorial validity: To test factorial validity of the model, a first-order CFA was conducted. The CFA consisted of the 10 manifest variables ("sub-domains") and five first-order factors ("domains"), as

Domain	Sub-Domain	Item
Psychological Symptoms ^c	Internal	(1) Lack of energy, fatigue, tiredness (2) Feeling out of control (3) Feeling aggravated or irritable (4) Feeling discouraged (5) Worried that won't fully recover from surgery (6) Motivation is low
	Interpersonal	(7) Feeling like a burden to others (8) Feeling aggravated or irritable
Physical Activities ^e	Basic	(9) Ability to sit up (10) Ability to stand up (11) Getting self to the bathroom (12) Dressing self
	Advanced	(13) Doing your day-to-day activities
		(14) Walking several blocks (15) Walking up a flight of stairs (16) Driving
General Symptoms ^a	Physical/ Neuropsychological	(17) Lack of physical strength (18) Felt clumsy or uncoordinated (19) Avoid strenuous activity (20) Trouble focusing on mental tasks (21) Noticed slurred speech
	Sleep	(22) Trouble staying awake during the day (23/24) Trouble falling or staying asleep
Bowel Symptoms ^c	Lower	(25) Feeling constipated (26) Straining during bowel movements (27) Gas pains (28) Problem passing gas (29) Bowel movements seem incomplete (30) Feeling rectal pressure or fullness (31) Bowel movements are unsatisfying
	Upper	(32) Felt nauseated (33) Belly feels bloated
Appetite Symptoms ^c	Pleasure	(34) Not able to enjoy favorite foods (35) Nothing tastes good (food or drink)
	Digestion	(36) Only able to eat small amounts of food at one time (37) Poor appetite

Note: Based on CFA, Upper Bowel sub-domain is associated with Bowel and Appetite domains.

^cResponse options for items within these domains range from "Not at all bothered" to "Extremely bothered."

^eResponse options for items within this domain range from "Not at all" to "Extremely limited."

^aResponse options for items within this domain range from "None" to "Extremely."

Table 2: PoRI Domains, Sub-Domains and Items.

shown in Table 5. Results from the CFA revealed a good fit for the model, $\chi^2=37.441$, $p=0.052$, CFI=0.981, RMSEA=0.072 [20,21]. Figure 2 presents the graphical depiction of the model, along with standardized estimates.

Study4: Cross Validation of the Final Version of PoRI

An often neglected step in scale development is the cross validation [14,15]. Cross validation ensures that the validity coefficients are retained on a new sample of respondents.

Methods

Participants: A new group of participants were recruited according to the same inclusion and exclusion criteria as used for Study 3.

Procedures: Participants completed the *beta* version of the PoRI and the same comparison measures as in the initial validation described in Study3. Fifty participants were randomly selected from the total sample to retake the measures three days later.

Analysis: Statistical analysis involved repeating the analytical

Scale, Domain, and Sub-Domain	Stage 3		Stage 4	
	Internal Consistency (Cronbach α)	Test-Retest (Pearson r)	Internal Consistency (Cronbach α)	Test-Retest (Pearson r)
Scale	0.955	0.782	0.948	0.793
Psychological Symptoms	0.932	0.700	0.927	0.767
Internal	0.917		0.913	
Interpersonal	0.775		0.762	
Physical Activities	0.931	0.881	0.916	0.757
Basic	0.907		0.896	
Advanced	0.872		0.897	
General Symptoms	0.873	0.738	0.813	0.660
Physical/Neuropsychological				
Symptoms	0.818		0.819	
Sleep	0.817		0.438	
Bowel Symptoms	0.932	0.742	0.920	0.814
Lower	0.933		0.923	
Upper	0.634		0.463	
Appetite Symptoms	0.896	0.719	0.898	0.790
Pleasure	0.802		0.797	
Digestion	0.799		0.812	

Table 3: Internal Consistency and Test-Retest Reliability Coefficients.

Domains	Comparison Measures									
	Psych		Physical		General		Bowel		Appetite	
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Psych Symptoms	0.787	0.749	0.455	0.582	0.738	0.704	0.525	0.585	0.439	0.370
Physical Activities	0.270	0.547	0.861	0.806	0.484	0.487	0.266	0.359	0.257	0.199
General Symptoms	0.667	0.760	0.450	0.636	0.774	0.773	0.543	0.567	0.449	0.394
Bowel Symptoms	0.503	0.497	0.355	0.420	0.637	0.594	0.783	0.748	0.385	0.363
Appetite Symptoms	0.276	0.445	0.245	0.310	0.359	0.312	0.248	0.254	0.531	0.566

Note: C1=Rotterdam Symptom Checklist psychological subscale; C2=Brief Fatigue Inventory; C3=Rotterdam Symptom Checklist Activity Subscale; C4=Oswestry Disability Scale; C5=Rotterdam Symptom Checklist Miscellaneous Subscale; C6=Opioid Side Effects Checklist; C7=PAC-SYM; C8=PAC-QOL; C9=PAC-QOL-Item 8 (Decreased Appetite); C10=Rotterdam Symptom Checklist Item 1 (Lack of Appetite)

Table 4: Validation Study (Stage 3) Correlations (Pearson r) between Domains and Comparison Measures.

Domains	Comparison Measures									
	Psych		Physical		General		Bowel		Appetite	
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Psych Symptoms	0.697	0.771	0.478	0.510	0.669	0.751	0.502	0.612	0.296	0.368
Physical Activities	0.156	0.503	0.831	0.649	0.346	0.353	0.138	0.261	0.191	0.224
General Symptoms	0.513	0.677	0.606	0.540	0.644	0.676	0.454	0.544	0.414	0.434
Bowel Symptoms	0.463	0.553	0.256	0.250	0.686	0.641	0.828	0.822	0.407	0.431
Appetite Symptoms	0.295	0.442	0.410	0.287	0.502	0.443	0.321	0.413	0.474	0.643

Note: C1=Rotterdam Symptom Checklist Psychological Subscale; C2=Brief Fatigue Inventory; C3=Rotterdam Symptom Checklist Activity Subscale; C4=Oswestry Disability Scale; C5=Rotterdam Symptom Checklist miscellaneous subscale; C6=Opioid Side Effects Checklist; C7=PAC-SYM; C8=PAC-QOL; C9=PAC-QOL-Item 8 (Decreased Appetite); C10=Rotterdam Symptom Checklist Item 1 (Lack of Appetite).

Table 5: Cross-Validation Study (Stage 4) Correlations (Pearson r) between Domains and Comparison Measures.

approach performed in Study3 on a new sample of participants.

Results

Participants: Two hundred individuals volunteered to participate, and a total of 132 met inclusion criteria, of which 129 had no missing data. This sample included 32 inpatients (25%) and 97 outpatients (75%). Mean age of patients was 48.2 years (SD=15.9); 34% were male with 85% Caucasian, 9% Hispanic, 5% African-American, 1% Asian, and 1% Native American. Most (73%) reported some post-secondary education. Thirty-four percent had undergone laparoscopic surgery, 28% knee replacement, 12% spine fusion, 9% abdominal surgery, 8% chest surgery, and 9% other.

Internal consistency: Internal consistency was obtained for each domain on the cross validation sample, yielding a coefficients ranging from 0.813 to 0.927 (Table 3). Cronbach α coefficients were also computed for sub-domains. These coefficients were also within the acceptable range (Range: 0.797 to 0.933, Table 3), except for the sub-domains, "Upper Bowel" (α =0.463) and "Sleep" (α =0.438). The low Cronbach alphas for these sub-domains were not surprising since these sub-domains had only two items each, which adversely impacts the α coefficient [19].

Test-retest reliability: As in Study 3, domain level test-retest reliability over a 3-day period was obtained for a subsample of 64 participants, yielding stability coefficients which ranged from 0.660 to 0.814 (Table 3).

Construct validity: Convergent and discriminant validity was

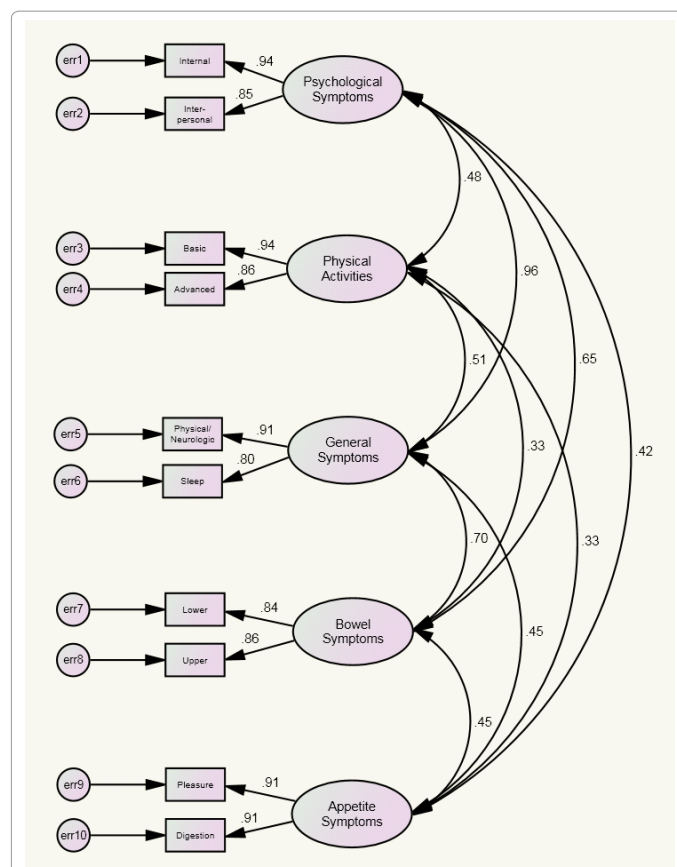


Figure 2: Graphical Depiction of PoRI CFA Model with Standardized Estimates on the Validation Sample.

similar to that observed in Study 3. Specifically, the average on-diagonal correlation was 0.70 (SD=0.11) was significantly greater than the average off-diagonal correlation (0.43, SD=15; $t=5.2$, $df=48$, $p<0.001$).

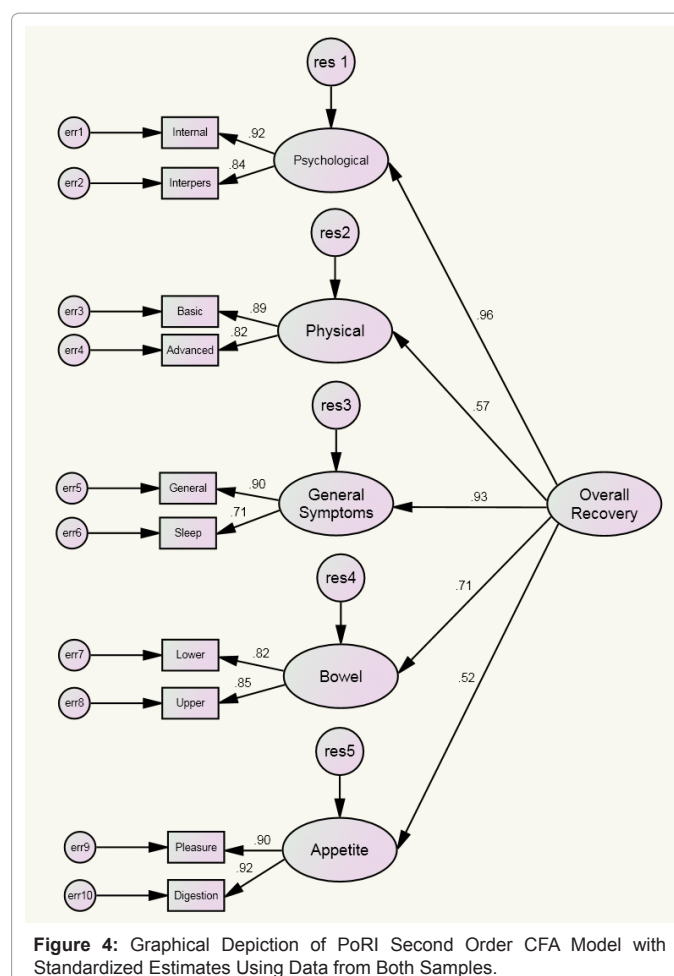
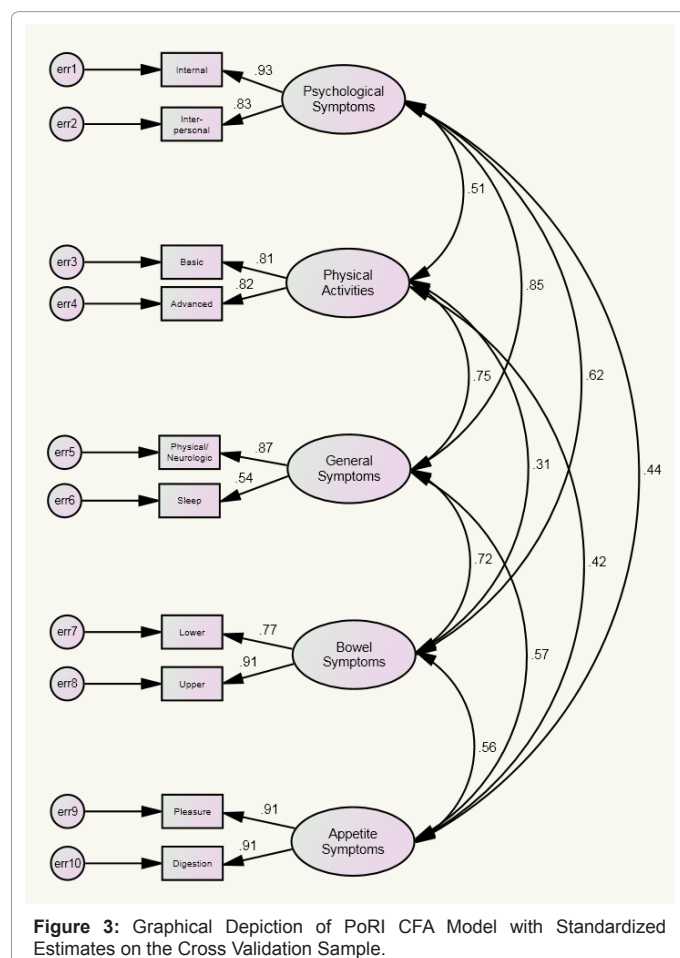
Factorial validity: To confirm the factorial validity of the model, the same first-order CFA performed in Study 3 was performed on the cross validation sample. As before, the CFA consisted of 10 manifest variables ("sub-domains") and five first-order factors ("domains"). Results from the CFA revealed a good fit for the model, $\chi^2=33.672$, $p=0.115$, CFI=0.988, RMSEA=0.052. Figure 3 shows the graphical depiction of the model, along with standardized estimates.

Study5: Exploratory Analyses and Proposed Scoring System

We hypothesized that there was an overarching (second-order) factor that accounted for the five (first-order) factors identified previously. In addition, in this section we describe scoring procedures for the PoRI.

Methods

Design: In order to explore this hypothesis, a second-order factor analysis was conducted on the entire data set (i.e., validation and cross validation data combined). Both data sets were combined given the added complexity of the second-order factor analytic model. Internal consistency and test-retest reliability associated with the Recovery factor (total scale) were examined separately for the validation and cross validation samples. A second-order factor analysis was conducted



to explore the factorial validity of incorporating an overarching Recovery factor into the first-order factor structure proposed in the previous stages.

Results

Second-order CFA: Results from the second-order CFA suggested adequate fit for the model, $\chi^2=81.440$, $p<0.001$, CFI=0.963, RMSEA=0.085. In order to achieve convergence, an equality constraint was placed on the residual variances associated with psychological and general symptoms factors [20]. See Figure 4 for the graphical depiction of the model, along with standardized estimates. Internal consistency and test-retest reliability of the total scale (overarching Recovery factor) conducted on each sample separately were deemed adequate (Table 3).

Subscale and scale scoring: Given the relatively strong psychometric properties of the PoRI, we have included a scoring system. We recommend scoring the subscales (a.k.a. first-order factors) by taking the arithmetic mean of all items that make up each subscale and scoring the scale (a.k.a. second-order factor) by taking the arithmetic mean of all 37 items. Higher scores reflect greater difficulty in post-operative recovery. Based on the item response options, we offer a scoring system for the scale and each subscale (Table 6).

Discussion

This article describes the development, validation and cross validation of the Postoperative Recovery Index (PoRI) an easy-to-use,

No Difficulty	1
Little Difficulty	>1 to <1.5
Moderate Difficulty	1.5 to <2.5
Considerable Difficulty	2.5 to <3.5
Extreme Difficulty	3.5 to 5

Table 6: PoRI Scale and Subscale Scoring System.

self-report measure of postoperative recovery that is clinically relevant and psychometrically sound. State-of-the-art scale development process included concept mapping to systematically establish content validity of the scale by integrating input of various stakeholders, including patients, providers, and pharmaceutical company researchers in the area of gastroenterology. A careful step-wise process involved item generation, item reduction, empirical testing and cross validation yielding a scale that may prove useful in evaluating the impact of interventions, medications, and procedural changes on the experience of patients following surgery.

Goals for the PoRI ensured that the scale would measure the quality of immediate post-surgical recovery and during the early recovery period (i.e., approximately 30 days following surgery) in patients undergoing a broad range of surgical procedures including open laparotomy, spinal fusion, total-knee replacement, full thoracotomy, and laparoscopic-colon resection. Further, the PoRI was intended to be multidimensional, capturing and scaling the quality of recovery in relevant dimensions. Measured domains and sub-domains were derived during the content validity stage and refined early in the empirical evaluation of items. Specifically, the scale captures psychological symptoms (internal symptoms and interpersonal concerns), physical activities (basic and advanced), general symptoms (physical/neuropsychological and sleep), bowel symptoms (upper and lower), and appetite (pleasure and digestion concerns). To be practical for use clinical and research settings, the PoRI was self-report, brief, and easy to administer and score. Careful attention was paid to the psychometric properties, including a complete cross validation to ensure that shrinkage of validity coefficients was minimal. Finally, an exploratory analysis examined the factor structure of the scale to determine the possible presence of a second-order, overall "Recovery" factor (total scale), and a scoring system for the PoRI was proposed.

There are several limitations to this study. First, the sample sizes were relatively small for the purposes of scale development and validation [21], and a number of surgery types were not included (e.g., cardiovascular, gynecology). Nevertheless, it should be noted that reliability, both internal and test-retest, and factorial validity were demonstrated when examined on the data obtained from the validation sample, and importantly, similar reliability and validity estimates were observed on the cross validation sample. In a similar vein, since the number of patients who had undergone any particular surgery widely varied, it was not feasible to test the factorial structure for each surgery type separately. Although the samples used in the conceptual and empirical studies presented here included representation of gender and race (i.e., Caucasian, African-American, Hispanic and Asian), subsample sizes were too small to test factorial equivalence (e.g., measurement or structural invariance [20]) of the scores from the PoRI across demographic groups. Nor was the PoRI tested in other cultures (i.e., countries outside the U.S.) We also did not assess the degree to which the scale is capable of detecting small, yet clinically meaningful changes in health status during the initial stages of recovery or how well the scale can predict health status and overall quality of life several months post-surgery. In light of these limitations, we recommend that further research be conducted to examine (1) the replicability of the

current findings by surgery type and specific populations (e.g., gender/race/ethnicity and countries outside the U.S.) and (2) the scale's short and long-term predictive validity of key criterion measures (e.g., quality of life).

Given the strong psychometric properties of the scores from the PoRI, it is recommended that the scoring system presented in Table 6 be used in current clinical practice in the United States. Still, as with any measuring instrument, scores should be interpreted within the context of a full clinical evaluation. Moreover, this is the first validation study conducted on the scores from the PoRI, and as a result, the authors recommend that future studies assess the replicability of the findings observed in the current study in both research and clinical settings.

In conclusion, the PoRI is offered as a valid, multidimensional measure of recovery after surgery and anesthesia with broad applicability in post-surgical settings. It is anticipated that the PoRI will be a useful tool for the empirical evaluation of surgical interventions, procedures and medications. The scale is currently being employed in several clinical trials. Use in clinical practice will be fostered by presentation of studies at professional meetings and other, relevant organizations.

Acknowledgement

The authors wish to thank Tom Thornhill for his assistance with the development of this scale, and Mirella Zulueta who assisted in editing and proofing the manuscript. Preparation of this article was supported in part by Inflexion, Inc., Newton, MA and Adolor Corporation (currently Cubist Pharmaceuticals, Lexington, MA 02421). The authors had sole editorial rights over the manuscript.

References

- Dexter F, Aker J, Wright WA (1997) Development of a measure of patient satisfaction with monitored anesthesia care. *Anesthesiology* 87:865-873.
- Myles PS, Hunt JO, Nightingale CE, Fletcher H, Beh T, et al. (1999) Development and psychometric testing of a quality of recovery score after general anesthesia and surgery in adults. *Anesth Analg* 88:83-90.
- Kluivers KB, Hendriks JC, Mol BW, Bongers MY, Vierhout ME, et al. (2008) Clinimetric properties of 3 instruments measuring postoperative recovery in a gynecologic surgical population. *Surgery* 144:12-21.
- Chung F, Chan VW, Ong D (1995) A post-anesthetic discharge scoring system for home readiness after ambulatory surgery. *J Clin Anesth* 7: 500-506.
- Royse CF, Newman S, Chung F, Stygall J, McKay RE, et al. (2012) Development and feasibility of a scale to assess postoperative recovery: the post-operative quality recovery scale. *Anesthesiology* 113: 892-905.
- Herrera F J, Wong JJ, Chung F (2007) A systematic review of postoperative recovery outcomes measurements after ambulatory Surgery. *Anesth Analg* 105: 63-69.
- Myles PS, Weitkamp G, Jones K, Melick J, Hensen S (2000) Validity and reliability of a postoperative quality of recover score: the QoR-40. *Br J Anaesth* 84: 11-15.
- Myles PS, Hunt JO, Fletcher H, Solly R, Kelly S, et al. (2001) Relationship between quality of recovery in hospital, and quality of life at three months after cardiac surgery. *Anesthesiology* 95: 862-867.
- Myles PS, Viira D, Hunt JO (2006) Quality of life at three years after cardiac surgery: relationship with preoperative status and quality of recovery. *Anaesth Intensive Care* 34: 176-183.
- Machata AM, Kabon B, Willschke H, Fässler K, Gustorff B, et al. (2009) A new instrument for pain assessment in the immediate postoperative period. *Anaesthesia* 64: 392-398.
- Singer S, Wollbrück D, Wulke C, Dietz A, Klemm E, et al. (2009) Validation of the EORTC QLQ-C30 and EORTC QLQ-H&N35 in patients with laryngeal cancer after surgery. *Head Neck* 31: 64-76.
- Wengritzky R, Mettho T, Myles PS, Burke J, Kakos A (2010) Development and validation of a postoperative nausea and vomiting intensity scale. *Br J Anaesth* 104:158-66.

13. Hollenbeck BK, Dunn RL, Wolf JS Jr, Sanda MG, Wood DP, et al. (2008) Development and validation of the convalescence and recovery evaluation (CARE) for measuring quality of life after surgery. *Qual Life Res* 17:915-26.
14. Anastasi A, UrbinaS (1997) Psychological testing. Prentice Hall, Upper Saddle River, New Jersey.
15. Nunnally J, Bernstein IH (1994) Psychometric Theory (3rdedn). McGraw Hill, New York.
16. Food and Drug Administration (2009) Guidance for Industry on Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims (2009).
17. Trochim W, Kane M (2005) Concept mapping: an introduction to structured conceptualization in health care. *Int J Qual Health Care* 17:187-191.
18. Butler SF (2004) The baby or the bath water? *Addiction* 99: 413-414.
19. Furr RM, Bacharach VR (2008) Psychometrics: An Introduction. Sage Publications, Thousand Oaks, California.
20. Byrne BM (2010) Structural equation modeling with AMOS: Basic Concepts, Applications, and Programming. (2ndedn), Lawrence Erlbaum Associates Mahwah, New Jersey.
21. Kline RB (2005) Principles and Practice of Structural Equation Modeling. (2ndedn), The Guilford Press, New York.
22. Fairbank JCT, Pynsent PB (2000) The Oswestry Disability Index. *Spine* 25: 2940-2953.
23. Frank LF, Flynn J, Rothman M (2001) Use of a self-report constipation questionnaire with older adults in long-term care. *The Gerontologist* 41: 778-786.
24. Slappendel R, Simpson K, Dubois D, Keininger DL (2006) Validation of the PAC-SYM questionnaire for opioid-induced constipation in patients with chronic low back pain. *Eur J Pain* 10: 209-217.
25. Jamison RN, Raymond SA, Slawsby EA, Nedeljkovic SS, Katz NP (1998) Opioid therapy for chronic noncancer back pain: A randomized prospective study. *Spine* 23: 2591-2600.
26. Tan G, Jensen MP, Thornby JI, Shanti BF (2004) Validation of the brief pain inventory for chronic non-malignant pain. *J Pain* 5: 133-137.
27. Mattsson NB, Williams HV, Rickels K, Lipman RS, Uhlenhuth EH (1969) Dimensions of symptom distress in anxious neurotic outpatients. *Psychopharmacol Bull* 5: 19-32.
28. McCorkle R, Young K (1978) Development of symptom distress scale. *Cancer Nurs* 1: 373-378.
29. Mendoza TR, Wang XS, Cleeland CS, Morrissey M, Johnson BA, et al. (1999) The rapid assessment of fatigue severity in cancer patients. *Cancer* 85: 1186-1196.
30. Ware JE, Kosinski M, GandekB (2003) SF-36 Health Survey: Manual and interpretation guide. The Health Institute, New England Medical Center, USA.