European Spine Journal (2020) 29:1935–1952
https://doi.org/10.1007/s00586-020-06462-z

ORIGINAL ARTICLE

Development of the "Core Yellow Flags Index" (CYFI) as a brief instrument for the assessment of key psychological factors in patients undergoing spine surgery

Anne F. Mannion[1] · Francine Mariaux[1] · Raluca Reitmeir[2] · Tamas F. Fekete[2] · Daniel Haschtmann[2] · Markus Loibl[2] · Dezsö Jeszenszky[2] · Frank S. Kleinstück[2] · François Porchet[2] · Achim Elfering[3]

Received: 5 March 2020 / Accepted: 10 May 2020 / Published online: 16 June 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

Background Depression, anxiety, catastrophising, and fear-avoidance beliefs are key "yellow flags" (YFs) that predict a poor outcome in back patients. Most surgeons acknowledge the importance of YFs but have difficulty assessing them due to the complexity of the instruments used for their measurement and time constraints during consultations. We performed a secondary analysis of existing questionnaire data to develop a brief tool to enable the systematic evaluation of YFs and then tested it in clinical practice.

Methods The following questionnaire datasets were available from a total of 932 secondary/tertiary care patients (61 ± 16 years; 51% female): pain catastrophising ($N=347$); ZUNG depression ($N=453$); Hospital Anxiety and Depression Scale (anxiety subscale) ($N=308$); fear-avoidance beliefs ($N=761$). The single item that best represented the full-scale score was identified, to form the 4-item "Core Yellow Flags Index" (CYFI). 2422 patients (64 ± 16 years; 54% female) completed CYFI and a Core Outcome Measures Index (COMI) before lumbar spine surgery, and a COMI 3 and 12 months later (FU).

Results The item–total correlation for each item with its full-length questionnaire was: 0.77 (catastrophising), 0.67 (depression), 0.69 (anxiety), 0.68 (fear-avoidance beliefs). Cronbach's α for the CYFI was 0.79. Structural equation modelling showed CYFI uniquely explained variance ($p<0.001$) in COMI at both the 3- and 12-month FUs ($\beta=0.11$ (women), 0.24 (men); and $\beta=0.13$ (women), $\beta=0.14$ (men), respectively).

Conclusion The 4-item CYFI proved to be a simple, practicable tool for routinely assessing key psychological attributes in spine surgery patients and made a relevant contribution in predicting postoperative outcome. CYFI's items were similar to those in the "STarT Back screening tool" used in primary care to triage patients into treatment pathways, further substantiating its validity. Wider use of CYFI may help improve the accuracy of predictive models derived using spine registry data.

Keywords SPINE surgery · Outcome · Yellow flags · Registry data · Core Yellow Flags Index (CYFI)

Introduction

"Yellow flags" are psychological factors and maladaptive beliefs that act as risk factors for persistent pain and prolonged disability in relation to musculoskeletal symptoms [1, 2]. They concern the features that affect how a person manages their situation with regard to their thoughts, feelings and behaviours. The flags are not a diagnosis or a symptom, but an indication that someone may not recover as expected. Some studies have shown that the presence of yellow flags, such as psychological distress/depression [3–5], fear-avoidance beliefs [6, 7] and anxiety [8], also increases the likelihood of a poor outcome after spine surgery. For this reason, such risk factors may influence

✉ Anne F. Mannion
anne.mannion@yahoo.com

[1] Spine Center Division, Department of Teaching, Research and Development, Schulthess Klinik, Lengghalde 2, 8008 Zurich, Switzerland

[2] Spine Center, Schulthess Klinik, Lengghalde 2, 8008 Zurich, Switzerland

[3] Institute for Psychology, University of Bern, Fabrikstrasse 8, 3012 Bern, Switzerland

clinicians' perceptions of the suitability of a patient for a surgical intervention [9] or their opinion of the "appropriateness of surgery" in individual cases [10]. However, even if spine surgeons are cognisant of the flag concept and its importance, many have difficulty detecting yellow flags during the consultation [11] and they rarely formally screen for them [9]. This may be the result of the length and scoring complexity of the current instruments, time constraints in routine consultations, or the perception of not being specifically trained to manage psychosocial attributes identified by such tests [12]. While established self-report instruments exist to evaluate most of the yellow flag constructs of interest, lengthy questionnaires are not suitable for use in the routine clinical setting, where the compliance/involvement of all patients is desired and brevity is of the essence. Further, although brief yellow flag screening instruments have been developed for use in primary care [13, 14] or outpatient physiotherapy [15], these may not be appropriate for use in surgical patients, who appear to be a distinct group with respect to their psychological status pre-treatment [16].

The aim of this study was to create a new, brief tool to routinely assess the yellow flag status of patients being considered for spine surgery, and to evaluate its predictive validity in relation to the outcome of surgery.

Methods

The development of the "yellow-flag" tool followed two phases, as summarised below (details regarding the specific questionnaires and the statistical procedures used are given later, in the respective sections).

Phase 1: strategy to select the "yellow-flag" single items

The multidimensional Core Outcome Measures Index (COMI) [17, 18] comprises single items covering the key outcome domains in patients with spinal disorders and has become a useful tool in the routine evaluation of patient outcome. In accordance with the philosophy behind the COMI of keeping responder burden to a minimum, we sought to develop a complementary set of single-item measures with standardised 5-point response options to assess four of the "core" yellow flags (depression, anxiety, catastrophising and fear-avoidance) [3, 7, 8, 13, 19]. Our previous outcome studies in patients with spinal disorders have provided us with many large datasets containing patients' individual item scores for full-length, established questionnaires addressing these four yellow flags. Table 1 gives a description of the patient samples and the references to the original studies from which the data were taken. The data were derived from a total sample of 932 patients (61 ± 16 years; 51% female; 64% surgical) presenting with spinal problems in secondary or tertiary care. Not all patients had completed each questionnaire, depending on the study they were involved in (see Table 1).

We carried out a secondary analysis of these datasets to select the item that in each case best represented the corresponding full questionnaire while also making sense as a stand-alone question for inclusion in a short set of yellow flag questions, to be coined the "Core Yellow Flags Index" (CYFI). Item quality was assessed using the criteria developed by Stanton et al. [20]. Final judgements about the clinical importance of the best single items for the four instruments were made by an expert group comprising spine surgeons, a methodologist and researchers in the field of spine outcome measures.

Phase 2: test of factor structure and prognostic validity of the four yellow-flag items

In a second phase, we tested the factor structure and prognostic validity of the CYFI using new clinical data collected from May 2015 to Apr 2018. A total of $N = 3344$ patients undergoing surgery of the thoracolumbar spine were asked to complete the CYFI and the COMI, preoperatively, and the COMI at 3- and 12-month follow-up (FU). Questionnaires were completed preoperatively by 2971 (89%) patients, and at 3-month and 12-month FU by 2940 (88%) and 2738 (82%), respectively. A total of 2422 (73%) patients (64.4 ± 15.8 years; 54% female) completed all questionnaires at all three time-points (baseline and both follow-ups). The "Main Pathology" as documented on the Spine Tango surgery form (v.2011; https://www.eurospine.org/forms.htm) was degenerative disease in 1963 (81%) patients, repeat surgery in 194 (8%) and various other pathologies (such as non-degenerative deformity or spondylolisthesis, fracture or trauma, inflammation, infection, tumour, other) in the remaining 265 (11%) patients.

The test–retest reliability of CYFI was assessed in a subgroup of 56 patients (66.3 ± 13.4 years; 55% female) who completed the questionnaire on two occasions preoperatively, 5 ± 9 days apart.

Questionnaires

The questionnaires used to identify the single item yellow flags included:

- the 6-item catastrophising sub-scale of the Coping Strategies Questionnaire (CSQ) [21], or the Pain Catastrophising Scale (PCS) [22, 23]
- the ZUNG Self-rated Depression questionnaire [24]
- the Hospital Anxiety and Depression Scale (HADS) Anxiety subscale [25, 26]

Table 1 Data sources for the secondary analyses used to identify core yellow-flag items for the CYFI

Study		Patients[a]	Brief description of patient group	Questionnaires used				
Label	References			Catastrophising		ZUNG	FABQ phys act	HADS anxiety
				PCS	Catastrophising subscale of CSQ			
A	Staerkle et al. [6]	187 surgical, 90 conservative	Patients with back pain presenting to various tertiary care hospitals and clinics	–	–	277	277	–
B	Meyer et al. [23]	108 conservative	Patients with chronic back pain presenting for care at University hospital departments of Rheumatology and Physical Medicine	108	–	108	108	–
C	Steurer et al. [52], Becker et al. [53]	241 surgical, 67 conservative	Patients with central lumbar spinal stenosis presenting to the Spine Unit of an orthopaedic hospital	–	–	–	308	308
D	Pulkovski et al. [54], Mannion et al. [55], Caporaso et al. [56]	68 conservative	Patients with periodic or chronic back pain presenting for care at University hospital departments of Rheumatology and Physical Medicine	68	–	68	68	–
E	Havakeshian et al. [7]	171 surgical	Patients undergoing surgical decompression (for spinal stenosis or herniated disc) in Spine Unit of orthopaedic hospital	–	171	–	–	–
				176	171	453	761	308

PCS Pain Catastrophising Scale, *CSQ* Coping Strategies Questionnaire, *ZUNG* Zung Self-rated Depression Scale, *FABQ phys act* Fear-Avoidance Beliefs Questionnaire, Physical activity subscale, *HADS anxiety* Hospital Anxiety and Depression Scale, anxiety sub-scale

[a]Includes the given study's data in part or in whole, with/without pilot data collected prior to the full study

- the physical activity sub-scale of the Fear-Avoidance Beliefs Questionnaire (FABQ) [6, 27], to assess beliefs about activity being a cause of the patient's back trouble and fears about the dangers of such activities when experiencing an episode of low back pain.

The questionnaires used to assess the concurrent validity of the single item yellow flags included:

- Visual Analogue Scale (VAS) or graphic/numeric rating scale (GRS/NRS) to measure representative (back or leg) spine-problem-related pain in the last week [28]
- Roland and Morris questionnaire (RMQ), a 24-item questionnaire that assesses disability due to low back pain in relation to various daily functions/activities [29, 30].

The longitudinal validity of the single item flags was evaluated in relation to the COMI.

- The COMI is a 7-item instrument scored 0–10 and comprises questions covering the domains: pain intensity (axial and peripheral, measured separately); function; symptom specific well-being; general quality of life; and social and work disability [17, 31].

All the questionnaires were either originally developed in German or had been adapted and validated for the German language prior to their use in the studies listed in Table 1.

Statistical analyses

Phase 1

Items were favoured for CYFI that: (a) showed a high corrected item–total correlation, i.e. the value of the item corresponded closely to the total scale score without the respective item, indicating the representativeness of the item score for the total scale and its adequacy in representing the construct as a single item; (b) did not display large floor or ceiling effects (i.e. high proportions of scores representing the lowest or highest score possible), that might otherwise indicate a lack of discriminative function, and (c) in Spearman rank correlation analyses, had a meaningful relationship with pain intensity and disability, the clinical outcome measures that have previously been shown to correlate with yellow flag items.

Phase 2

The new sample of data from 2422 surgical patients was analysed using structural equation modelling (SEM). Confirmatory factor analysis was carried out on the preoperative CYFI data, to examine whether the single items corresponded to a single yellow-flag factor, i.e. had a one-dimensional factorial structure with high item loadings on a common factor. Cronbach's alpha was used to assess the internal consistency of the CYFI (≥ 0.70 considered good, [32]).

The hypothesis involving longitudinal data (i.e. that CYFI would add to the prediction of follow-up COMI scores, over and above baseline COMI scores) was tested using SEM by examining the longitudinal directional paths between CYFI at baseline and COMI scores at follow-up, controlling for age, and spinal pathology; this was entitled the "prospective risk path". We estimated risk paths separately for men and women because the prevalence of yellow flags seems to differ between men and women and because the first test of a model that did not allow their risk paths to differ was a worse fit to the empirical data than a model that allowed differences in risk paths. Path coefficients were considered small (0.10), moderate (0.30) and large (0.50) in relation to the effect size classification of Cohen [33].

The reproducibility of single yellow-flag item scores was tested using quadratic weighted Kappas and that of the whole CYFI score was tested with intraclass correlation coefficients (ICC) (in each case, ≥ 0.60 is considered substantial [34]).

The analyses were performed using IBM SPSS (IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp), and AMOS 18.0 software (for the confirmatory factor analysis (CFA) and prospective risk path analyses).

Results

Selection of best yellow-flag single items representing their scales (phase 1)

For PCS, two datasets were available (studies B and D) and for the CSQ catastrophising subscale, one (E) (Table 2). The item "It's terrible and I think it's never going to get any better" (present in both CSQ and PCS) proved to be the best for representing catastrophising. It showed the most consistently high corrected item–total correlations for all studies (0.75, 0.80, and 0.66 for B, D and E, respectively). Compared with most other items of the PCS, floor effects were in the mid-range (33.6%, 39.4% and 33.5%, respectively); there were a few items with lower floor effects, but these were poor in other item characteristics. The chosen item had consistent correlations with pain (0.31, 0.20 and 0.33, respectively) and with the RMQ score (0.52, 0.37 and 0.21, respectively). Finally, the item was verified in the expert group to be one of the best items to represent the pain catastrophising construct as a "stand-alone" item.

The ZUNG Depression scale consists of 20 items. For this construct, data from 3 independent samples were analysed (studies A, B and D) (Table 3). The best stand-alone item for the depression scale was found to be "I feel down-hearted, blue and sad". The item represents the construct very well

Table 2 Results of the statistical analyses to identify the best item representing the domain catastrophising. Item 3, "It's terrible and I think it's never going to get any better" (highlighted in bold), was chosen as the best

Construct	Study sample	Item content	Cronbach's alpha sample B/sample D	Corrected item–total correlation sample B	Corrected item–total correlation sample D	% Floor effect[a] sample B	% Floor effect[a] sample D	Correlation with pain[b] sample B	Correlation with pain[b] sample D	Correlation with disability[c] sample B	Correlation with disability[c] sample D
Pain Catastrophising scale	B ($N=107$) D ($N=67$)		0.925 0.928								
Item 1		I worry all the time about whether the pain will end		0.700	0.788	15.9	23.9	0.208	0.237	0.488	0.340
Item 2		I feel I can't go on		0.711	0.635	37.4	50.0	0.270	0.336	0.419	0.490
Item 3		**It's terrible and I think it's never going to get any better**		**0.747**	**0.798**	**33.6**	**39.4**	**0.314**	**0.203**	**0.517**	**0.373**
Item 4		It's awful and I feel that it overwhelms me		0.762	0.702	43.0	51.5	0.340	0.060	0.491	0.334
Item 5		I feel I can't stand it anymore		0.746	0.647	37.4	39.4	0.325	0.190	0.538	0.378
Item 6		I become afraid that the pain will get worse		0.590	0.662	18.9	25.8	0.148	0.223	0.285	0.288
Item 7		I keep thinking of other painful events		0.375	0.671	52.8	63.6	0.263	0.259	0.314	0.465
Item 8		I anxiously want the pain to go away		0.771	0.596	14.0	15.2	0.318	0.040	0.456	0.138
Item 9		I can't seem to keep it out of my mind		0.724	0.728	28.7	40.9	0.114	-0.020	0.370	0.167
Item 10		I keep thinking about how much it hurts		0.716	0.789	34.0	39.4	0.201	0.096	0.416	0.291

Table 2 (continued)

Construct	Study sample	Item content	Cronbach's alpha sample B/sample D	Corrected item–total correlation sample B	Corrected item–total correlation sample D	% Floor effect[a] sample B	% Floor effect[a] sample D	Correlation with pain[b] sample B	Correlation with pain[b] sample D	Correlation with disability[c] sample B	Correlation with disability[c] sample D
Item 11		I keep thinking about how badly I want the pain to stop		0.744	0.711	17.8	25.8	0.258	0.077	0.458	0.296
Item 12		There's nothing I can do to reduce the intensity of the pain		0.545	0.612	39.6	25.8	0.277	0.279	0.500	0.352
Item 13		I wonder whether something serious may happen		0.556	0.552	31.5	34.8	0.045	0.022	0.163	0.166

Construct	Study sample	Item-label	Cronbach's alpha sample E	Corrected item–total correlation sample E	% Floor effect[a] Sample E	Correlation with pain[b] sample E	Correlation with disability[c] sample E
Coping Strategies Questionnaire, subscale catastrophising	E ($N = 164$)		0.855				
Item 5		**It is terrible and I feel it is never going to get any better**		**0.661**	**33.5%**	**0.327**	**0.209**
Item 12		It is awful and I feel it overwhelms me		0.712	46.3%	0.355	0.309
Item 14		I feel my life isn't worth living		0.574	64.6%	0.253	0.295
Item 28		I worry all the time about whether it will end		0.568	22.0%	0.238	0.255
Item 38		I feel I can't stand it any more		0.645	26.4%	0.386	0.321
Item 42		I feel I can't go on		0.737	41.7%	0.356	0.364

Study sample: See Table 1. [a] > 15% = high, > 70% = adverse [57]; [b] representative pain in the last week; [c] Roland–Morris disability questionnaire

Corrected item–total correlation = correlation between the item and the total scale score that was built of all other items of the scale; high corrected item–total correlation means item represents the scale well (is adequate to represent the construct as a single item)

(corrected item–total correlations in the three samples were 0.67, 0.69 and 0.66, respectively). Floor effects were large (30.6%, 53.0% and 46.7%), but compared with most other items of the ZUNG they were in the mid-range. Correlations with pain in the last week were relatively low but consistent (0.14, 0.19 and 0.17, respectively), whereas those with Roland–Morris disability scores were moderately high and also consistent (0.30, 0.41 and 0.37, respectively). In addition, the item was verified in the expert group to be the most useful stand-alone item for representing the depression construct. Item 20 also showed good item quality in sample A, though less good in B and D, but we considered it unclear whether "not enjoying the things I used to enjoy" might be reflecting the lack of pleasure due to physical pain rather than the depressed mood.

The anxiety subscale of the Hospital Anxiety and Depression Scale (HADS) consists of 7 items, and data from one study (C) were analysed to identify the best fitting single item (Table 4). The item that performed best was item 5 "Worrying thoughts go through my mind". The item showed the highest corrected item–total correlation of all items in the scale (0.69), confirming that it represented the total anxiety score very well. Floor effects were large (52.3%), but about in the mid-range of values for all the seven items (32–76%). The correlation between this item and pain in the last week was the second highest of all the seven items (0.19), and its correlation with disability was third highest (0.22, with the highest correlation being 0.30). Item 1 "I feel tense or 'wound up'" also showed good item quality, but it was felt the colloquialism "being wound up" may have made it unsuitable for use as a stand-alone item, and perhaps caused difficulties with later translations into other languages. Hence, with item 5 ("worrying thoughts…") having the highest item–total correlation, and wording suitable for a stand-alone item, the experts rated this as the best to represent anxiety.

The physical activity subscale of the Fear-Avoidance Beliefs Questionnaire comprises four items, and data were available from four data-sets (studies A, B, C, and D) (Table 5). The item "Physical activity might harm my back" was chosen as the best. It was not "the best" in any of the criteria, but it was always good and more consistently good across the four samples than were other items (respectively, corrected item–total correlation: 0.75, 0.66, 0.62, 0.61; floor effects: 20.6%, 16.1%, 22.7%, 9.0%; correlation with pain: 0.17, 0.23, 0.29, 0.19; correlation with disability: 0.40, 0.45, 0.45, 0.37). Experts rated the item as the best and most credible as a stand-alone item in representing the FABQ-Activity subscale.

The final wording of the CYFI items in English and other languages (official national languages or native languages commonly spoken by patients attending the authors' Spine Center, for which published versions of the full-length questionnaires were available) is shown in Table 6.

Test of factor structure and prognostic validity of the four yellow-flag items (phase 2)

Confirmatory factor analysis showed that the 4 yellow flag items represented a common latent construct (CYFI), with age and pathology being controlled for, and with the 4 CYFI-item loadings on the common CYFI factor being constrained to be the same for men and women (RMSEA = 0.05, CFI = 0.96, χ^2 (19) = 141.60, χ^2/df = 7.45). Cronbach's alpha for the four yellow-flag items was 0.79, showing good internal consistency.

The test of prognostic validity for CYFI included a structural equation model with CYFI predicting COMI at 3-month follow-up and 12-month follow-up while controlling for preoperative COMI and pathology (Fig. 1). On a cross-sectional basis, preoperative CYFI and COMI scores were highly correlated (Fig. 1: $\beta = 0.52$ for men, $\beta = 0.42$ for women; each $p < 0.001$). CYFI explained a significant proportion of the variance in COMI at 3-month FU ($\beta = 0.24$, approximately 8% variance explained in men and $\beta = 0.11$, approximately 2% variance in women, $p < 0.001$; Fig. 1), i.e. CYFI contributed to a small but significant extent to explaining the treatment effect. The stability between COMI at baseline and COMI at 3-month FU was low—due to the treatment—with $\beta = 0.15$ in men, $\beta = 0.20$ in women (Fig. 1). The stability between COMI at 3-month FU and COMI at 12-month FU was high ($\beta = 0.61$ in men, $\beta = 0.55$ in women, $p < 0.001$; Fig. 1). Nonetheless, CYFI added significantly and independently to the prediction of COMI at 12-month FU ($\beta = 0.14$ in men, approx. 4% variance explained, $p < 0.001$; $\beta = 0.13$ in women, approx. 3% variance explained, $p < 0.001$; Fig. 1) and explained variation in the COMI at 12-month FU that was not explained by individual differences in COMI existing at either baseline or 3-month FU. The fit of the model was good (RMSEA = 0.04, CFI = 0.97, χ^2 (39) = 216.92, χ^2/df = 5.56).

Test retest reliability for each item of the CYFI was 0.60–0.76 and for the CYFI whole score, 0.72 (95% CI 0.58–0.86).

Discussion

Our study showed that the newly developed 4-item CYFI constitutes a simple, practicable, reliable and valid tool for routinely assessing key psychological attributes in patients undergoing treatment for spinal disorders in tertiary care. The brevity of the CYFI should make it a useful addition to the brief COMI in the self-assessment of baseline status before surgery. It may be used by clinicians to orientate

Table 3 Results of the statistical analyses to identify the best item representing the domain depression. Item 1, "I feel down-hearted, blue and sad" (highlighted in bold), was chosen as the best

Construct	Study sample	Item-label	Cronbach's Alpha Sample A/ Sample B/ Sample D	Corrected item–total correlation sample A	Corrected item–total correlation sample B	Corrected item–total correlation sample D	% Floor effect[a,d] Sample A	% Floor effect[a,d] Sample B	% Floor effect[a,d] Sample D	Correlation with Pain[b] Sample A	Correlation with Pain[b] Sample B	Correlation with Pain[b] Sample D	Correlation with disability[c] Sample A	Correlation with disability[c] Sample B	Correlation with disability[c] Sample D
Zung Depression scale	A($N=272$) B ($N=107$) D ($N=67$)		0.844 0.880 0.854												
Item 1		**I feel down-hearted, blue and sad**		**0.665**	**0.694**	**0.656**	**30.6**	**53.0**	**46.7**	**0.141**	**0.192**	**0.165**	**0.304**	**0.412**	**0.370**
Item 2		Morning is when I feel the best [d]		0.213	0.145	0.379	14.8	12.1	20.5	0.025	-0.032	0.071	0.090	0.038	0.192
Item 3		I have crying spells or feel like it		0.556	0.492	0.502	52.3	81.8	64.8	0.177	0.237	0.152	0.367	0.346	0.276
Item 4		I have trouble getting to sleep at night		0.426	0.441	0.426	21.3	28.8	33.6	0.160	0.300	0.225	0.289	0.337	0.447
Item 5		I eat as much as I used to [d]		0.361	0.279	0.345	44.4	37.9	50.7	0.266	0.043	0.179	0.311	0.183	0.285
Item 6		I still enjoy sex [d]		0.257	0.251	0.470	51.0	40.0	35.8	0.294	0.203	0.213	0.507	0.279	0.349
Item 7		I notice that I am losing weight		0.415	0.151	0.169	67.0	83.3	71.9	0.129	0.029	0.085	0.144	0.225	0.172
Item 8		I have trouble with constipation		0.392	0.442	0.238	56.1	60.6	59.0	0.282	0.043	0.110	0.386	0.101	0.251

Table 3 (continued)

Construct	Study sample	Item-label	Cronbach's Alpha Sample A/ Sample B/ Sample D	Corrected item–total correlation sample A	Corrected item–total correlation sample B	Corrected item–total correlation sample D	% Floor effect[a,d] Sample A	% Floor effect[a,d] Sample B	% Floor effect[a,d] Sample D	Correlation with Pain[b] Sample A	Correlation with Pain[b] Sample B	Correlation with Pain[b] Sample D	Correlation with disability[c] Sample A	Correlation with disability[c] Sample B	Correlation with disability[c] Sample D
Item 9		My heart beats faster than usual		0.370	0.369	0.397	62.6	75.8	55.9	0.073	0.188	0.117	0.128	0.295	0.256
Item 10		I get tired for no reason		0.620	0.340	0.538	37.0	51.5	35.9	0.216	0.235	0.120	0.427	0.288	0.404
Item 11		My mind is as clear as it used to be [d]		0.563	0.480	0.243	54.2	60.6	72.0	0.107	0.072	0.013	0.274	0.225	0.043
Item 12		I find it easy to do the things I used to [d]		0.677	0.639	0.629	23.6	27.7	29.0	0.322	0.119	0.249	0.509	0.287	0.534
Item 13		I am restless and can't keep still		0.425	0.490	0.398	32.4	42.4	32.2	0.168	0.112	0.103	0.231	0.281	0.248
Item 14		I am hopeful about the future [d]		0.569	0.567	0.514	41.5	36.4	46.2	0.044	-0.061	0.049	0.231	0.075	0.218
Item 15		I am more irritable than usual		0.613	0.606	0.434	34.9	50.0	31.9	0.253	0.143	0.131	0.380	0.400	0.295
Item 16		I find it easy to make decisions [d]		0.576	0.577	0.414	34.3	28.8	37.1	0.229	0.074	0.021	0.346	0.075	0.191

Table 3 (continued)

Construct	Study sample	Item-label	Cronbach's Alpha Sample A/ Sample B/ Sample D	Corrected item–total correlation sample A	Corrected item–total correlation sample B	Corrected item–total correlation sample D	% Floor effect[a,d] Sample A	% Floor effect[a,d] Sample B	% Floor effect[a,d] Sample D	Correlation with Pain[b] Sample A	Correlation with Pain[b] Sample B	Correlation with Pain[b] Sample D	Correlation with disability[c] Sample A	Correlation with disability[c] Sample B	Correlation with disability[c] Sample D
Item 17		I feel that I am useful and needed [d]		0.550	0.604	0.528	54.8	47.0	61.3	0.212	0.007	0.046	0.191	0.136	0.256
Item 18		My life is pretty full [d]		0.610	0.623	0.544	50.9	39.4	65.3	0.125	0.100	0.100	0.248	0.285	0.297
Item 19		I feel others would be better off if I were dead		0.325	0.398	0.437	89.6	95.5	90.4	0.175	0.089	0.112	0.251	0.274	0.202
Item 20		I still enjoy the things I used to [d]		0.679	0.475	0.513	51.9	47.0	60.1	0.305	0.175	0.071	0.413	0.348	0.176

Study sample: See Table 1. [a,d]> 15% = high, > 70% = adverse [57]; [b]representative pain in the last week; [c]Roland–Morris disability questionnaire; [d]item is phrased positively, so the scoring is first reversed before calculating floor effects

Corrected item–total correlation = correlation between the item and the total scale score that was built of all other items of the scale; high corrected item–total correlation means item represents the scale well (is adequate to represent the construct as a single item)

Table 4 Results of the statistical analyses to identify the best item representing the domain anxiety. Item 5, "Worrying thoughts go through my mind" (highlighted in bold), was chosen as the best

Construct	Study sample	Item-label	Cronbach's alpha	Corrected item–total correlation	Floor effect[a]	Correlation with Pain[b]	Correlation with disability[c]
Hospital Anxiety and Depression Scale, subscale anxiety	C ($N=308$)		0.809				
Item 1		I feel tense or "wound up"		0.648	32.1%	0.195	0.295
Item 3		I get a sort of frightened feeling as if something awful is about to happen		0.578	49.0%	0.058	0.212
Item 5		**Worrying thoughts go through my mind**		**0.689**	**52.3%**	**0.190**	**0.224**
Item 7		I can sit at ease and feel relaxed		0.414	38.0%	0.165	0.234
Item 9		I get a sort of frightened feeling like 'butterflies in the stomach'		0.616	61.4%	0.132	0.223
Item 11		I feel restless, as if I have to be on the move		0.393	48.1%	0.000	− 0.023
Item 13		I get sudden feelings of panic		0.570	75.6%	0.117	0.154

Study sample: See Table 1. [a]> 15% = high, > 70% = adverse [57]; [b]representative pain in the last week; [c]Roland–Morris disability questionnaire

Corrected item–total correlation = correlation between the item and the total scale score that was built of all other items of the scale; high corrected item–total correlation means item represents the scale well (is adequate to represent the construct as a single item)

themselves with regard to the yellow flag status of their patients, and its data may be able to strengthen the existing predictor models of surgical outcome.

A number of brief tools exist to assess yellow flags, but these have focused on chronic LBP patients in primary care, occupational health or physical therapy settings [13–15, 35]. Several factors provided the impetus for us to create a new tool designed to be used with surgical patients. Patients in tertiary care are intrinsically different from those in primary care, in terms of both their symptom severity and degree of psychological disturbance [16]. In creating our own tool, we wished to use, as a basis, questionnaires that had previously been used with patients in secondary and tertiary care study settings. We also wanted to select items from questionnaires that were available in our 3 national languages (German, French and Italian) as well as English and other languages spoken in our country for which a version of COMI exists (see Table 6). Further, rather than employing a binary response option (yes/no to whether the statement applies), as used for example in the STarT Back, we wanted to offer a 5-point graded scale that would be consistent with the items in the COMI. Nonetheless, in considering the final items for inclusion in our tool, we attempted to align with the STarT Back, where feasible and supported by the item-quality analyses. The STarT Back items did not all come from the same full-length questionnaires as used in the present study: they were the same for anxiety (i.e. HADS) and catastrophising (i.e. PCS), and the same two items were considered to be most representative of these domains in both studies. The depression item in the STarT Back ("in general, I have not enjoyed all the things I used to enjoy") came from the Patient Health Questionnaire (PHQ-2) rather than the ZUNG. The ZUNG contains a similar item (item 20) and, although it showed good item quality in our sample A, it was not consistently good for samples C and D (Table 3). Moreover, when presented as a stand-alone item, we considered that "not enjoying the things I used to enjoy" was too unspecific as a depression item, liable to inadvertently capture the impact of pain on the enjoyment of activities rather than the mental state of being depressed and losing interest (especially in surgical patients with their higher pain levels). The fear item in the STarT Back ("not safe for a person with a condition like mine to be physically active") originates from the Tampa Kinesiophobia questionnaire and could perhaps be considered a more unwieldy way of saying "Physical activity might harm my back" (our chosen FABQ item), albeit with some ambiguity in the interpretation of the word "safe". Rasch analyses have previously identified this Tampa

Table 5 Results of the statistical analyses to identify the best item representing the domain fear-avoidance beliefs. Item 2, "Physical activity might harm my back." (highlighted in bold), was chosen as the best

Construct	Study sample	Item-label	Cronbach's alpha samples A/B/C/D	Corrected item–total correlation sample A	Corrected item–total correlation sample B	Corrected item–total correlation sample C	Corrected item–total correlation sample D	Floor effect[a] sample A	Floor effects[a] sample B	Floor effect[a] sample C	Floor effect[a] sample D	Correlation with pain[b] sample A	Correlation with pain[b] sample B	Correlation with pain[b] sample C	Correlation with pain[b] sample D	Correlation with disability[c] sample A	Correlation with disability[c] sample B	Correlation with disability[c] sample C	Correlation with disability[c] sample D
FABQ Activity	A ($N=270$) B ($N=107$) C ($N=305$) D ($N=66$)		0.802 0.852 0.819 0.782																
Item 1		Physical activity makes my pain worse		0.513	0.501	0.538	0.573	9.3	5.9	16.7	1.9	0.478	0.408	0.222	0.200	0.533	0.379	0.446	0.389
Item 2		**Physical activity might harm my back**		**0.753**	**0.656**	**0.617**	**0.611**	**20.6**	**16.1**	**22.7**	**9.0**	**0.171**	**0.225**	**0.293**	**0.188**	**0.403**	**0.447**	**0.450**	**0.368**
Item 3		I should not do physical activities which (might) make my pain worse		0.773	0.763	0.608	0.665	23.4	13.1	18.2	5.6	0.116	0.302	0.108	0.158	0.391	0.489	0.298	0.363
Item 4		I cannot do physical activities which (might) make my pain worse		0.752	0.653	0.596	0.624	23.4	13.2	21.2	8.3	0.149	0.371	0.224	0.304	0.501	0.522	0.355	0.490

Study sample: See Table 1. [a]> 15% = high, > 70% = adverse [57]; [b]representative pain in the last week; [c]Roland–Morris disability questionnaire

Corrected item–total correlation = correlation between the item and the total scale score that was built of all other items of the scale; high corrected item–total correlation means item represents the scale well (is adequate to represent the construct as a single item)

Table 6 German, English, French, Italian, Spanish, Portuguese and Hungarian versions of the CYFI (see footnote for further details)

Language	Introduction[a]	Responses[a]	Catastrophising[b]	Depression[b]	Anxiety[b]	Fear-avoidance beliefs[b]
German (study language)	Die folgende Liste enthält Aussagen von Personen, die an Rückenproblemen leiden. Möglicherweise treffen auch auf Sie einige Aussagen zu. Kreuzen Sie bitte an, wie sehr jede der Aussagen auf Sie zutrifft.	(1) Trifft gar nicht zu (2) Trifft eher nicht zu (3) Teils-teils (4) Trifft eher zu (5) Trifft genau zu	Es ist schrecklich und ich denke, dass es nie mehr besser wird	Ich fühle mich bedrückt, schwermütig und traurig	Mir gehen beunruhigende Gedanken durch den Kopf	Körperliche Aktivität könnte meinem Rücken schaden
English	The following list contains statements from people with back problems. Some of these statements may also reflect your own thoughts/feelings. Please tick the box that best describes the extent to which each statement reflects your own thoughts/feelings at the moment.	(1) Not at all (2) To a slight degree (3) To a moderate degree (4) To a great degree (5) Totally/completely	It's terrible and I think it's never going to get any better	I feel down-hearted, blue and sad	Worrying thoughts go through my mind	Physical activity might harm my back
French	La liste ci-dessous contient des affirmations de patients souffrant de problèmes de dos. Certaines de ces affirmations reflètent peut-être aussi vos propres pensées/sentiments. Veuillez s'il vous plaît cocher la case qui décrit au mieux à quel point chaque affirmation reflète vos pensées/sentiments en ce moment.	(1) Pas du tout (2) Un peu (3) Modérément (4) Beaucoup (5) Tout à fait	C'est terrible et je pense que cela ne s'améliorera jamais	Je me sens découragé et mélancolique	Je me fais du souci	L'activité physique pourrait abîmer mon dos
Italian	La seguente lista contiene affermazioni espresse da persone con problemi alla schiena. Alcune di queste affermazioni potrebbero rispecchiare anche i suoi pensieri/le sue sensazioni. La prego di marcare la casella che al meglio descrive la misura in cui ogni affermazione rispecchia i suoi pensieri/le sue sensazioni al momento.	(1) Per niente (2) Leggermente (3) Moderatamente (4) Significativamente (5) Totalmente	È orribile e penso che non finirà mai	Mi sento scoraggiato, depresso e triste	Dei pensieri preoccupanti mi attraversano la mente	L'attività fisica potrebbe danneggiare la mia schiena

Table 6 (continued)

Language	Introduction[a]	Responses[a]	Catastrophising[b]	Depression[b]	Anxiety[b]	Fear-avoidance beliefs[b]
Spanish	La siguiente lista contiene afirmaciones de personas con problemas de espalda. Algunas de ellas pueden también reflejar sus propios pensamientos/sentimientos. Por favor, marque la casilla que mejor representa sus pensamientos/sentimientos en cada frase en este momento.	(1) Nada en absoluto (2) Un poco (3) Moderadamente (4) Mucho (5) Totalmente	Es terrible y pienso que esto nunca va a mejorar	Me siento triste y deprimido/a	Tengo la cabeza llena de preocupaciones	La actividad física podría dañar mi espalda
Portuguese	A lista de termos apresenta afirmações relacionadas com as pessoas que possuem problemas na coluna vertebral. Escolha a melhor afirmação que corresponde aos seus sintomas ou situação no momento.	(1) De modo nenhum (2) Em pequeno grau (3) Em grau moderado (4) Em grande grau (5) Totalmente	É terrível e penso que nunca mais vai melhorar	Eu me sinto triste e "para baixo"	Tenho a cabeça cheia de preocupações	A atividade física pode machucar minhas costas
Hungarian	Az alábbi állítások hát/ deréktáji problémákkal küzdő személyektől származnak. Néhány közülük önre is vonatkozhat. Kérem jelölje, hogy melyik jellemzi/tükrözi leginkább az ön jelenlegi állapotát.	(1) Egyáltalán nem (2) Egy kicsit (3) Közepesen (4) Számottevően (5) Teljes mértékben	Ez rettenetes és azt gondolom, hogy sohasem fog egy kicsit sem javulni	Lehangolt és csüggedt vagyok	Aggasztó gondolatok járnak a fejemben	A fizikai aktivitás káros lehet a hátamnak

[a]The introductory text and the response options were developed based on the wording of the German versions of the FABQ [6] and PCS [23] and were formulated to make sense in relation to all of the flag items. They were translated into English by two native English speakers who were fluent in German, also taking into account the wording used in the English versions of the FABQ and PCS. For the other languages, the introduction and response options were then translated from English by native speakers, and discussed within a multilingual expert group; however, none of the other language versions have undergo formal cross-cultural adaptation using the established guidelines

[b]The CYFI items themselves were taken from published versions of the corresponding full-length questionnaires in the given language (further details of original sources available on request)

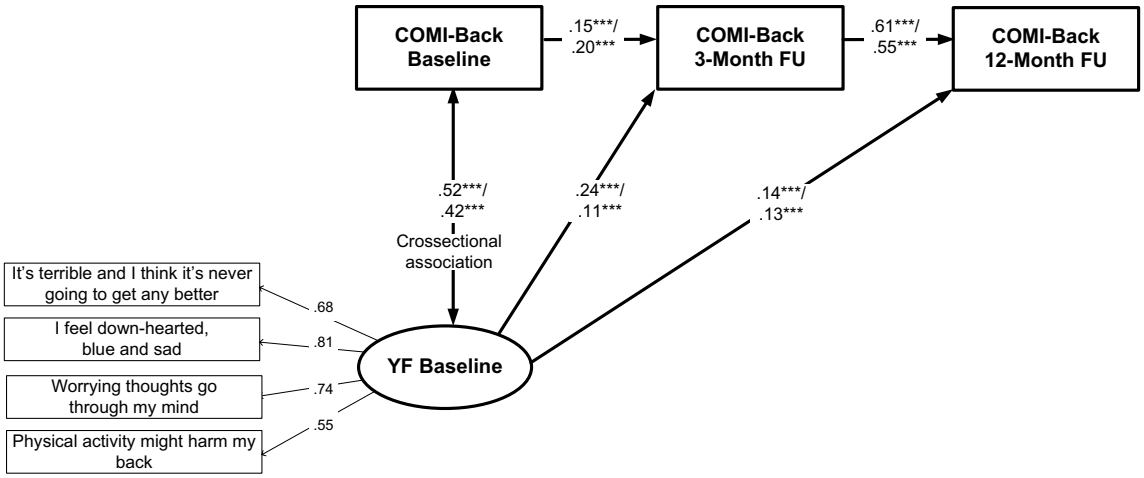

Fig. 1 Results of the structural equation modelling showing the factor analysis of the CYFI and the correlations between CYFI at baseline and COMI at follow-up (FU), controlling for preoperative COMI, sex and age. The fit of the model was good (RMSEA = 0.04, CFI = 0.97, $\chi^2 = 216.92$, $df = 39$, $\chi^2/df = 5.56$). The first coefficient in each pathway indicates the standardised regression coefficient for men, and the second, for women *** all $p < 0.001$, two-tailed

Kinesiophobia item as being psychometrically poor [36] and showing differential item functioning with respect to gender [37]. Interestingly, recent qualitative analyses performed by the STarT Back group revealed that the STarT Back depression and fear items were considered "cumbersome" by both patients and general practitioners alike [38]. This substantiates our aforementioned misgivings about these two items. Despite the above differences, test–retest reliabilities were similar for the two tools: the quadratic weighted kappa for the psychosocial subscale of the STarT Back completed by all 53 patients studied was 0.69 (0.51–0.81) and, for 23 of their patients reporting stable symptoms, 0.76 (0.52–0.89) [13]; for the CYFI, the corresponding value was 0.72 (95% CI 0.58–0.86).

Identifying a need to include a yellow flag measure in the baseline assessment of back pain patients, Cedraschi et al. [35] added two yellow flag questions to the COMI, to assess depression and anxiety. The wording was created by the authors, rather than being extracted from established questionnaires, and simply enquired "how much did you feel anxious?" and "how much did you feel depressed?", with a list of 5–6 thoughts and feelings being provided for each question as examples of what it might mean to feel anxious or depressed. Such "double/multiple-barrelled" (or compound) questions that enquire about many feelings/thoughts within one and the same question can pose difficulties, since respondents wishing to endorse only one of the options might be confused how to answer [32]. Moreover, the predictive validity of their flag items in relation to outcome was not evaluated. It was suggested that the items be incorporated into the existing COMI to provide a modified-COMI with a psychological dimension, by taking the higher of the two scores (anxiety or depression) and averaging it with the remaining COMI item scores. We see numerous problems with this. Firstly, it would cause confusion with respect to the scoring of the COMI as an outcome instrument and would render incomparable the scores from studies with and without the flag questions. Secondly, the psychological items do not constitute key outcomes for many spinal disorders; they may be important predictors or screening items, but they are not "core outcomes" [39], which means inclusion of their scores in the overall COMI score would likely reduce the responsiveness of the instrument (as was seen in [35]). For the CYFI, our recommendation is to view it as an independent tool, calculating an unweighted sum-score for its four items, since in factor analysis all made a reasonable contribution to the latent variable "yellow flags" (Fig. 1).

We showed that the CYFI made a significant independent contribution to the prediction of COMI scores at 3- and 12-month follow-up. Our findings were hence in keeping with the numerous studies that have shown that higher scores on yellow flag questionnaires generally predispose to poorer outcome [40–42]. In the present study, the proportion of variance in outcome accounted for by CYFI (2–8%, depending on gender and follow-up time-point) was greater than that reported for the psychological variables in some previous studies (0–2% [6, 43, 44]) and lower than that reported in others (15–20% [4]). In many studies, only the statistical significance of the effect or the variance accounted for by the whole model was reported, rather than the size of the effect for the psychological variables per se, making it difficult to draw comparisons [45, 46] (and see reviews in [40, 41, 47]). Also, some of the published studies were not truly prospective and most omitted from their models the cross-sectional

relationship between psychosocial factors and baseline outcome scores. In the present study, COMI and CYFI were highly correlated at baseline, meaning that the unique contribution of CYFI in predicting COMI at follow-up—beyond that explained by COMI at baseline—was somewhat limited. In our prediction of 12-month COMI, there was, in addition to the direct effect of CYFI, also the indirect effect of CYFI on COMI at 12 months that was mediated by COMI at 3 months. The strong correlation between baseline COMI and CYFI probably indicates that the psychological status of patients at baseline is closely related to their ongoing pain problem and reflects to a lesser extent psychological problems beyond this. In other words, the yellow flags measured in the current sample have a more "situational" origin, driven by current pain and disability, and less of a "stable" trait-like origin reflecting psychological problems unrelated to current pain and disability. The situational component of CYFI is probably less powerful in predicting outcome compared with the more stable component. It is also highly likely that in some patients the psychological factors play a major role, whereas in other patients they have no significance. This has been reported in the literature before, where psychological factors appear to have a greater part to play in more "contentious" diagnoses for which the indication for surgery is less certain, compared with those for which the indication is more clear-cut [41]. Further investigations in this area are warranted such that we might direct our future attention to those patients whose outcome is especially influenced by psychological factors. It is difficult to do true experimental studies in this field to prove causality; however, the future collection of CYFI data also at follow-up, in addition to COMI data, and the use of cross-lagged panel correlations, might provide a method for identifying the source, direction and extent of the associations.

The observation that psychological variables significantly influence outcome often provokes the discussion as to whether, having identified that a patient demonstrates significant yellow flags, surgery should still be recommended. We do not believe that the effect size (in the present study, small to moderate; see above) is great enough to promote the CYFI as a tool to be used to deny operative procedures to patients who otherwise have a clear clinical indication for surgery. Indeed, to the authors' knowledge, no such psychological screening tool currently exists, and it is well known that many high-scoring patients still derive great benefit from surgery. Instead, we believe the current findings provide an impetus for administering the CYFI as part of a systematic collection of baseline data, along with numerous other risk factors, such that these can be included in predictive analytical models to improve the accuracy of individual outcome prediction. Many factors ultimately contribute to explaining the variance in individual outcomes; the more variables we are able to identify that make a significant contribution, the more accurate our overall predictions should be. Having a knowledge of the preoperative CYFI score for individual patients may also be useful in daily clinical practice to open a dialogue about these issues with the patient and to better manage their expectations of treatment. This may minimise the subsequent dissatisfaction with outcome that can follow from having overly optimistic expectations [48]. The findings might also be considered as support for more research on the clinical benefit of cognitive behavioural therapy (CBT) accompanying surgery. A number of studies [49–51] have shown positive effects, and this is a field of ongoing study, particularly in relation to the selection of appropriate cases.

Our study had a number of limitations. First, the data used in the development of the CYFI were from patients in secondary or tertiary care; the majority, but not all, were surgical patients. Second, the CYFI contains only "negative items" and there are no items enquiring about positive affect, coping strategies or resilience. Although these attributes are often believed to be the "opposite" of the yellow flag attributes, in some studies of spine surgery patients they have been shown to contribute to the prediction of outcome [43]. Third, in the longitudinal study, questionnaires were not completed by all patients at baseline (11% failed to complete one, mostly due to language problems, administrative errors, and emergency admissions) and other patients did not return a questionnaire at 3-month or 12-month follow-up (12–18%). This may have introduced attrition bias in the findings. Fourth, the reason that sex-specific models showed better fit currently eludes us. However, it is important to appreciate that yellow flags do not operate in isolation from other factors [2], and more elaborate models will ultimately be required. Further, such models should be externally validated (i.e. tested for their predictive ability in a separate population of patients), a step that was beyond the scope of the present study. The CYFI items were taken from published versions of the corresponding full-length questionnaires in each language. Nonetheless, confirmation of the adequacy of the different language versions as a group of items and of the corresponding introduction and response options, which have not been formally validated (Table 6), along with further evaluation of the performance of the CYFI (internal and test–retest reliability, construct and longitudinal validity, etc.) in each language, is encouraged. And finally, we cannot yet advise on the cut-offs required for indicating that a patient is "yellow flag positive", on a binary basis; we hope to address this in future studies.

In summary, the 4-item CYFI proved to be a simple, practicable, reliable and valid tool for routinely assessing key psychological attributes in spine surgery patients. The CYFI made a statistically significant contribution to the prediction of patient outcome after surgery. In this way, its widespread use may assist in developing better outcome prediction

tools, based on the systematic collection of baseline data, e.g. in spine registries. The brevity of the instrument makes it suitable for implementation in everyday clinical practice, as part of the baseline assessment of patients undergoing spine surgery.

Acknowledgements We thank all the patients of the Schulthess Klinik who contributed their data. We thank Dave O'Riordan, Gordana Balaban, Stéphanie Dosch, and Selina Nauer for their administration of the Spine Tango surgery forms and patient-rated outcome measures. We thank Astrid Junge for her help and advice in finalising the wording of the introductory text and response options for the German version of the CYFI and Melissa Wilhelmi for her English translations of the same. We thank the authors of the original studies from which data were derived to develop the new CYFI.

Compliance with ethical standards

Conflict of interest The authors declare that they have no potential conflict of interest.

References

1. Kendall NA (1999) Psychological approaches to the prevention of chronic pain: the low back paradigm. Baillieres Best Pract Res Clin Rheumatol 13:545–554
2. Nicholas MK, Linton SJ, Watson PJ, Main CJ, Decade of the Flags" Working G (2011) Early identification and management of psychological risk factors ("yellow flags") in patients with low back pain: a reappraisal. Phys Ther 91:737–753. https://doi.org/10.2522/ptj.20100224
3. Block AR, Ohnmeiss DD, Guyer RD, Rashbaum RF, Hochschuler SH (2001) The use of presurgical psychological screening to predict the outcome of spine surgery. Spine J 1:274–282
4. Mannion AF, Elfering A, Staerkle R, Junge A, Grob D, Dvorak J, Jacobshagen N, Semmer NK, Boos N (2007) Predictors of multidimensional outcome after spinal surgery. Eur Spine J 16:777–786
5. Wilhelm M, Reiman M, Goode A, Richardson W, Brown C, Vaughn D, Cook C (2017) Psychological Predictors of Outcomes with Lumbar Spinal Fusion: A Systematic Literature Review. Physiother Res Int. https://doi.org/10.1002/pri.1648
6. Staerkle R, Mannion AF, Elfering A, Junge A, Semmer NK, Jacobshagen N, Grob D, Dvorak J, Boos N (2004) Longitudinal validation of the fear-avoidance beliefs questionnaire (FABQ) in a Swiss-German sample of low back pain patients. Eur Spine J 13:332–340
7. Havakeshian S, Mannion AF (2013) Negative beliefs and psychological disturbance in spine surgery patients: a cause or consequence of a poor treatment outcome? Eur Spine J 22:2827–2835. https://doi.org/10.1007/s00586-013-2822-5
8. de Groot KI, Boeke S, van den Berge HJ, Duivenvoorden HJ, Bonke B, Passchier J (1997) The influence of psychological variables on postoperative anxiety and physical complaints in patients undergoing lumbar surgery. Pain 69:19–25
9. Grevitt M, Pande K, O'Dowd J, Webb J (1998) Do first impressions count? A comparison of subjective and psychologic assessment of spinal patients. Eur Spine J 7:218–223
10. Mannion AF, Pittet V, Steiger F, Vader JP, Becker HJ, Porchet F (2014) Development of appropriateness criteria for the surgical treatment of symptomatic lumbar degenerative spondylolisthesis (LDS). Eur Spine J 23:1903–1917. https://doi.org/10.1007/s00586-014-3284-0
11. Lee KC, Patel S, Sell P (2014) Identification of obstacles to recovery in secondary care. Eur Spine J Suppl 1:S125
12. Kent P, Mirkhil S, Keating J, Buchbinder R, Manniche C, Albert HB (2014) The concurrent validity of brief screening questions for anxiety, depression, social isolation, catastrophization, and fear of movement in people with low back pain. Clin J Pain 30:479–489. https://doi.org/10.1097/AJP.0000000000000010
13. Hill JC, Dunn KM, Lewis M, Mullis R, Main CJ, Foster NE, Hay EM (2008) A primary care back pain screening tool: identifying patient subgroups for initial treatment. Arthritis Rheum 59:632–641. https://doi.org/10.1002/art.23563
14. Linton SJ, Nicholas M, MacDonald S (2011) Development of a short form of the Orebro Musculoskeletal Pain Screening Questionnaire. Spine (Phila Pa 1976) 36:1891–1895. https://doi.org/10.1097/BRS.0b013e3181f8f775
15. Lentz TA, Beneciuk JM, Bialosky JE, Zeppieri G Jr, Dai Y, Wu SS, George SZ (2016) Development of a Yellow Flag Assessment Tool for Orthopaedic Physical Therapists: Results From the Optimal Screening for Prediction of Referral and Outcome (OSPRO) Cohort. J Orthop Sports Phys Ther 46:327–343. https://doi.org/10.2519/jospt.2016.6487
16. Abbott AD, Tyni-Lenne R, Hedlund R (2010) The influence of psychological factors on pre-operative levels of pain intensity, disability and health-related quality of life in lumbar spinal fusion surgery patients. Physiotherapy 96:213–221. https://doi.org/10.1016/j.physio.2009.11.013
17. Mannion AF, Porchet F, Kleinstück F, Lattig F, Jeszenszky D, Bartanusz V, Dvorak J, Grob D (2009) The quality of spine surgery from the patient's perspective: Part 1. The Core Outcome Measures Index (COMI) in clinical practice. Eur Spine J 18:367–373
18. Deyo RA, Battié M, Beurskens AJHM, Bombardier C, Croft P, Koes B, Malmivaara A, Roland M, Von Korff M, Waddell G (1998) Outcome measures for low back pain research. A proposal for standardized use. Spine 23:2003–2013
19. Abbott AD, Tyni-Lenne R, Hedlund R (2011) Leg pain and psychological variables predict outcome 2–3 years after lumbar fusion surgery. Eur Spine J 20:1626–1634. https://doi.org/10.1007/s00586-011-1709-6
20. Stanton JM, Sinar EF, Balzer WK, Smith PC (2002) Issues and strategies for reducing the length of self-report scales. Pers Psychol 55:167–194
21. Rosenstiel AK, Keefe FJ (1983) The use of coping strategies in chronic low back pain patients: relationship to patient characteristics and current adjustments. Pain 17:33–44
22. Sullivan MJL, Stanish W, Waite H, Sullivan M, Tripp D (1998) Catastrophizing, pain, and disability in patients with soft-tissue injuries. Pain 77:253–260
23. Meyer K, Sprott H, Mannion AF (2008) Cross-cultural adaptation, reliability, and validity of the German version of the Pain Catastrophizing Scale. J Psychosom Res 64:469–478
24. Zung WW (1965) A Self-Rating Depression Scale. Arch Gen Psychiatry 12:63–70
25. Zigmond AS, Snaith RP (1983) The Hospital Anxiety and Depression Scale. Acta Psychiatr Scand 67:361–370
26. Herrmann C, Buss U (1994) Vorstellung und Validierung einer deutschen Version der "Hospital Anxiety and Depression Scale" (HAD-Skala): Ein Fragebogen zur Erfassung des psychischen Befindens bei Patienten mit körperlichen Beschwerden. Diagnostica 40:143–154
27. Waddell G, Newton M, Henderson I, Somerville D, Main CJ (1993) A Fear-Avoidance Beliefs Questionnaire (FABQ) and the

role of fear-avoidance beliefs in chronic low back pain and disability. Pain 52:157–168
28. Haefeli M, Elfering A (2006) Pain assessment. Eur Spine J 15(Suppl 1):S17–24
29. Roland M, Morris R (1983) A study of the natural history of back pain. Part 1: Development of a reliable and sensitive measure of disability in low-back pain. Spine 8:141–144
30. Exner V, Keel P (2000) Erfassung der Behinderung bei Patienten mit chronischen Rückenschmerzen. Schmerz 14:392–400
31. Mannion AF, Elfering A, Staerkle R, Junge A, Grob D, Semmer NK, Jacobshagen N, Dvorak J, Boos N (2005) Outcome assessment in low back pain: how low can you go? Eur Spine J 14:1014–1026
32. Streiner DL, Norman GR (1995) Health Measurement Scales: a practical guide to their development and use. Oxford University Press Inc., Oxford
33. Cohen J (1977) Statistical power analysis for the behavioral sciencies, Revised edn. Academic Press, San Diego
34. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159–174
35. Cedraschi C, Marty M, Courvoisier DS, Foltz V, Mahieu G, Demoulin C, Gierasimowicz Fontana A, Norberg M, de Goumoens P, Rozenberg S, Genevay S, Section Rachis de la Societe Francaise de R (2016) Core Outcome Measure Index for low back patients: do we miss anxiety and depression? Eur Spine J 25:265–274. https://doi.org/10.1007/s00586-015-3935-9
36. Woby SR, Roach NK, Urmston M, Watson PJ (2005) Psychometric properties of the TSK-11: a shortened version of the Tampa Scale for Kinesiophobia. Pain 117:137–144. https://doi.org/10.1016/j.pain.2005.05.029
37. Damsgard E, Fors T, Anke A, Roe C (2007) The Tampa Scale of Kinesiophobia: A Rasch analysis of its properties in subjects with low back and more widespread pain. J Rehabil Med 39:672–678. https://doi.org/10.2340/16501977-0125
38. Hill JC, Tooth S, Coopers V, Chen Y, Lewis M, Walthall S, Saunders B, Bartlam B, Protheroe J, Chudyk A, Dunn KM, Foster NE (2019) Stratified care for patients with back, neck, knee, shoulder or multi-site pain: the STarT MSK Feasibility/Pilot Randomised Controlled Trial (Isrctn15366334). In: Society for back pain research. Sheffield, UK
39. Chiarotto A, Deyo RA, Terwee CB, Boers M, Buchbinder R, Corbin TP, Costa LO, Foster NE, Grotle M, Koes BW, Kovacs FM, Lin CW, Maher CG, Pearson AM, Peul WC, Schoene ML, Turk DC, van Tulder MW, Ostelo RW (2015) Core outcome domains for clinical trials in non-specific low back pain. Eur Spine J 24:1127–1142. https://doi.org/10.1007/s00586-015-3892-3
40. Celestin J, Edwards RR, Jamison RN (2009) Pretreatment psychosocial variables as predictors of outcomes following lumbar surgery and spinal cord stimulation: a systematic review and literature synthesis. Pain Med 10:639–653. https://doi.org/10.1111/j.1526-4637.2009.00632.x
41. Mannion AF, Elfering A (2006) Predictors of surgical outcome and their assessment. Eur Spine J 15(Suppl 1):S93–108
42. den Boer JJ, Oostendorp RA, Beems T, Munneke M, Oerlemans M, Evers AW (2006) A systematic review of bio-psychosocial risk factors for an unfavourable outcome after lumbar disc surgery. Eur Spine J 15:527–536
43. Seebach CL, Kirkhart M, Lating JM, Wegener ST, Song Y, Riley LH 3rd, Archer KR (2012) Examining the role of positive and negative affect in recovery from spine surgery. Pain 153:518–525. https://doi.org/10.1016/j.pain.2011.10.012
44. Katz JN, Stucki G, Lipson SJ, Fossel AH, Grobler LJ, Weinstein JN (1999) Predictors of surgical outcome in degenerative lumbar spinal stenosis. Spine 24:2229–2233
45. Sinikallio S, Aalto T, Airaksinen O, Herno A, Kroger H, Savolainen S, Turunen V, Viinamaki H (2007) Depression is associated with poorer outcome of lumbar spinal stenosis surgery. Eur Spine J 16:905–912
46. Trief PM, Grant W, Fredrickson B (2000) A prospective study of psychological predictors of lumbar surgery outcome. Spine 25:2616–2621
47. Aalto TJ, Malmivaara A, Kovacs F, Herno A, Alen M, Salmi L, Kroger H, Andrade J, Jimenez R, Tapaninaho A, Turunen V, Savolainen S, Airaksinen O (2006) Preoperative predictors for postoperative clinical outcome in lumbar spinal stenosis: systematic review. Spine 31:E648–663
48. Mannion AF, Junge A, Elfering A, Dvorak J, Porchet F, Grob D (2009) Great expectations: really the novel predictor of outcome after spinal surgery? Spine 34:1590–1599
49. Monticone M, Ferrante S, Teli M, Rocca B, Foti C, Lovi A, Brayda Bruno M (2014) Management of catastrophising and kinesiophobia improves rehabilitation after fusion for lumbar spondylolisthesis and stenosis. A randomised controlled trial. Eur Spine J 23:87–95. https://doi.org/10.1007/s00586-013-2889-z
50. Abbott AD, Tyni-Lenne R, Hedlund R (2010) Early rehabilitation targeting cognition, behavior, and motor function after lumbar fusion: a randomized controlled trial. Spine (Phila Pa 1976) 35:848–857. https://doi.org/10.1097/BRS.0b013e3181d1049f
51. Archer KR, Devin CJ, Vanston SW, Koyama T, Phillips SE, Mathis SL, George SZ, McGirt MJ, Spengler DM, Aaronson OS, Cheng JS, Wegener ST (2016) Cognitive-behavioral-based physical therapy for patients with chronic pain undergoing lumbar spine surgery: a randomized controlled trial. J Pain 17:76–89. https://doi.org/10.1016/j.jpain.2015.09.013
52. Steurer J, Nydegger A, Held U, Brunner F, Hodler J, Porchet F, Min K, Mannion AF, Michel B (2010) LumbSten: the lumbar spinal stenosis outcome study. BMC Musculoskelet Disord 11:254. https://doi.org/10.1186/1471-2474-11-254
53. Becker HJ, Nauer S, Porchet F, Kleinstuck FS, Haschtmann D, Fekete TF, Steurer J, Mannion AF (2017) A novel use of the Spine Tango registry to evaluate selection bias in patient recruitment into clinical studies: an analysis of patients participating in the Lumbar Spinal Stenosis Outcome Study (LSOS). Eur Spine J 26:441–449. https://doi.org/10.1007/s00586-016-4850-4
54. Pulkovski N, Mannion AF, Caporaso F, Toma V, Gubler D, Helbling D, Sprott H (2012) Ultrasound assessment of transversus abdominis muscle contraction ratio during abdominal hollowing: a useful tool to distinguish between patients with chronic low back pain and healthy controls? Eur Spine J 21(Suppl 6):S750–759. https://doi.org/10.1007/s00586-011-1707-8
55. Mannion AF, Caporaso F, Pulkovski N, Sprott H (2012) Spine stabilisation exercises in the treatment of chronic low back pain: a good clinical outcome is not associated with improved abdominal muscle function. Eur Spine J 21:1301–1310. https://doi.org/10.1007/s00586-012-2155-9
56. Caporaso F, Pulkovski N, Sprott H, Mannion AF (2012) How well do observed functional limitations explain the variance in Roland Morris scores in patients with chronic non-specific low back pain undergoing physiotherapy? Eur Spine J 21(Suppl 2):S187–195. https://doi.org/10.1007/s00586-012-2255-6
57. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC (2007) Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 60:34–42

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.