

The quality of spine surgery from the patient's perspective. Part 1: The Core Outcome Measures Index in clinical practice

Anne F. Mannion · F. Porchet · F. S. Kleinstück ·
F. Lattig · D. Jeszenszky · V. Bartanusz ·
J. Dvorak · D. Grob

Received: 22 December 2008 / Revised: 2 March 2009 / Accepted: 5 March 2009 / Published online: 25 March 2009
© Springer-Verlag 2009

Abstract The Core Outcome Measures Index (COMI) is a short, multidimensional outcome instrument, with excellent psychometric properties, that has been recommended for use in monitoring the outcome of spinal surgery from the patient's perspective. This study examined the feasibility of implementation of COMI and its performance in clinical practice within a large Spine Centre. Beginning in March 2004, all patients undergoing spine surgery in our Spine Centre (1,000–1,200 patients/year) were asked to complete the COMI before and 3, 12 and 24 months after surgery. The COMI has one question each on back (neck) pain intensity, leg/buttock (arm/shoulder) pain intensity, function, symptom-specific well being, general quality of life, work disability and social disability, scored as a 0–10 index. At follow-up, patients also rated the global effectiveness of surgery, and their satisfaction with their treatment in the hospital, on a five-point Likert scale. After some fine-tuning of the method of administration, completion rates for the pre-op COMI improved from 78% in the first year of operation to 92% in subsequent years (non-response was mainly due to emergencies or language or age issues). Effective completion rates at 3, 12 and 24-month follow-up were 94, 92 and 88%, respectively. The 12-month global outcomes (from $N = 3,056$ patients) were operation helped a lot, 1,417 (46.4%); helped, 860 (28.1%); helped only little, 454 (14.9%); did not help, 272 (8.9%); made things worse,

53 (1.7%). The mean reductions in COMI score for each of these categories were 5.4 (SD2.5); 3.1 (SD2.2); 1.3 (SD1.7); 0.5 (SD2.2) and -0.7 (SD2.2), respectively, yielding respective standardised response mean values (“effect sizes”) for each outcome category of 2.2, 1.4, 0.8, 0.2 and 0.3, respectively. The questionnaire was feasible to implement on a prospective basis in routine practice, and was as responsive as many longer spine outcome questionnaires. The shortness of the COMI and its multidimensional nature make it an attractive option to comprehensively assess all patients within a given Spine Centre and hence avoid selection bias in reporting outcomes.

Keywords COMI · Outcome · Spine surgery · Clinical practice

Introduction

Few would dispute the necessity for quality control measures to monitor the effectiveness and safety of treatments delivered to patients in routine clinical practice. However, the issue of how “quality” should be defined, and by whose standards it should be judged, is less clear. The proportion of positive outcomes after spinal surgery depends to a large extent on the manner in which outcome is assessed (see review [11]), and there is no single, universally accepted method. In the past, clinicians typically judged the outcome from their own perspective, using simple rating schemes such as “excellent, good, moderate, and poor”. The technical success of the operation also lent itself to evaluation by means of sophisticated imaging at follow-up. However, most of the time, these measures proved to be only weakly associated with outcomes of relevance to the patient and to society [6].

A. F. Mannion (✉)
Spine Center Division, Department of Research and
Development, Schulthess Klinik, Lengghalde 2,
8008 Zurich, Switzerland
e-mail: anne.mannion@kws.ch; anne@annemannion.com

F. Porchet · F. S. Kleinstück · F. Lattig · D. Jeszenszky ·
V. Bartanusz · J. Dvorak · D. Grob
Spine Center, Schulthess Klinik, Zurich, Switzerland

It is now accepted that the focus should be placed on patient-orientated measures and that the patient should be the main judge of outcome, with the result that clinician-based methods have been superseded by a diverse range of patient self-assessment questionnaires. However, the emergence of so many new instruments, some of which have not been fully validated [19], and the lack of their standardised use have served to compromise meaningful comparisons across studies and patients. In recognition of this, a standardised set of outcome measures for use with back patients was proposed in 1998 by a multinational group of experts [6]. There was general consensus that the most appropriate core outcome measures should include the domains pain, back specific function, generic health status (well-being), work disability, social disability and patient satisfaction [4, 6]. Accordingly, the group proposed a parsimonious set of six questions that would cover each of these domains, yet be brief enough to alleviate respondent burden, and hence be practical for routine clinical use and quality management. Three separate research groups have now examined the psychometric properties of this “Core Outcome Measures Index” (COMI) and have, between them, documented its reliability, validity, sensitivity to change and ability to be predicted by known risk factors [7, 13, 14, 18]. It has since been argued that the availability of simple instruments such as the COMI should help encourage clinicians to collaborate with national and international registries [2]. However, the performance of the COMI in relation to its utility for everyday quality control purposes in typical practice settings has yet to be evaluated.

The aim of the present study, described in two parts, was to examine the feasibility of using the COMI for quality management purposes in a large Spine Centre. Part 1 details the practicalities of its implementation, including the difficulties encountered and the solutions devised to overcome these. It also documents the COMI scores recorded before surgery and at regular intervals up to 2 years later. Finally, it compares the constructs “global effectiveness of the treatment” and “satisfaction with treatment of the back problem”, in an attempt to tease out possible subtle differences in these two retrospective patient-rated measures of quality. Part 2 then goes on to establish the minimal clinically relevant change score for both improvement and deterioration in “everyday practice”.

Methods

The study group comprised all German or English speaking patients undergoing spine surgery in the Spine Center of our hospital from March 2004.

The SSE Spine Tango registry, supported by an in-house custom-made database, was used to document the relevant data. The surgical form, which was completed at various stages from admission through to discharge, enquired about pathology, previous treatment, patient morbidity status, surgical details and surgical complications. These data will not be further discussed in the present article, except to say that the distribution of Spine Tango “main pathology” categories in the available data was as follows: 78.9% degenerative disease, 6.7% spondylolisthesis, 4.9% deformity, 3.5% failed surgery, 2.0% tumour, 1.6% fracture/trauma, 0.8% inflammation, 1.6% others; in the group with degenerative disease, decompression alone was carried out in 56.7%, decompression and fusion in 36.5%, fusion alone in 5.0%, and none of these in 1.8%. Before and 3, 12 and 24 months after surgery, patients were requested to complete a questionnaire containing the multidimensional COMI [7, 13, 14]. The COMI comprises a series of questions covering the domains of pain (back and leg/buttock pain intensity, each measured separately on a 0–10 numeric graphic rating scale; or for cervical spine patients, neck and arm/shoulder pain, respectively), and function, symptom-specific well-being, general quality of life, social disability and work disability (each on a 5-point Likert scale). At follow-up, in addition to the COMI questions, there were four questions inquiring about re-operations, the occurrence and nature of any complications that had arisen following surgery (this is dealt with by Grob et al. [9] in this issue and will not be considered further here), overall satisfaction with treatment of the back/neck problem in the hospital (5-point Likert scale ranging from “very satisfied” to “very dissatisfied”) and the global effectiveness of surgery (“how much did the operation help your back/neck problem?; 5-point Likert scale from “helped a lot” to “made things worse”).

The pre-operative questionnaire was sent to the patient at home, along with the information about their forthcoming hospital stay, and they were asked to complete it and hand it in during admission. Completion of the questionnaire at home ensured that the information provided by the patient was free of any care-provider influence. For the same reasons, the follow-up questionnaires were sent out by post, along with a stamped-addressed envelope for them to be returned to the Hospital’s Research Unit.

Statistical analyses

Descriptive data are presented as mean \pm standard deviations (SD).

The COMI sum score was calculated as described in the original validation paper [12]; briefly, the items that scored 1–5 [function, symptom-specific well-being, general QOL,

disability (average of social and work disability)) were first re-scored on a 0–10 scale (raw score -1 , multiplied by 2.5). These items and the pain score (the highest value out of leg pain and back pain; already scored 0–10) were then averaged to provide a COMI index score ranging from 0 to 10.

Repeated measures analysis of variance (ANOVA) was used to examine the significance of the change in mean scores from pre-surgery to 3, 12 and 24 months post-surgery. All remaining analyses were done on the pre-op and 12 month post-op data, since this provided the largest group at a sufficiently long follow-up.

Effect sizes were calculated as the standardised response mean by taking the mean of the individual change scores and dividing this by the corresponding standard deviation of these change scores [3]. The effect size was also calculated in relation to the five categories of the “global effectiveness of surgery” question. The correlation between the instrument change-scores and the (ordinal) global outcome scale gave a further indication of responsiveness [17].

The analyses were conducted using Statview 5.0 (SAS Institute Inc, San Francisco, CA, USA) and statistical significance was accepted at the $P < 0.05$ level.

Results

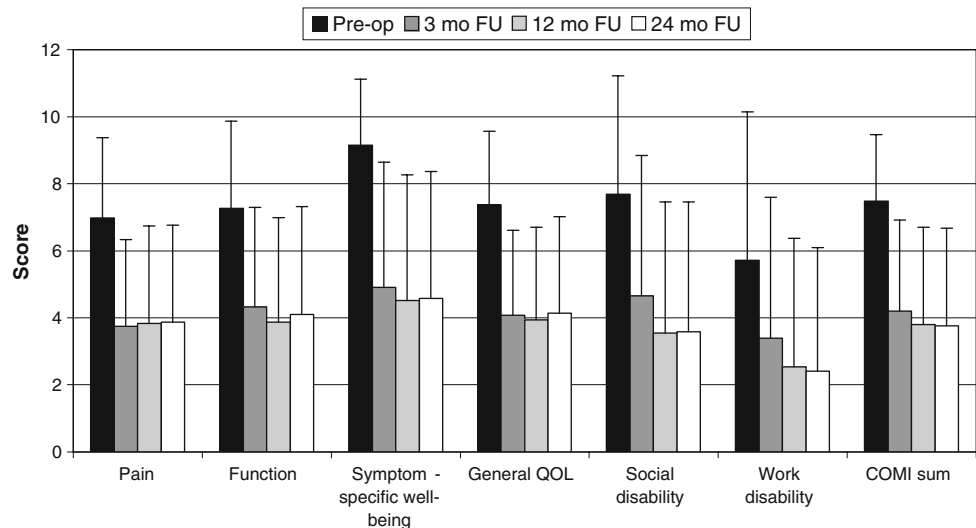
Feasibility of implementation

During the first year of implementation, the pre-operative COMI questionnaire was completed by approximately 78% of all patients undergoing surgery. This rather suboptimal completion rate was the result of difficulties in the administration of the system, and inadequate control mechanisms. Our first attempts involved sending the questionnaire with the appointment for the operation and asking patients to deposit the questionnaire at the reception upon admission; whatever questionnaires were deposited were then forwarded to the research department. It transpired that some questionnaires turned up on the wards, some were posted by the patient to the research or clinical department (either before or after the hospital stay), and some were not returned at all. It was apparent that a tighter system of control was necessary, to include a check of the questionnaires that were actually due each day; in this way, any missing forms could be tracked down more easily before the operation. The system was hence optimised as follows: (1) The questionnaire was sent out in the same way (i.e. by the central hospital appointments office), but a list of ID numbers of all the patients sent a questionnaire was handed to the research department each day. A case was then opened for the patient in a custom-built database,

created in Filemaker Pro (Filemaker Inc, Santa Clara, CA, USA). (2) The hospital reception was asked to mark on their admissions list all the patients scheduled for spine surgery that day, and, when those patients arrived, to explicitly ask them for their questionnaire. If a patient arrived with no questionnaire or an incompletely filled-out questionnaire, the research department was called to come and deal with the patient. (3) The research department had a copy of the operation list for the following day, and in the late afternoon, when all admissions were complete, the research assistant picked up the questionnaires from the reception, cross-checking that all patients on the list had handed in a questionnaire. In this way, only patients who categorically refused to complete the questionnaire for any reason, individuals with language or learning difficulties, small children, and the occasional emergency admissions were not integrated into the system (although their operation and case details were still recorded in the local database, such that follow-up could still be attempted, where appropriate). With this tighter control from all sides, the pre-operative completion rate rose to 93% in the second year, and remained at $\sim 92\%$ in subsequent years. This is the system currently in use within our Spine Centre today. The patients that were missed by the system in the first year were slightly younger than those that were captured (55.7 ± 18.5 years vs 58.8 ± 16.0 years, respectively; $P = 0.03$), but there was no difference in their gender distribution (53.5 vs 54.1% female, respectively; $P = 0.88$). Further, their scores at the first follow-up did not differ significantly (3-month COMI score, 4.3 ± 2.8 vs 4.4 ± 3.0 , respectively). We hence consider it unlikely that the overall findings were severely biased by these missing baseline data.

The questionnaire data were immediately entered into the database, along with other patient-relevant details and the admission date. Approximately 1 week after admission, the database was checked to ensure that all patients who were admitted actually went on for surgery. Only at this point were the operation date and other brief details (indication, operation and surgeon identity) entered into the system and the patient counted as a “case”. This prevented the data from non-operated patients from remaining unnoticed in the system and patients who did not ultimately undergo surgery from unintentionally being sent a post-operative questionnaire. The due dates for the 3, 12 and 24-month follow-ups were automatically calculated in the database. Searches could then be done of given date-ranges, on a weekly basis, to identify all patients in need of a follow-up questionnaire that week. The follow-up questionnaires were sent from and returned to the research department by post. Non-responders were sent a reminder/called by phone. They were given the option of completing the short questionnaire over the telephone, and especially

Fig. 1 Scores for each of the COMI domains and the COMI sum-score before surgery and 3, 12 and 24 months after surgery. In each case, significant reductions ($P < 0.05$) were observed from pre-surgery to 3 months post-surgery, after which no further significant changes occurred. Data at each time point show the mean and SD from the same group of patients (those with scores at every time point ($N = 1,582$)) analysed with repeated measures ANOVA



encouraged to do so if the research assistant perceived that they were unlikely to post the questionnaire back. A “comments” box in the database allowed the research assistant to note these reminders and hence recognise “persistent offenders” and be aware of their likely behaviour in future follow-ups.

The effective compliance at follow up (i.e., proportion of those actually sent a questionnaire) currently stands at 94% for the 3-month FU, 92% for the 12-month FU, and 88% for the 24-month FU.

Outcome scores: COMI

Figure 1 shows the typical mean COMI scores, recorded in each outcome domain before surgery and at 3, 12 and 24 months after surgery for patients with all of these follow-ups (as of March 2008). For all domains, there was a significant reduction in scores from pre-surgery to 3-months post-surgery, with the values then remaining stable up to 2 years post-surgery.

The multidimensional COMI sum-score reduced from 7.5 (SD 2.0) before surgery to 4.2 (SD 2.7) at 3 months, 3.8 (SD 2.9) at 12 months and 3.8 (SD 2.9) at 24 months. The corresponding effect sizes (standardised response mean) were 1.14 at 3 months, 1.22 at 12 months and 1.22 at 24 months.

On an individual basis, there was a highly significant correlation between the change in COMI score recorded after 3 months and that recorded after 12 months ($r = 0.68$, $P < 0.0001$; Fig. 2), and between the change recorded after 3 months and that recorded after 24 months ($r = 0.61$, $P < 0.0001$; data not shown). In other words, the early outcome was a good predictor of the longer-term outcome.

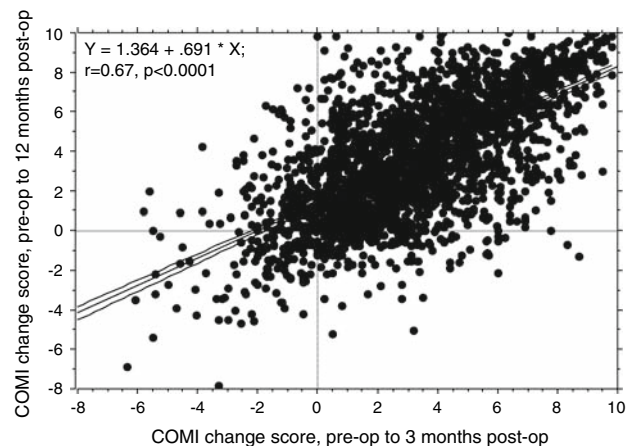


Fig. 2 Relationship between the change score for the COMI from baseline to 3-months’ follow-up and from baseline to 12 months’ follow-up ($N = 2,600$ cases)

Outcome scores: treatment effectiveness and satisfaction with medical care

At the 12-month follow-up ($N = 3,056$ patients), the distribution of answers for treatment effectiveness (“how much did the surgery help your back/neck problem?”) was as follows: operation helped a lot, 1,417 (46.4%); helped, 860 (28.1%); helped only little, 454 (14.9%); did not help, 272 (8.9%); made things worse, 53 (1.7%). The proportions in each outcome category did not differ between the lumbar and cervical patients.

Two thousand one hundred and eighty-five (71.3%) patients declared being very satisfied with the overall treatment of their back/neck problem in our hospital, 421 (13.8%) satisfied, 228 (7.4%) neither satisfied nor dissatisfied, 138 (4.5%) dissatisfied and 91 (3.0%) very dissatisfied.

Satisfaction and treatment effectiveness were highly related (Spearman Rho = 0.64, $P < 0.0001$), but there were also some incongruous findings. For example, in the group of “very dissatisfied” patients, ~20% nonetheless reported that the operation had “helped or helped a lot”. Conversely, in the group of “very satisfied” patients ~9% had actually declared that the operation had “helped only little”, and ~5% that it “had not helped” or had “made things worse”. Qualitative investigation of a random selection of these discrepant cases revealed, in the case of “dissatisfied patients with a good surgical result”, issues with certain aspects of the nursing care, the attitude of the treating doctor, the time taken to get to the point of a good result, etc., and, in the case of patients with “high satisfaction but a poor result”, appreciation in relation to these same aspects (e.g. the caring nature of the nurses and doctors, the efforts invested to try to bring about a good result, etc.).

Outcome scores: COMI change scores in each outcome category

The COMI score before and after surgery (12 months FU) for each global outcome category is shown in Fig. 3.

The mean reductions in the COMI score from pre-op to 12 months post-op (max possible reduction = 10 points) for each outcome category were: 5.4 points (SD 2.5) for the group “helped a lot”, 3.1 (SD 2.2) for “helped”, 1.3 (SD 1.7) for “helped only little”, 0.5 (SD 2.2) for “didn’t help”, –0.7 (SD 2.2) for “made things worse”. Hence, the change score and the effect size was greatest for patients in the best outcome category (effect size 2.16) diminishing as one progressed down to “didn’t help” (effect size 0.23).

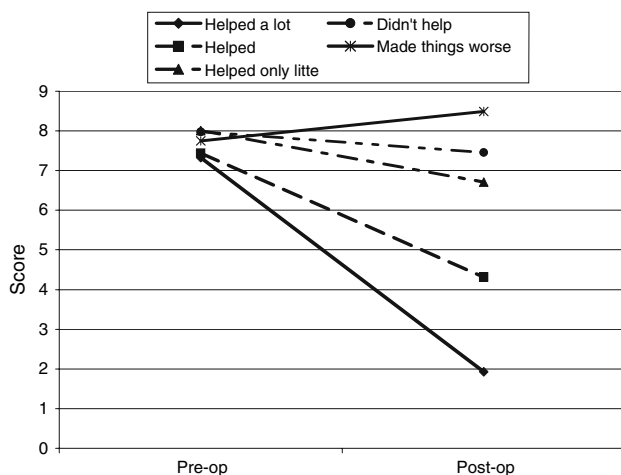


Fig. 3 COMI sum-score change for each outcome category 12 months after surgery

Discussion

Feasibility of the system

The present study showed that it was perfectly feasible to implement the COMI, as a multidimensional patient-oriented instrument, alongside the daily clinical practice of a busy Spine Centre dealing with over 1,000 cases per year. However, it requires the establishment of a “dedicated person” (e.g. secretary, study nurse, research assistant; full-time or part-time depending on the size of the practice) whose job it is to keep track of the patients and their questionnaire data.

Some teething difficulties were experienced in the first year, whilst trying to establish a workable system within the existing infrastructure of the hospital, but it was ultimately possible to produce a smooth-running prospective evaluation of multidimensional outcome in almost every single patient.

Although the system was based on the SSE Spine Tango registry, we found it essential to create an additional database to manage the follow-up for the patient questionnaires and to maintain a clear overview of the system. The database allowed the easy identification of patients who were due a questionnaire, had failed to respond, died, been re-operated, or moved abroad, etc., and provided an easily accessible overview of the patient’s status, specific spine problem, language, etc., which assisted in any future contacts with the patient. It is to be hoped that, in the not too distant future, the SSE Spine Tango system itself will offer such an application as a supplement to the existing registry. A system of work was also developed for dealing with patients who returned a questionnaire expressing extreme dissatisfaction: they were contacted first by the research assistant (who explained that the call was being made “independently”, from the Research Department, and in the interests of quality management) and asked whether they would like to elaborate on the situation and would like a further appointment to be made for them with the treating surgeon. This had the benefit that it prevented the patients from feeling they were simply submitting feedback that would never really be examined or acted on in relation to their own case, yet it respected (to an extent of their own choosing) their “anonymity” as far as their treating clinician was concerned. Possibly, this resulted in sustained compliance with the whole project (greater conviction in the system on behalf of the patient) and hence higher follow-up rates than might be expected in a registry with direct submission of the data and no “human contact”. Undoubtedly, it also resulted in a more accurate and less biased depiction of longer-term outcome for the whole group of patients—albeit including a higher proportion of poor outcomes—because those with a poor result did not

just disappear out of the system in disgust, after submitting their first negative follow-up rating.

Outcome measures: COMI scores

An interesting finding in the present study was that the mean scores in all outcome domains remained pretty stable from the 3-month follow-up onwards (Fig. 1), and on an individual basis there was a strong correlation between the change scores for successive follow-up intervals (Fig. 2). In other words, the early result was highly predictive of the longer-term outcome. This has been reported before, in relation to surgery for degenerative diseases of the lumbar spine [1, 10, 11, 15], though never in such large numbers. The vast majority of patients undergoing surgery in our Spine Centre do so in connection with degenerative disorders (see Grob et al. [9] this supplement), and surgery typically serves a “mechanical” purpose, aiming to relieve pain by removing a physical obstruction (e.g. by decompression in the case of spinal stenosis/herniated disc) or stabilising an unstable segment (e.g. by fusion in the case of spondylolisthesis). Hence, as far as the main symptoms are concerned, the success (or otherwise) of the operation should be evident relatively early on. In this respect, the indiscriminate insistence in the scientific literature on a minimum 2-year follow-up for the publication of spine-surgery outcome studies may in many instances be inappropriate, and may result in a critical phase of the follow up being overlooked. There may be exceptions, of course, in the case of, e.g. longer fusions for degenerative scoliosis, and we will look at these in more detail when the group sizes are large enough for sub-group analyses. However, in general, as the early post-operative results appear to herald the longer-term outcome, the practical conclusion from our findings is that a ‘wait and see policy’ in patients with a poor initial outcome is not advocated. The systematic documentation of early results is imperative for the development of timely re-interventions (either conservative or operative) in unimproved patients, in an attempt to avert the development of chronic “failed back” problems and long-term disability. Obviously this needs to be dealt with on a case-by-case basis: if a surgically remediable cause for the continuing symptoms can be found, then re-operation should be considered; if not, then the initiation of some kind of pain management programme might be indicated.

Outcome scores: global treatment effectiveness and satisfaction with medical care

The patient’s perception of the global benefit of surgery is typically assessed in terms of either the “overall effectiveness of treatment” or “satisfaction with the treatment delivered”. Whilst these two constructs are clearly related,

they are not synonymous [5, 8], and, as the results of the present study showed, they sometimes throw up incongruous responses on an individual basis, where satisfaction with care is not accompanied by a corresponding improvement in the condition being treated, and vice versa. The use of these two different indices can also lead to different proportions of “good results” being declared for a given group of patients. In the present study, “success” seemed to be higher when the question was phrased in terms of satisfaction (85.2% satisfied/very satisfied) as opposed to the effectiveness of the treatment (74.5% operation helped/helped a lot). Most likely, the rating of satisfaction is heavily influenced by the patient-provider relationship and includes an expression of appreciation for the surgeon “having done his best”, even if the final result is not ideal. In contrast, the “effectiveness of treatment” rating focuses quite clearly on the therapeutic improvement, in terms of how much the surgery helped the back problem. These subtle differences should be borne in mind when interpreting the outcome results of different studies. Satisfaction with care is an extremely important outcome to the treatment provider and is an integral part of internal quality control. We hence strongly recommend this item for inclusion in the outcome assessment. However, its relevance as an objective outcome measure outside of the institution in which it is assessed, e.g. in the context of multicentre trials or for use in meta-analyses, may be limited. It is recommended that, especially in studies that seek to examine the quality of new or alternative techniques/implants, and in systematic reviews of a given treatment method, the focus be placed on the effectiveness of the treatment.

Outcome measures: COMI scores and global effectiveness of surgery

The literature is replete with studies reporting the outcome of surgery in terms of the proportion of patients with a “good” or “poor” result. Dichotomised outcomes like these are usually built by collapsing the data from multiple-category items in which responses range from extremely positive to extremely negative. However, a proportion of patients will always declare a middling result and—though the decision can heavily influence the overall success rate declared for a given procedure—it is often not clear to which main outcome group these should belong. The present data shed some light on this phenomenon and provide some indication as to where the “line should be drawn”. In the original test-retest studies to determine the measurement error associated with the COMI, the “minimal *detectable* change” for the index was reported to be 1.7 points [12]. In other words, for score changes less than 1.7 it is difficult to distinguish with any certainty between

“real change” and measurement error. In the present study, the mean change in COMI score was 1.3 for the outcome group “helped only little”. We hence recommend that in future studies in which outcomes must be dichotomised, only the top two ratings “helped a lot” and “helped” should be considered as a “good” outcome, and the category of “helped only little” and lower—or similar categories that reflect only negligible improvement—should be considered as “poor”. This concurs with the findings of other groups, using similarly ranked global effectiveness scales [16]. Intuitively, and especially for elective surgical procedures where the main therapeutic goals are symptom relief and functional improvement, this categorisation would also seem to be appropriate; a “slight improvement” can scarcely justify the time and effort, risks and costs of the procedure, for either patient or care-provider alike.

General comments, limitations of the study, and directions for the future

The present study was intended to show how to collect, and what can be learnt from, huge and highly representative datasets gathered within the framework of a quality management scheme using a short but psychometrically sound outcome instrument. In presenting the data, we made no attempt to split up the patient population in relation to specific diagnoses, types of intervention, gender, age-groups, etc. Such analyses should, however, be performed once the register has grown sufficiently to provide adequately sized sub-groups. Future studies should examine whether the instrument is as responsive in these different sub-groups.

The COMI has been cross-culturally adapted in a growing number of languages; the psychometric properties of these different language versions and the practicability of their use in hospitals with different infrastructures, in different countries, should be subject to further investigation.

Conflict of interest statement None of the authors has any potential conflict of interest.

References

- Amundsen T, Weber H, Nordal HJ, Magnaes B, Abdelnoor M, Lilleas F (2000) Lumbar spinal stenosis: conservative or surgical management?: a prospective 10-year study. *Spine* 25:1424–1435. doi:10.1097/00007632-200006010-00016
- Balagué F, Mannion AF, Pellise F, Cedraschi C (2007) Clinical update: low back pain. *Lancet* 369:726–728. doi:10.1016/S0140-6736(07)60340-7
- Beurskens AJHM, de Vet HCW, Köke AJA (1996) Responsiveness of functional status in low back pain: a comparison of different instruments. *Pain* 65:71–76. doi:10.1016/0304-3959(95)00149-2
- Bombardier C (2000) Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. *Spine* 25:3100–3103. doi:10.1097/00007632-200012150-00003
- Carlson J, Youngblood R, Dalton JA, Blau W, Lindley C (2003) Is patient satisfaction a legitimate outcome of pain management? *J Pain Symptom Manage* 25:264–275. doi:10.1016/S0885-3924(02)00677-2
- Deyo RA, Battie M, Beurskens AJHM, Bombardier C, Croft P, Koes B, Malmivaara A, Roland M, Von Korf M, Waddell G (1998) Outcome measures for low back pain research. A proposal for standardized use. *Spine* 23:2003–2013. doi:10.1097/00007632-199809150-00018
- Ferrer M, Pellise F, Escudero O, Alvarez L, Pont A, Alonso J, Deyo R (2006) Validation of a minimum outcome core set in the evaluation of patients with back pain. *Spine* 31:1372–1379. doi:10.1097/01.brs.0000218477.53318.bc (discussion 1380)
- George SZ, Hirsch AT (2005) Distinguishing patient satisfaction with treatment delivery from treatment effect: a preliminary investigation of patient satisfaction with symptoms after physical therapy treatment of low back pain. *Arch Phys Med Rehabil* 86:1338–1344. doi:10.1016/j.apmr.2004.11.037
- Grob D, Kleinstuck F, Lattig F, Porchet F, Bartanusz V, Jeszenszky D, Mannion AF (2009) The patient’s perspective on complications after spine surgery. *Eur Spine J* (this issue)
- Hakkinen A, Ylinen J, Kautiainen H, Airaksinen O, Herno A, Kiviranta I (2003) Does the outcome 2 months after lumbar disc surgery predict the outcome 12 months later? *Disabil Rehabil* 25:968–972. doi:10.1080/0963828031000122258
- Mannion AF, Elfering A (2006) Predictors of surgical outcome and their assessment. *Eur Spine J* 15(Suppl 1):S93–S108. doi:10.1007/s00586-005-1045-9
- 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. doi:10.1007/s00586-005-0911-9
- Mannion AF, Denzler R, Dvorak J, Muntener M, Grob D (2007) A randomised controlled trial of post-operative rehabilitation after surgical decompression of the lumbar spine. *Eur Spine J* 16:1101–1117. doi:10.1007/s00586-007-0399-6
- 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. doi:10.1007/s00586-006-0255-0
- McGregor AH, Hughes SPF (2002) The evaluation of the surgical management of nerve root compression in patients with low back pain. Part 1: The assessment of outcome. *Spine* 27:1465–1470. doi:10.1097/00007632-200207010-00018
- Ostelo RW, de Vet HC (2005) Clinically important outcomes in low back pain. *Best Pract Res Clin Rheumatol* 19:593–607. doi:10.1016/j.berh.2005.03.003
- Stratford PW, Spadoni G, Kennedy D, Westaway C, Alcock GK (2002) Seven points to consider when investigating a measure’s ability to detect change. *Physiother Can* 54:16–24
- White P, Lewith G, Prescott P (2004) The core outcomes for neck pain: validation of a new outcome measure. *Spine* 29:1923–1930. doi:10.1097/01.brs.0000137066.50291.da
- Zanoli G, Strömqvist B, Padua R, Romanini E (2000) Lessons learned searching for a HRQoL instrument to assess the results of treatment in persons with lumbar disorders. *Spine* 25:3178–3185. doi:10.1097/00007632-200012150-00013