

Charles G. Greenough

## Outcome assessment: recommendations for daily practice

Received: 1 November 2005  
Accepted: 1 November 2005  
Published online: 29 November 2005  
© Springer-Verlag 2005

C. G. Greenough  
Consultant Spinal Surgeon,  
James Cook University Hospital,  
Marton Road, TS4 3BW  
Middlesbrough, UK  
E-mail: Charles.Greenough@stees.nhs.uk  
Tel.: +44-1642-854311  
Fax: +44-1642-282649

**Abstract** The choice of instruments for the assessment of outcome in spinal surgery is bewildering. For day-to-day practice, however, consideration of the purpose for which information is required allows construction of simple strategies for data collection. Recommendations are made for short and convenient data sets for use in personal audit, clinical governance, benchmarking, patient selection and business planning. No simple data set can measure in detail every aspect of

practice, but use of these recommendations will provide information that will be of great value to the spinal surgeon and ultimately to his patients.

**Keywords** Outcome measurement · Spinal surgery · Spine · Outcome · Assessment · Review · Low-back pain · Low-back outcome score · Oswestry disability index · Roland and morris questionnaire · SF36 · EuroQol

### Introduction

The detailed analysis of the various outcome assessment techniques contained within the papers of this special edition are thought provoking but in addition they fill the ordinary spinal surgeon with despair. Within these papers 82 back specific disability scales [15], six quality of life measurements [17], 24 pain scales [10] and ten or more work related outcome measures [4] have been reviewed. The validation of each of these is highly variable and none satisfy every criterion that has been suggested. All these instruments have advantages and all have flaws. Confounding factors exist in all the domains of measurement and in any case the individual patient over a follow-up period of 2 years will be undergoing a large number of changes unrelated to the disease or its treatment, which will potentially have an impact on the measurement instrument chosen.

The purpose of this review is not to expound all these arguments again, but to provide some framework that

the ordinary spinal surgeon can use to decide how he is going to take measurements in his spinal practice. It is not designed to assist in the selection of instruments for major research projects. For such projects measurement instruments will be finely tuned to the exact needs of the specific research. A number of factors will assist in the choice of instruments for daily use.

### Who wants to know?

It is clear that data gathered from routine spinal practice will be of interest to many different constituencies and it is essential that the eventual end user be considered from the start. The interest of each constituency will be in different outcome domains. For example society at large will be less interested in the individual satisfaction of a patient but would like to see a reduction in the support an individual might require and would certainly be interested if an individual was able to return to work and again become a contributing member of society

(Table 1). Thus if the data was principally collected to support a bid for funding of a particular treatment, participation and work related measures would be most appropriate together with a health related quality of life measure.

### How will the data be collected?

In daily practice it is absolutely essential that data collection is simple, rapid and undemanding of surgical, nursing, secretarial and clerical resources. It must also allow the expression of patient's true opinion as far as possible without any perceived pressure from the clinical team. In daily practice it is impractical to have an independent observer gather this information from the patient who should be provided with a private area to complete the selected assessment.

The most effective method of data collection is the use of an interactive computer based questionnaire. In spinal practice as well as in many other areas of medicine the use of computers with aids such as touch screens have been found to be very acceptable for patients [21]. There is also the advantage that no transcription of data is required. Further, the programme can be arranged so that it is impossible to miss out questions of scored sets, a not infrequent problem with paper-based collection. The patient can complete questionnaires in their own time without the need for supervision which saves the staff time and also allows the patient the feeling of complete privacy when answering questions. Web based applications for the patient to complete in their own home are available but compliance has not yet been fully reported and there is variable degree of computer literacy within the population. From the patient's perspective and for use in busy out patient setting the length of total computer interview will require careful consideration.

### The patient population

There is no measurement instrument that will perform well in every patient population. Patients in surgical clinics will generally express more pain and disability than patients attending physiotherapy. The patient population will inform the choice of instrument to avoid possible floor or ceiling effects.

**Table 1** The importance of various outcome domains to the constituency involved

|               | Patient | Family | Society |
|---------------|---------|--------|---------|
| Satisfaction  | ++      | +      |         |
| Disability    | ++      | ++     | +       |
| Participation | ++      | ++     | ++      |

The age of the population will also have impact. There are no disability measures in back pain described for children and for the elderly population questions relating to some domains, for example participation in work, are less apposite.

### When should data be gathered?

Baseline measure is an absolute requirement prior to treatment. This will allow comparison of severity and will also allow the question of threshold of treatment to be addressed. Following surgical treatment it has been traditional to maintain a 2-year follow-up period. However, as Mannion suggests [14], for surgery such as decompression for spinal stenosis it would seem that a period such as 6 months would be enough to allow the end result to be observed. Some reports indicate that the 2 years results are accurately predicted by the 1-year results. It would appear logical to allow a sufficient period of time to pass for the objective of surgery to be achieved. Thus fusion may require a longer period of follow-up than discectomy.

If the longevity of implants such as total disc replacement is of interest, then very long-term follow-up will be necessary. The practical difficulties of recalling patients for review or achieving compliance in postal or web-based follow-up are significant.

### Domains of measurement

There is increasing agreement that for comprehensive assessment of outcome a number of different domains need to be addressed. These include:

- Pain
- Back specific disability
- Participation
- Work related disability
- Health care usage
- General health status and quality of life
- Satisfaction

To measure all of these domains in daily practice presents a considerable challenge in terms of the number of instruments required and the time taken to complete them. Which of these domains are most appropriate for any individual surgeon's practice depends largely on the purpose for which the surgeon wishes to collect the data.

Attempts have been made to suggest ways of covering these domains in a minimum data set. The core set proposed by Deyo et al. [3] covers five domains in five individual questions (pain, function, symptomatic well being, back related disability and disability (social role)) together with an optional question on satisfaction. The separation

of “function” and “back related disability” is not completely clear. This set has not yet been evaluated. A modification of this set has proven responsive and correlates well with other established instruments [13]. However no results in patient groups are available at present.

Patient satisfaction with the procedure has been proposed as a global outcome measure [11].

However, there is no universally accepted format for the assessment of work related disability or patient satisfaction.

### Work related disability

As described by Elfering [4], there are a number of assessment tools that may be used to address work related disability. However these are generally more suited to research projects. The two most important for simple assessments are work status and sickness absence.

Work status must distinguish subjects not working because of back-related disability from those not working from other causes. No generally accepted assessment of work status has been published but the question from the low-back outcome score (LBOS) [9] is part of a validated outcome measure and covers the important considerations [1] (Table 2).

The difficulties of measuring sickness absence have been detailed by Elfering [4] in this supplement, together with detailed recommendations based on the suggestions of Hensing [12]. For day-to-day use the consultation of records is impractical and for assessment in an individual a measure such as days lost from work in the previous follow-up period may be quick, simple and effective. The Deyo core set [3] directly provides such a measure of sickness absence. The modification of Mannion [13] will do so but only if the work status is already known.

### Satisfaction

It is important to recognise that there are at least two variables within global satisfaction. The first is the patient satisfaction with the care process, the way in which their surgeon approached them and their

**Table 2** Assessment of work status

| Regarding your back pain    |  |
|-----------------------------|--|
| At present, are you working | Full time at your usual job ()<br>Full time at a lighter job ()<br>Part time ()<br>Not working—disability ()<br>Housewife ()<br>Retired (not disability) ()<br>Unemployed ()<br>Student () |

perception of how the surgeon tried to help them. The second area of satisfaction lies with their assessment of the actual effect of the treatment on their condition. These have previously been shown to be quite separate, but if only one question is provided then the opinion of the actual effects of treatment is contaminated by the satisfaction with the treatment process and treatment provider, in effect an expression of gratitude [16]. It is, therefore, important to have at least two questions, so that each of these areas may be addressed separately.

Satisfaction with overall care may be addressed as a visual analogue scale with anchors of “completely dissatisfied” and “completely satisfied” or as a Likert scale as in the “core set”. If a Likert scale is used for the results of the effect of treatment, then this should be balanced around “no change” (Fig. 1). The modified core set proposed by Mannion et al. [13] does provide separation of satisfaction with the process of care from satisfaction with the results of the treatment itself. However, the responses on the Likert scale for the results of treatment are unbalanced with only one option being clearly worse than before and the mid point implying improvement. The scale is therefore biased to improvement.

### Why is the data being collected?

This is the most important question that the surgeon has to address. The purpose for which the data is being gathered will make a significant contribution to the decision as to which instruments to utilise. The recommendations made here are based on the purpose for which the data is to be collected.

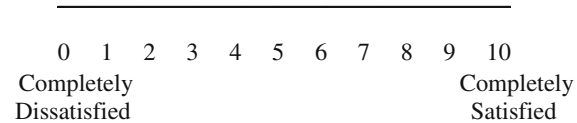
### Recommendations

#### Audit of personal practice

The data may simply be for personal use and the audit of personal practice. Such data would also be valuable when informing patients of the results of treatment in their particular surgeon’s hands. A simple and effective data set would comprise a visual analogue pain scale, a simple measure of back-related disability, work disability and a satisfaction rating. The first three elements are contained within the LBOS [9] that is short and has been well validated. The sum score will also allow some comparison with the literature where it has been previously used and also directly compared to other well-used instruments such as the Oswestry Disability Index (ODI) and the Roland and Morris Questionnaire (RMQ). The LBOS performs better than the ODI in patients with lesser disabilities and may be used in conservatively treated as well as operatively treated patients.

**Fig. 1** Example of a satisfaction measure

Please mark on the line below how satisfied you are with the overall assessment and care you received for your back or leg pain :-



Please indicate how your back or leg condition has changed as a result of your treatment.

---      ---      ---      ---      ---      ---      ---  
 Worse than I    Much    A bit    No change    A bit    Much    Completely  
 could imagine    Worse    Worse               Better    Better    Better

The ODI [6] and RMQ [19] are also well validated and widely used disability questionnaires but would require the addition of further instruments to provide information on employment, participation and health care usage.

In addition to these it may be desired to examine generic health status or quality of life. The SF 36 [24] and the EuroQol (EQ-5D) [5] are both well validated and widely reported. The EQ-5D has the additional advantage of being short.

The core set of Deyo [3] represents another option for use, but many surgeons would require some assessment of work status to enable patient counselling. The satisfaction question must address both components described above, but the modification by Mannion et al is unbalanced in favour of a good response to therapy. If audit against an external standard is required then results using the “core set” are generally unavailable at present.

The programme used by the author for computerised collection of simple audit data is available free of charge on application.

#### Comparison with the literature and benchmarking

If the surgeon wishes to bench mark his practice against published results then the use of a widely used instrument would be recommended. The ODI and RMQ have both been widely used and have been recommended by a previous review [2]. The ODI appears to have better responsiveness in more severely disabled patients where the RMQ might be preferred for a less disabled population. The LBOS performs satisfactorily in surgically and conservatively treated patients but has been less widely utilised.

When using the ODI or RMQ in addition the visual analogue pain scale would be required together with an assessment of work related disability and perhaps health care usage. These data are already contained within the LBOS, which therefore represents a compact option.

Comparison of results with the literature prior to treatment will allow the surgeon to assess his own threshold for treatment, although the usual caveats of homogeneity of patient population and consideration of confounding factors will apply.

All these instruments will need the addition of an instrument such as the EuroQol if any financial aspects are to be considered.

Most surgeons would also employ a simple satisfaction scale addressing both satisfaction domains.

The core set of Deyo has yet to be employed in many studies and as yet will not allow benchmarking. However, subject to the caveats above, it may well be more widely employed in the future.

#### Clinical governance and benchmarking

It is increasingly important to address the issues of clinical governance within individual practice. It is here the use of a register is most appropriate as it will enable the surgeon to bench mark his practice against his peers. To this end a National Registry would be the instrument of choice and satisfy governance requirements.

If the use of a registry is selected then the individual surgeon will have little individual choice of the outcome measures utilised which are normally agreed by consensus amongst the instigators and users of the register. The who? when? and where? of data entry will require close consideration. If a national registry is not available then use of Spine Tango, supported by the Spine Society of Europe, should be considered.

To undertake governance and bench marking without use of a register the surgeon should consider the recommendations in “comparison with the literature” above.

### Predicting results and patient selection

Mannion, in this issue, has addressed this important aspect. Although a number of factors that are not directly related to the disease or treatment have significant impact on outcome, it is not yet possible to confidently withhold surgical treatment on the basis of these predictive factors alone. However, they can be of great value in informing discussion of the likelihood of success in an individual patient.

Simple and effective data collection would include compensation, anxiety and depression (see confounding factors below) and duration of sickness absence.

### Long-term follow-up

Long-term results are particularly important in the assessment of implants such as total disc replacement. Failure and revision are important end points but, as with total joint replacement, occur late. To be certain that a particular prosthesis is performing badly, for example, requires larger numbers than are available in one centre. All implants such as total disc replacement should be entered into a registry, as this is the most effective surveillance tool available at present.

The methodical use of two or three simple assessment tools as in “audit” above will allow longer term results to be obtained on an “ad hoc” basis in any practice. This will be of great value for audit, but is no substitute for a well-designed prospective research study if a consistent time line is required.

### Support for funding/business case

Many surgeons now will have need to present a business case in support of new or existing treatments. Such business cases are built partly on the available literature, but local purchasing authorities are always most interested in local services.

A back related disability scale will be important, but general health and quality of life measures, work related disability and patient satisfaction will be crucial to demonstrate cost effectiveness or cost utility [22].

The LBOS contains work status but sickness absence must be added. It also contains information on health care utilisation. The ODI and RMQ are widely used but require addition of work status and sickness absence and contain no health care utilisation data. In addition the

EQ-5D could be utilised together with a pain scale (VAS) and a satisfaction assessment.

### Specific outcomes

In day-to-day practice, patients are often counselled prior to surgery in terms of simple specific activities or results; e.g., walking distance in spinal stenosis and leg pain in prolapsed intervertebral disc. Surgeons use these factors because they are thought to be specific to the pathology being treated. It would, therefore, seem appropriate to use such measures as part of a day-to-day measurement of practice. There is no agreed format for these questions and in the main they would remain useful only within a specific clinic. However, in the assessment of outcome of an individual patient, such measures can be extremely valuable.

### Confounding factors

It is clear that a number of factors not directly related to the disease or its management have a substantial impact on the overall outcome in an individual [14]. Compensation, anxiety or depression, coping strategies, fear avoidance, socio-economic group and work related factors have all been demonstrated in surgical treatments to have variable influence on results. In addition to outcome measures described, careful consideration will be given to assessment of such factors.

Socio-economic group is derived usually from occupation, which can easily be made available, but there is a cost in time and expertise converting occupation to socio-economic group.

The question of compensation may be determined in one or two simple questions (Table 3). It appears important to distinguish lump sum compensation from wage replacement [7].

Anxiety and depression are conveniently assessed by a combination of the MSPQ and Zung that can either be summed and a cut-off applied [8] or reported as the DRAM [18]. The hospital anxiety and depression questionnaire is also short and convenient [25].

The factors above are critical in assessment, but in addition the fear avoidance beliefs questionnaire

**Table 3** Assessment of compensation staTUS

|   |        |
|---|--------|
| Are you claiming legal compensation for your back pain? | Yes/No |
| If so, has the case been settled?                       | Yes/No |
| Are you receiving any sick pay or disability benefits?  | Yes/No |



(FABQ) [23] or the Coping Strategies Questionnaire [20] could be considered if collection time permits.

## Conclusion

There is no solution to the problems of day-to-day data collection that is perfect. There are, however, a number

of perfectly serviceable options that can be recommended for day-to-day practice. As can be seen in the detailed analysis contained in this supplement no recommendation will find favour in every quarter. However, the use of such simple strategies will allow a very substantial amount of analysis to be carried out in an individual practice and in the end will significantly benefit both the surgeon and his patients.

## References

- Amick BC III, Lerner D, Rogers WH, Rooney T, Katz JN (2000) A review of health related work outcome measures and their uses, and recommended measures. *Spine* 25:3152–3160
- Bombardier C (2000) Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. *Spine* 25:3100–3103
- 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
- Elfering A (2005) Work-related outcome assessment instruments. *Eur Spine J* (in press)
- EuroquoL-Group (1990) EuroquoL—a new facility for the measurement of health-related quality of life. *Health Pol* 16:199–208
- Fairbank J (1995) Use of Oswestry Disability Index (ODI). *Spine* 20–13:1535–1537
- Greenough CG (1993) Recovery from low back pain. 1–5 year follow-up of 287 injury related cases. *Acta Orthopaedica Scandinavica (Suppl)*. S254(vol 64):1–33
- Greenough CG, Fraser RD (1991) A comparison of eight psychometric instruments in low back pain. *Spine* 16:1068–1074
- Greenough CG, Fraser RD (1992) Assessment of outcome in patients with low-back pain. *Spine* 17–1:36–41
- Haefeli M, Elfering A (2005) Pain assessment. *Eur Spine J* (in press)
- Hagg O, Fritzell P, Oden A et al (2002) Simplifying outcome measurement: evaluation of instruments for measuring outcome after fusion surgery for chronic low back pain. *Spine* 27:1213–1222
- Hensing G, Alexanderson K, Allebeck P, Bjurulf P (1998) How to measure sickness absence? Literature review and suggestion of five basic measures. *Scand J Soc Med* 26:133–144
- 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* (in press), published online: 4 June 2005
- Mannion A, Elfering A (2005) Predictors of surgical outcome and their assessment. *Eur Spine J* (in press)
- Müller U, Ruder C, Greenough CG (2005) Back-related outcome assessment instruments. *Eur Spine J* (in press)
- Murray MM, Holmes M, Greenough CG (2002) Results of a spinal assessment clinic. *J Bone Joint Surg* 84B(Suppl II):144
- Németh G (2005) Health related quality of life outcome instruments. *Eur Spine J* (in press)
- Main CJ, Wood PLR, Hollis S, Spanswick CC, Waddell G (1992) The distress and risk assessment method a simple patient classification to identify distress and evaluate the risk of poor outcome. *Spine* 17:42–52
- Roland M, Morris R (1983) A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. *Spine* 8–2:141–144
- Rosenstiel AK, Keefe FJ (1983) The use of coping strategies in chronic low back pain patients: relationship to patient characteristics and current adjustment. *Pain* 17:33–44
- Schaeren S, Bischoff-Ferrari HA, Knupp M, et al (2005) A computer touch-screen version of the North American Spine Society outcome assessment instrument for the lumbar spine. *J Bone Joint Surg Br* 87:201–204
- van Tulder M, Boos N (2005) Economic evaluations—a new avenue of outcome assessment in spinal disorders. *Eur Spine J* (in press)
- 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
- Ware JE, Sherbourne CD (1992) The MOS 36-item short-form health survey (SF-36). I Conceptual framework and item selection. *Med Care* 30–36:473–483
- Zigmond AS, Snaith RP (1983) The hospital anxiety and depression scale. *Acta psychiatr scand* 67:361–370