

13–17

Is there value in routine administration of outcome questionnaires?

Authors Julie Agel, Richard Bransford

Institution Orthopaedic and Sports Medicine, Harborview Medical Center, Seattle, WA, USA

Final class of evidence-prognosis

Study design	
RCT	
Cohort	
Case control	
Case series	•
Methods	
Patients at similar point in course of treatment	
F/U ≥ 85%	
Similarity of treatment protocols for patient groups	•
Patients followed up long enough for outcomes to occur	•
Control for extraneous risk factors	NA
Overall class of evidence	IV

The definition of the different classes of evidence is available on page 63.

ABSTRACT

Study design: Retrospective case-series of prospective routinely collected Short Form 36v2 (SF-36v2) data in an outpatient spine clinic.

Objective: To determine if there is value in routine administration of outcome questionnaires for the patient and/or clinician without a targeted population or monitoring of follow-up.

Method: Retrospective review of Health Outcomes Scoring database.

Results: During an 18-month period 1,863 patients completed 3,124 SF-36v2 questionnaires. The extent of diagnoses and the uneven timing of follow-ups of the completed questionnaires render this data useless both to the patient and for aggregate analysis.

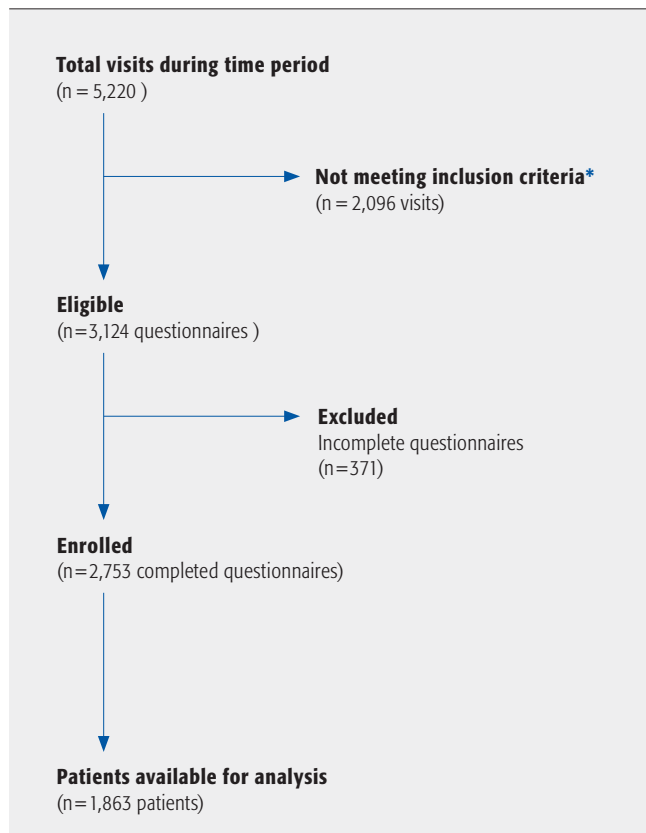
Conclusion: There is no value in routine administration of outcome questionnaires to patient care or clinical research.

The authors have no conflict of interest for this study.
No funding was received for research.

STUDY RATIONALE

The routine use of patients' outcomes scoring systems has become common in various specialties, such as spine surgery. The goal of such undertakings is to get a better understanding of patient well-being from a patient perspective as it is impacted by disorders, diseases, and hopefully being helped by interventions. To get a better handle on patient outcomes it has become normal to routinely collect these outcomes data at every clinic visit to hopefully enable a retrospective analysis at a later date. This practice is increasingly doubtful due to the cost involved and the incomplete nature of meaningful outcomes data actually collectable. In our outpatient spine clinic beginning in 2000, a set menu of accepted Health-Related Quality of Life (HRQOL) questionnaires was given to all patients for all visits. Despite aggressive promotion and monitoring of this data collection during 2 years (2003–2004), this may not be an effective use of patient time and clinic resources when critically reviewed for value and economic cost.

Fig 1 Patient sampling and selection.



OBJECTIVE

To determine if there is any value in the administration of outcome questionnaires designed to track a patient's recovery after treatment when they are given as part of routine clinical care [1].

METHODS

Study design: Retrospective audit of prospectively collected SF-36v2 data [2] in an outpatient spine clinic during a fixed interval. We recorded completed questionnaires, date of completions, and quality of patient responses during the audit time to assess if any meaningful conclusions could be drawn from these data gatherings.

Inclusion criteria: Presence of a scorable SF-36v2 completed between September 2008 and February 2011 in the outpatient spine clinic at the University of Washington, Seattle (**Fig 1**).

Exclusion criteria: Incomplete SF-36v2 submitted to the data collection center with too many missing responses to score based on the criteria established by the Medical Outcome Study (n = 371).

Patient population: Patients with outpatient appointments in the spine clinic at the University of Washington between September 2008 and February 2011.

Outcomes: Proportion of patients with three or more completed SF-36v2.

Analysis: Descriptive.

RESULTS

- During the study period we identified 5,220 individual visits to our spine clinic, all of which were expected to be accompanied by completed paper-based SF-36v2 and then electronically scanned questionnaires.
- We identified satisfactory completion of 3,124 SF-36v2 questionnaires (60% of expected) from 1,863 distinct patients.
- 2,744 (88%) of 3,124 eligible SF-36v2 had enough questions completed to allow for actual scoring of the physical summary scale.
- 2,753 (88%) of 3,124 SF-36v2 had enough questions completed to allow for actual scoring of the mental health summary scales.
- Only 300 of the 1,863 patients had three or more completed SF-36v2 questionnaires during the course of treatment.
- A review of patients who had completed four questionnaires (n=63) demonstrated that the ranges around the average results demonstrate the misleading nature of data collection outside of fixed predetermined time (**Table 1**). Each patient's index questionnaire was verified to be his/her baseline questionnaire.
- Evaluating the questionnaire completion by quartiles (**Table 2**) shows that (with no relationship to any known intervention: surgical or pharmaceutical or rehabilitative) 25% of patients had completed their four questionnaires within 154 days of their first questionnaire; thus the data gathering was the accumulation of clinically irrelevant information.

Table 1 Average and range of questionnaires by patient count.

	Average	Earliest	Latest
Questionnaire 2	87 days	7 days	474 days
Questionnaire 3	180 days	23 days	632 days
Questionnaire 4	311 days	49 days	735 days

Table 2 Questionnaire completed by quartiles.

	25th%	50th%	75th%
Questionnaire 2	28 days	56 days	105 days
Questionnaire 3	63 days	140 days	245 days
Questionnaire 4	154 days	278 days	426 days

DISCUSSION

- To measure a change in function, administration of HRQOL tests, such as the SF-36v2 delivered through forms or questionnaires, should also have a corresponding time spread. This is especially relevant for spine surgery where baseline data is most importantly obtained during the first visit and a delay of at least 3 months for a first postoperative HRQOL test is recommended and 1-year follow-up or more is vital for research.
- Patients with multiple questionnaire administrations over a short period do not provide useful information as (1) most regularly used HRQOL questionnaires do not have enough sensitivity to measure relevant weekly change and (2) functional status does not undergo a meaningful change in most cases.
- Patients with questionnaire administration at 'random' intervals based on time to return to clinic schedules, as opposed to pre-established intervals, means that those with similar diagnosis or treatments cannot be grouped in a useful manner to create population-based data.
- According to an audit of clinical records, patients in this study frequently had multiple unnecessary pre-treatment questionnaires over short periods for visits pertaining to diagnostic workup, or preoperative clinic visits as opposed to appropriate time-gaps allowing for obvious clinical improvement after treatment.
- Patients in this review had multiple postoperative questionnaires in the immediate postoperative time representing discharge and 3- and 6-week visits; time when functional change is not anticipated or measurable with HRQOL techniques, such as the SF-36v2.
- In addition to not receiving useful data, there is a substantial economic cost to the distribution and collection of outcomes forms and data. In an era when outcomes data are increasingly becoming part of daily clinical life it is imperative that gathering of HRQOL data be done in a cost-efficient and meaningful way [3].
- The SF-36v2 is a commonly used outcome tool in spine surgery studies. However, if chosen, it should be administered in the setting of a well-thought question/study protocol and be undertaken at particular periods when clinical change is expected as opposed to an attempt at routine administration at every clinic visit. Overall, the use of SF-36v2, while considered synonymous with good-quality outcomes research, is currently undergoing a critical reevaluation relative to its responsiveness to spine surgery [4].

Strengths

- We surveyed a large sample of functional status evaluations of patients seen in an outpatient spine clinic.
- The importance of a periodic review of routine data collection cannot be overemphasized to verify value of the data.

Limitations

- The SF-36v2 proprietary data entry program does not allow fields, such as patient age and time from diagnosis or treatment to be added into the system so an additional data entry program must be created or the SF-36v2 data merged into an existing file which may require additional expertise, data collection, and expense.
- Patients usually will return to clinic when appropriate and not according to a predetermined time, which makes un-directed outcome data gathering subject to problems with not only missing data but also data not clinically meaningful.
- Forty percent of patients seen did not return a completed SF-36v2. It is possible that a more complete response rate would improve the utility of the data received.

SUMMARY AND CONCLUSION

- Routine administration of HRQOL questionnaires in a regular outpatient spine clinic does not seem to yield meaningful data retrieval.
- A viable plan is needed to identify what type of outcomes data is being collected at which specific intervals before actually deploying such data-gathering efforts. The HRQOL questionnaires should only be used based on demonstrated ability of the technique in question to actually reflect the condition or intervention to be studied. The SF-36v2 or its derivatives are mainly designed for impact studies of major diseases and may not be useful if applied for specific conditions and procedures that are commonly investigated in spine surgery.
- It seems advisable for any HRQOL data-gathering process, regardless of in-study or routine format, to regularly critically audit the data-collection process to ensure that initially set goals are being met and that it is cost-effective and truly meaningful.
- Clinical and ancillary staff must be involved in HRQOL data gathering and understand the goal of handing out functional outcomes questionnaires.
- There are clear differences in methodology applied for data collected solely for individual patient follow-up or marketing purposes versus data collected for practice review of patient recovery or research protocols. The latter usually follow a far more stringent protocol and can be expected to yield more meaningful results rather than the ‘scatter approach’ used by the former.
- Future developments with routine electronic data capture through web-based applications, when available, and more automated-auditing prospects may change the outlook of routine application of HRQOL in daily clinical practices.
- A primary limitation of routine outcomes data gathering is that it does not capture those patients who fail to return for subsequent care; creating bias in any aggregate data analysis.

A more detailed description can be found in the Web Appendix at www.aospine.org/ebsj.

REFERENCES

1. **Deyo RA, Patrick DL** (1989) Barriers to the use of health status measures in clinical investigation, patient care, and policy research. *Med Care*; 27(3):S254–S268.
2. Quality Metric Health Outcomes Scoring Software 3.0. Copyright 2004–2009.
3. **Mirza SK, Campbell B, Ching RP, et al** (1998) Influence of comorbid medical conditions on pretreatment. *J Surg Outcomes*; 1(1):31–40.
4. **Chapman JR, Norvell DC, Hermsmeyer JT, et al** (2011) Evaluating common outcomes for measuring treatment success for chronic low back pain. *Spine (Phila Pa 1976)*; 36(21 Suppl):S54–68.

EDITORIAL PERSPECTIVE

Reviewers welcomed the question posed by the authors and described it as an ‘overdue critical review’ of the widely held practice to reflexively collect patient outcomes data, such as the SF-36v2, at every outpatient visit in hopes to possibly garner something useful from this data pile at some future time. The reviewers challenged the authors on a number of points:

- *Are there any scenarios where routine HrQoL data collection would be useful? Would it not seem almost ‘standard of care’ to document patient well-being in functional, pain and psychological dimensions from the patient’s perspective at every visit?*
- *Are there any less onerous tests present or emerging, which may be less cumbersome and more meaningful to carry out routinely than the SF-36v2? Of more involved tests, such as the SF-36v2, should perhaps only preoperative baseline and 1 and 2 year intervals be collected?*
- *What specific research questions should be asked prior to deciding to collect HrQoL data?*
- *Finally, the reviewers pointed out that the analysis provided was largely observational and lacked more elaborate statistical analysis – such as an investigation into which patient tended to be ‘completers’ versus ‘non-completers’ and the potential influence of success of treatment on compliance with questionnaires.*

While these were all valid points we allowed the study to proceed with publication based on its ‘strength in numbers’ but hope that the authors will be stimulated to show how to improve cost efficient and meaningful data gathering for routine spine surgery patients beyond pure research studies.