
 METHODOLOGIC ISSUES

The SF-36 health survey: a valid measure of changes in health status after injury

Branko Kopjar

Abstract

Objectives—The aim of this study is to evaluate the criterion validity and responsiveness to changes over time of the Medical Outcome Study Short Form 36 (MOS SF-36) measure.

Methods—A consecutive sample of 775 patients 16 to 78 years treated for an unintentional injury at the hospital or emergency clinic in Drammen, Norway was selected for the study. Data about activity restrictions and health status measured by SF-36 were obtained by a postal questionnaire 6–10 weeks after the injury. A follow up survey was sent 24–28 weeks later to all who reported activity restriction at the time of the first survey. Fifty two of these replied (63%).

Results—469 patients responded to the survey questionnaire and of these, 82 experienced some restriction of activity. These scored lower ($p < 0.01$) on all eight SF-36 health dimensions (physical functioning, social functioning, role limitation (physical), role limitation (emotional), bodily pain, mental health, vitality, and general health) than the 387 patients without activity restriction. Scores on physical functioning, social functioning, role limitation (physical), bodily pain, and vitality significant improved ($p < 0.01$) among the 52 patients who were followed up. Scores on the other dimensions, however, showed no significant changes over time.

Conclusion—The MOS SF-36 appears to be a valid instrument, responsive to changes in health status over time among unintentionally injured adult people. Thus it may be possible to use the SF-36 to describe changes in health due to injury. The applicability of this or similar measures for injured children remains to be established.

(*Injury Prevention* 1996; 2: 135–139)

Keywords: SF-36, follow up studies, injury sequelae, disability.

Injury is a sudden event resulting in an acute decline of health. Knowing the type and amount of this decline is important for several

reasons. It may lead to improvements in the provision of health and social services for the injured, and to more rational allocation of resources for injury prevention.

Evaluating the impact of injury on health is a complex task in part because injuries represent a large number of diagnoses and because patients with the same diagnosis may differ in the severity of their injury. Moreover, some patients sustain multiple injuries. Although most recover completely from falls, for example, a minority have long lasting disabilities. Similarly, some patients with limb injury who experience disability improve over time.¹ Conversely, some injuries produce long lasting consequences that affect several dimensions of health. Physical sequelae and psychological distress were found among victims of traffic accidents as long as two years after the injury.² Due to these variations, classifying injuries according to the severity of their consequences is a complex task.

A useful measure of decline in health due to injury should provide a common descriptor for various injury diagnoses; it should be responsive to changes in health status over time; and it should be able to describe changes in several health dimensions.

Classification systems for severity/consequence of injury

Several classification systems describe specific types of injury related effects on health. The abbreviated injury severity (AIS) score classifies injuries according to their immediate threat to life.³ In contrast, the injury impairment score (IIS) categorizes injuries according to long term (residual) impairment,⁴ whereas the functional capacity index (FCI) is based on long term impact on functional capacity.⁵ Although useful for many purposes, these scales have many limitations. The AIS, primarily a threat to life scale, is not sufficiently sensitive to classify injuries according to impairment or functional limitation.⁶ Empirical studies of the IIS have resulted in concerns about its validity,⁷ and the FCI has still not been subjected to extensive evaluation. All these scales ignore the multidimensional nature of health; they suffer from low discriminatory value (most patients have ceiling scores)¹; they are insensitive to changes in health over time; and they represent a professionals' view of the problem rather than the patient's.

Department of
Population Health
Sciences, National
Institute of Public
Health, Oslo, Norway

Correspondence to:
Dr Branko Kopjar, Senior
Research Scientist, National
Institute of Public Health,
Department of Population
Health Sciences,
Geitmyrsveien 75, 0462 Oslo,
Norway.

Generic measures of health and health related quality of life

Generic measures of health status and health related quality of life reflect the dimensions of health conceptualized by the general public. There is a growing interest in the development and application of such measures.⁸ Extensive reviews providing the theoretical and conceptual basis for the development of these measures are available.^{9,10} Unfortunately, the development of similar instruments suitable for children is still in progress.¹¹

Currently, the most commonly used generic measure of health status is the Medical Outcome Study Short Form 36 health survey (MOS SF-36).^{12,13} The origin of this is the Rand Corporation's health insurance experiment questionnaire, originally consisting of 108 items.¹⁴ Although the original instrument was valid and sensitive, it was too long for general use. Two shorter versions appeared useful — the SF-20 and SF-36 and currently the latter is the most popular.

The SF-36 is a self administered questionnaire consisting of 36 items, requiring 5–10 minutes to complete. The 36 items generate a profile of scores across eight dimensions of health: physical functioning (10 questions), social functioning (two questions), role limitations (physical problems) (four questions), role limitation (emotional problems) (three questions), mental health (five questions), vitality (four questions), pain (two questions), general health (five questions). One question addressing health change (transition) is not scored nor represented in the eight dimensions. Several studies have shown that the SF-36 has high internal consistency, good criterion validity, replicability and reliability.^{15–17} The evidence about the responsiveness of the SF-36 to changes in health status over time is still inconclusive.¹⁸ Two studies found it to be responsive to such changes^{19,20} while one other found the opposite.²¹

Generic health measures are potentially useful for evaluating health decline due to injury and for describing several dimensions of

health. They may provide a common descriptor for various types of injuries and permit injuries to be compared with other diseases. Dollar figures can be attached to changes in health status thus facilitating economic evaluations. Despite these many appealing characteristics, the applicability of the SF-36 for injured people has not previously been evaluated.

The aim of this study is to evaluate the validity of the SF-36 health survey and its ability to describe changes in health status among injured people. More specifically, this study aims to determine if the SF-36 can differentiate between injuries resulting in various degrees of activity restriction (criterion validity), and if the SF-36 is responsive to changes in health status over time among injured adults.

Methods

SAMPLE AND DATA COLLECTION INSTRUMENT

Cases were selected prospectively from patients seen for unintentional injury at the Central Hospital and the emergency clinic in Drammen, Norway. The inclusion criteria were: age between 16 and 78 years; main diagnosis ICD-9 800 to 995 (except codes 905–909: late consequences of injuries); the injury was unintentional, and the visit to the hospital/emergency clinic was the first visit for that injury. We recruited a consecutive sample of 829 people injured during the three month period ending 15 May 1994.

We administered a mailed questionnaire consisting of the following: demographic data; length, type, and duration of injury related activity restriction (disability), the SF-36,¹³ and the EuroQol²² (another generic measure of health status). As no Norwegian version of the SF-36 was available at the time of this study we made our own translation. The translation used in this study was marginally different from the official translation that appeared later as a part of the international effort.²³

The addresses of 54 patients (7%) could not be located. The 775 patients (829–54) were mailed the questionnaire six weeks after the injury. Non-respondents were sent a reminder four weeks later. Replies were received from 469 patients (61%). Of these, 82 patients were on sick leave, could not perform household tasks, or could not participate in leisure time activities at the time when they answered the questionnaire. These 82 patients were sent a follow up questionnaire at 24 weeks after the injury, with a follow up to non-respondents four weeks later. The follow up questionnaire was of the same form and content as the first questionnaire. Answers were received from 52 patients (63%). Of these, 24 still suffered from some type of activity restriction (work, household services, or leisure time activities), whereas 28 did not, that is, were free from activity restriction.

ANALYSIS

Scores for the SF-36 health dimensions were calculated according to the MOS SF-36 coding

Table 1 Activity restriction status 6–10 weeks after injury

	No activity restriction (n = 387)	Any type of activity restriction* (n = 82)	
		Non-respondents (n = 30)†	Respondents (n = 52)†
Age (years)			
Mean	41.0	44.8	46.9
SD	18.5	16.1	18.2
Sex			
Female	182	12	20
Male	205	18	32
Occupation			
Employed	210	18	30
Unemployed	22	2	0
Housewife	24	3	4
Student	52	1	4
Retired	70	6	12
Other	6	0	0
No answer	4	0	2
Education			
Primary	78	8	16
Secondary	179	11	24
University	76	6	9
No answer	53	5	5
Treatment			
Hospitalized	56	17	37
Non-hospitalized	331	13	15

*On sick leave or complete or partial restriction to perform household services or participate in leisure time activities. †Follow up at 24–48 weeks.

scheme.²⁴ We performed three groups of analyses: first, we evaluated differences in the health dimension scores between the 389 patients without activity restriction and the 82 with activity restriction at the time of the first survey. This involved analysis of variance, using age as a covariate. The aim of this analysis was to assess criterion validity. Second, we examined changes in the scores over time (6–10 weeks *v* 24–28 weeks) among the 52 patients who were followed up. This analysis was by activity restriction status at 24–28 weeks (24 patients with *v* 28 patients without activity restriction). The goal here was to evaluate responsiveness of the SF-36 to changes in health status over time. For this we used analysis of variance for repeated observations. Responsiveness was expressed as a standardized response mean (SRM),¹⁹ that is, the mean change score divided by the standard deviation of changes in scores so that a higher SRM indicates greater sensitivity to change. Finally, we analyzed differences in the SF-36 dimensions between the 389 patients without activity restriction at 6–10 weeks and the 28 patients with such restrictions at that time, but not 24–28 weeks after the injury. Again, analysis of variance was used, in this instance to evaluate relative sensitivity of the SF-36 to changes in health status among injured persons.

Results

Patient demographic data are shown in table 1. Patients suffering from activity restriction were somewhat older than those without activity restriction (46.1 and 41.0 years on average, respectively) ($p < 0.05$). Fourteen per cent of patients without activity restriction were hos-

pitalized compared with 66% of patients with activity restriction ($p < 0.01$). The 52 respondents and 30 non-respondents to the follow up questionnaire did not differ significantly on any of the characteristics in table 1.

Distribution of patients by diagnosis is shown in table 2. Two diagnoses dominate among cases with activity restriction — fractures and sprain/strains.

SF-36 DIMENSIONS 6–10 WEEKS AFTER THE INJURY

Scores on SF-36 health dimensions are shown in table 3. Patients with activity restriction had statistically significantly ($p < 0.01$) lower values for all eight dimensions than patients without activity restriction. Differences were largest for dimensions physical functioning, social functioning, role limitation (physical), role limitation (emotional), and bodily pain. The differences in mental health, vitality, and general health were less pronounced.

RESPONSIVENESS TO CHANGES OVER TIME

Changes in the SF-36 among the 52 patients with activity restriction 6–10 weeks after the injury are shown in table 4. The table shows *p* values for differences and for interaction. The *p* (difference) is the statistical significance of the difference between the values observed 6–10 weeks and 24–28 weeks after the injury. The *p* (interaction) shows the likelihood that the amount of change in particular SF-36 dimension (6–10 weeks *v* 24–28 weeks) is equal among both groups of patients (that is 24 patients still having activity restriction, and 28 patients free from activity restriction at 24–28 weeks).

As shown, several dimensions improved significantly ($p < 0.01$), while others did not. The amount of improvement in physical functioning, social functioning, and vitality was similar among the 24 patients with and 28 patients without activity restriction at 24–28 weeks (p (interaction) > 0.05 , for each dimension). Likewise, the improvement in role limitation (physical) was, as expected, statistically significantly (p (interaction) < 0.01) higher among patients who were free of activity restriction than among those who still suffered from activity restriction (SRM 0.82 and 0.30, respectively). Finally, the improvement in bodily pain was higher (p (interaction) < 0.01) among patients without activity restrictions compared with those with such restrictions (SRM 0.96 and 0.15, respectively).

Table 2 Distribution of patients by diagnosis; values are number (%)

	No activity restriction (n = 387)	Any type of activity restriction* (n = 82)	
		Non-respondents (n = 30)	Respondents (n = 52)
Fracture	97 (25)	17 (57)	29 (56)
Dislocation	3 (1)	—	—
Sprain/strain	127 (33)	8 (27)	9 (17)
Brain concussion	13 (3)	—	6 (12)
Internal injury	3 (1)	—	1 (2)
Open wound	64 (17)	1 (3)	4 (8)
Superficial injury	5 (1)	—	—
Contusion	29 (7)	1 (3)	—
Foreign body	33 (9)	—	—
Burn	7 (2)	—	—
Multiple injury	3 (1)	3 (10)	3 (6)
Other	3 (1)	—	—
Total	387 (100)	30 (100)	52 (100)

*On sick leave or complete or partial restriction to perform household services or participate in leisure time activities.

Table 3 SF-36 health dimensions at 6–10 weeks after injury by activity restriction status

	No activity restriction				Any type of activity restriction						
	No	Mean	SD	% at ceiling	% at floor	No	Mean	SD	(%) at ceiling	(%) at floor	<i>p</i> Value
Physical functioning	373	81.7	27.3	45	3	79	44.8	25.6	0	4	<0.01
Social functioning	379	79.8	24.1	41	2	82	54.0	24.7	11	2	<0.01
Role limitation (physical)	357	72.1	40.0	62	18	74	13.9	30.7	8	80	<0.01
Role limitation (emotional)	350	83.4	32.3	76	9	70	44.5	47.0	39	49	<0.01
Bodily pain	380	66.4	26.8	26	1	82	37.7	22.6	2	6	<0.01
Mental health	368	71.4	15.3	1	0	80	63.5	16.0	0	0	<0.01
Vitality	367	60.7	17.5	0	0	81	52.3	19.6	0	0	<0.01
General health	385	72.3	23.7	11	1	82	63.5	26.3	2	0	<0.05

Table 4 Average values for SF-36 health dimensions 6–10 weeks and 24–28 weeks after injury among the 52 patients with activity restriction 6–10 weeks after the injury

	Activity restriction status 24–28 weeks after injury							
	Any type of activity restriction* (n = 24)			No activity restriction (n = 28)			p Value	
	6–10 weeks	24–28 weeks	SRM§	6–10 weeks	24–28 weeks	SRM§	Difference†	Interaction‡
Physical functioning	34.3	45.1	0.54	50.5	71.2	1.11	<0.01	<0.05
Social functioning	44.3	47.4	0.14	57.9	70.8	0.50	<0.05	<0.05
Role limitation (physical)	5.7	10.2	0.30	22.8	51.5	0.82	<0.01	<0.01
Role limitation (emotional)	35.1	35.1	0.00	49.2	60.3	0.27	<0.05	<0.05
Bodily pain	34.5	37.6	0.15	39.1	61.3	0.96	<0.01	<0.01
Mental health	58.5	57.6	0.05	67.9	67.0	0.08	<0.05	<0.05
Vitality	45.3	48.1	0.15	55.6	61.5	0.50	<0.05	<0.05
General health	57.4	53.8	0.21	69.2	70.7	0.07	<0.05	<0.05

*On sick leave or complete or partial restriction to perform household services or participate in leisure time activities. †Difference between values observed 6–10 weeks and 24–28 weeks after the injury. ‡Interaction effect between the amount of change in the score (6–10 v 24–28 weeks) and the activity restriction status at 24–28 weeks (with and without restriction). §Standardized response mean: mean change in score divided by the standard deviation of individuals' changes in scores.

We also compared SF-36 dimensions among the 28 patients who were free of activity restriction at 24–28 weeks, and the 389 patients who did not already have restricted activity at the first measurement (6–10 weeks after the injury). These results are not shown in a separate table. At 6–10 weeks, the 28 patients with activity restriction scored lower than the 389 patients without activity restriction in physical functioning, social functioning, role limitation (physical), role limitation (emotional), and bodily pain ($p < 0.01$ for each dimension). Eighteen weeks later, however, the differences in physical functioning, social functioning, and bodily pain were no longer evident, but differences in role limitation (physical), and role limitation (emotional) persisted.

Discussion

This study suggests that the SF-36 health survey can be used to describe changes in health status due to injury. This important finding may lead to a wider application of generic measures of health among injured people. The limitations of this measure are, however, still numerous. The SF-36 generates eight scores, each representing one health dimension. It is still unresolved how these could be combined to generate a single index of health.¹⁶ Most questions in the measure refer to health status in the last four weeks. To avoid confusion between health before and after an injury, the instrument should be administered at the earliest five to six weeks after injury. As shown in this study, by that time most injured people will have recovered or show no further injury related decline in health. Further research on the applicability of the SF-36 among injured patients needs to address these and other limitations.

This study documents the responsiveness of the SF-36 to changes in health over time. Along with recovery measured by cessation of activity restriction, the scores for SF-36 dimensions improved. The amount of improvement paralleled the healing process. Patients whose activity level was fully restored showed more improvement on the SF-36 dimensions than other patients. This confirms the findings of two earlier studies showing the responsiveness to changes in health status of the SF-36^{17,18} but

differs from one study that found the opposite.¹⁹

Our study also provides further evidence of criterion validity of this measure. People with activity restriction at 6–10 weeks after the injury, scored significantly lower on all eight SF-36 health dimensions than those without activity restriction. This corroborates findings from other studies showing that the SF-36 can reliably differentiate among the groups of patients with different levels of health.¹⁵

Often, activity restriction is used as an indicator of consequences or severity of injury, particularly in cost calculations. It has even been proposed that injuries be classified according to the amount of activity limitation they produce.¹ This study, however, suggests that activity limitation is only a crude measure of injury consequences. People who showed no changes in their activity limitation status (that is, those still suffering from activity limitation at 24–28 weeks after injury) nevertheless showed significant improvements in some SF-36 dimensions. On the other hand, people who returned to normal activity still scored lower on several health dimensions than those free from activity restrictions. This suggests that 'activity restriction' or 'disability' represent less sensitive indices to describe the consequences of injury than the SF-36.

We found a significant polarization among patients with activity restriction according to role limitation (emotional) — approximately 40% of the patients had ceiling scores, and 50% had baseline scores. Other studies found that the emotional response to injury is dichotomous — some people react highly while others do not. One study found that type of emotional response to injury depends more on life circumstances before injury than upon objective severity of the injury.²⁵

LIMITATIONS

There are several possible limitations to our study. At the time it was conducted, there was no official translation or validation of the SF-36 in Norway. We made our own translation of the SF-36 which differed somewhat from the official translation that appeared later. Studies from other countries, however, indicate that small linguistic differences do not influence significantly the results obtained from similar

generic health measures.⁸ Another possible limitation is the influence of cultural differences upon concept validity. The SF-36 was developed in the US and later successfully applied in the UK. However, it is unknown to what extent the concepts represented in the SF-36 also apply to the Norwegian population. Among those without activity restriction in our study, the mean values for health perception, vitality, and mental health were similar to values for a sampe population in Great Britain.^{13 15}

The response rate of 61% is somewhat lower than that obtained in the UK studies,^{13 15} but similar to rates obtained in other mailed surveys performed in Norway. The rate for the follow up survey (63%) is also similar to what is obtained in other such studies.¹⁴ The design of this study ensures that this reply rate does not pose a serious threat to validity of our conclusions.

Conclusions

The SF-36 health survey appears to be valid instrument responsive to changes in health status over time among unintentionally injured adult people. A generalisability of these findings to population of injured children requires further investigation.

The SF-36 is not applicable among those younger than 16 years.¹⁷ There were too few teenagers in our sample to permit valid analyses by age. Further research is needed to evaluate possible effects of age and sex on validity and responsiveness over time of the SF-36. Although the measures are not applicable among children, who undoubtedly also experience a deterioration in health after injury, it is not known which dimensions of health are most affected or to what extent. It might be that injured children experience a different profile of decline in health than adults. Further research using similar measures suitable for children is needed.

This study was supported by a grant from the Norwegian Research Council. The author is indebted to Dr Jarl Jonland at the Central Hospital in Drammen for help with data collection.

1 Yates DW, Heath DF, Mars E, Taylor RJ. A system for measuring the severity of temporary and permanent disability after injury. *Accid Anal Prev* 1991; 23: 323-9.

- 2 Andersson AL, Dahlbaeck LO, Allebeck P. Psychosocial consequences of traffic accidents: a two year follow-up. *Scand J Soc Med* 1994; 22: 299-302.
- 3 Committee on Injury Scaling. *The abbreviated injury scale — 1984 revision*. Des Plaines, IL: American Association of Automotive Medicine, 1985.
- 4 Committee on Injury Scaling. *Injury impairment scale 1994*. Des Plaines, IL: American Association of Automotive Medicine, 1994.
- 5 McKenzie EJ, Damiano AM, Ditunno JF, Luchter S, Miller TR. *The development of the functional capacity index (FCI)*. Final report. Baltimore: The Health Services Research And Developmental Center And The Johns Hopkins Injury Prevention Center, 1994.
- 6 Nygren A, Hansson PG, Tingvall C. Acute injury scaling related to residual disability. *Acta Neurochir (Wien)* 1986; suppl 36: 25-7.
- 7 Association for the Advancement of Automotive Medicine. *38th Annual Proceedings*. Lyon, France, 1994.
- 8 Anderson RT, Aaronson NK, Wilkin D. Critical review of the international assessments of health-related quality of life. *Quality of Life Research* 1993; 2: 369-95.
- 9 McDowell I, Newell C. *Measuring health. A guide to rating scales and questionnaires*. Oxford: Oxford University Press, 1987.
- 10 Patric DL, Erickson P. *Health status and health policy. Allocating resources to health care*. Oxford: Oxford University Press, 1993.
- 11 Starfield B, Riley AW, Green BF, et al. The adolescent child health and illness profile. A population-based measure of health. *Med Care* 1995; 33: 553-66.
- 12 Stewart AL, Hays RD, Ware JE. The MOS short-form general health survey. Reliability and validity in a patient population. *Med Care* 1988; 26: 453-61.
- 13 Ware JE, Sherbourne CD. The MOS 36-Item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992; 30: 473-83.
- 14 Ware JE, Brook RH, Williams KN, Stewart AI, Davies-Avery A. *Conceptualization and measurement of health for adults in the health insurance study. Vol 1. Model of health and methodology*. Santa Monica, California: Rand Corporation, 1980. (Publication No R-1987/1-HEW.)
- 15 Braizer JE, Harper R, Jones NMB, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ* 1992; 305: 160-4.
- 16 Garratt AM, Ruta DA, Abdalla MI, Buckingham JK, Russell IT. The SF-36 health survey questionnaire: an outcome measure suitable for routine use within the NHS? *BMJ* 1993; 306: 1440-4.
- 17 Jenkinson C, Wright L, Coulter A. Criterion validity and reliability of the SF-36 in a population sample. *Quality of Life Research* 1994; 3: 7-12.
- 18 Braizer J. The SF-36 health survey questionnaire — a tool for economists. *Health Economics* 1993; 2: 213-5.
- 19 Katz JN, Larson MG, Phillips CB, Fossel AH, Liang MH. Comparative measurement sensitivity of short and longer health status instruments. *Med Care* 1992; 30: 917-25.
- 20 Garratt AM, Ruta DA, Abdalla MI, Russell IT. SF 36 health survey questionnaire: II. Responsiveness to changes in health status in four common clinical conditions. *Quality in Health Care* 1994; 3: 186-92.
- 21 Mangione CM, Phillips RS, Lawrence MG, Seddon JM, Orav EJ, Goldman L. Improved visual function and attenuation of declines in health-related quality of life after cataract extraction. *Arch Ophthalmol* 1994; 112: 1419-25.
- 22 EuroQol Group. EuroQol a new facility for the measurement of health-related quality of life. *Health Policy* 1990; 16: 199-208.
- 23 Aaronson NK, Acquardo C, Alonso J, et al. International quality of life assessment (IQOLA) project. *Quality of Life Research* 1992; 1: 349-51.
- 24 Medical Outcomes Trust. *How to score the SF-36 health survey*. Boston: Medical Outcomes Trust, January 1994.
- 25 Malt FU, Olafsen OM. Psychological appraisal and emotional response to physical injury: a clinical, phenomenological study of 109 adults. *Psychiatr Med* 1992; 10: 117-34.