

# Evidence for validity of a health status measure in assessing short term outcomes of cholecystectomy

M J Bardsley, C W Venables, J Watson, J Goodfellow, P D Wright

## Abstract

**Objective** — To assess the validity of the Nottingham health profile (NHP) as an indicator of short term outcome of cholecystectomy.

**Design** — Prospective assessment of outcome.

**Setting** — One teaching hospital.

**Patients** — 161 consecutive patients admitted for cholecystectomy between January 1989 and September 1990.

**Main measures** — Patients' reported symptoms and self assessed NHP scores before cholecystectomy and at follow up at three and 12 months (76 patients); assessment before admission (19).

**Results** — Complete data were obtained preoperatively and at three months' follow up from 154 patients; seven did not respond to the follow up questionnaire. 76/84(90%) patients in the study 12 months or more answered the 12 month follow up questionnaire; eight did not respond. Significant changes in score before and at three months after the operation were observed for four of the six dimensions: energy (35.34 v 19.53,  $p < 0.0001$ ), pain (27.38 v 9.8,  $p < 0.0001$ ), sleep (26.99 v 17.51,  $p = 0.0002$ ), and emotional reactions (16.12 v 7.56,  $p = 0.001$ ). The mean scores for 76 patients followed up at three and 12 months showed little subsequent change. Scores in readmitted patients were all significantly higher, suggesting poor health. Patients with five reported symptoms had significantly worse scores for all dimensions. Scores were similar before cholecystectomy whether the questionnaire was completed before or after admission.

**Conclusion** — The NHP is an appropriate tool for monitoring changes in health after cholecystectomy.

## Introduction

The process of evaluating the outcomes of health care in terms which are directly related to patients' experience is becoming increasingly advanced. In particular the past few years have seen several instruments applied to assess health status defined in terms of physical, psychological, and social wellbeing (see review of Spilker *et al*<sup>1</sup>). It is argued that such instruments better reflect the impacts of medical care on patients' everyday lives than traditionally based technical measures of

success.<sup>2</sup> Such health status measures are relevant in clinical trials,<sup>3,4</sup> in methods of auditing care,<sup>5</sup> and, ultimately, in questions of health service management and organisation.<sup>6</sup>

The CASPE/Freeman outcomes study has been examining the way information on patient outcome can be collected and used to guide clinical practice and health service management. We report results on one aspect of that analysis — namely, the validity and responsiveness of the Nottingham health profile as a measure of health status when used for patients having a cholecystectomy.

The Nottingham health profile (NHP) was developed in the 1970s as a general measure of perceived health<sup>7</sup> which attempted to overcome some of the problems of the instruments then available. In this study it was selected as a measure which could potentially summarise patient wellbeing and include psychological and social aspects in that assessment. The main part of the profile consists of scores ranging from 0 to 100% on six dimensions: energy, pain, emotional reactions, sleep, social isolation, and mobility. Scores in this part of the profile are calculated as the weighted sum of positive responses to a total of 38 items dealing with issues of general health. Part II of the instrument asks specifically about the effects of health on several issues of daily living and is scored separately. The profile can be administered by interview or as a patient completed questionnaire.

Several studies have shown the reliability and validity of the instrument in British subjects.<sup>7,8</sup> In cross sectional analyses the profile differentiated between disease groups,<sup>9</sup> though it has been criticised for, among other things, being insensitive to change in a relatively healthy population.<sup>10</sup> Some studies have used the profile to determine longitudinal changes after heart and lung transplantation<sup>11</sup> and, more recently, after coronary artery bypass grafting.<sup>12</sup> Despite these studies there is still considerable uncertainty about applicability of the instrument to different patient groups and its responsiveness to clinically important changes after treatment.

Though many measures of general health status are available,<sup>1</sup> the NHP was chosen because it had been validated with populations in the United Kingdom; it had been used in various types of patients; it included dimensions of health that were considered relevant for this case type; and it was available as a fairly short, self administered ques-

CASPE Research,  
Kings Edward's  
Hospital Fund,  
Bayswater, London  
W2 4HT

M J Bardsley, research  
manager

Department of  
General Surgery and  
CASPE/Freeman  
Outcomes Study,  
Freeman Hospital,  
Newcastle upon Tyne  
NE7 7DN

C W Venables,  
consultant surgeon in  
gastroenterology  
J Watson, research nurse,  
general surgery  
J Goodfellow, research  
assistant  
P D Wright, consultant  
surgeon

Correspondence to:  
Dr M J Bardsley, Outcomes  
Project Manager, East  
Anglian Regional Health  
Authority, Union Lane,  
Chesterton,  
Cambridge CB4 1RF

Accepted for publication  
10 January 1992

tionnaire. Its value in this study, however, was predicated on its ability to:

(a) be administered successfully to all patients having cholecystectomy, before and after the operation; (b) discriminate between good and bad outcomes defined by some other criteria – in this instance readmissions and the level of problems or symptoms reported by patients; and (c) be sufficiently sensitive to show a (beneficial) change in the populations of patients after cholecystectomy.

### Methods

All patients admitted to this hospital between January 1989 and September 1990 for cholecystectomy were considered eligible for the study. While they were in hospital a nurse explained the nature of the study to them and sought their consent to be included. They were given an explanatory letter and asked to complete an assessment of their symptoms and the NHP before the operation. In addition, details of the indications for cholecystectomy and comorbid conditions were noted on a proforma.

Surgeons completed a simple form detailing any perioperative complications. Postoperative complications were noted at discharge. Three months after the operation all patients were sent the same questions about their symptoms and the NHP, together with an addressed and postage paid envelope. To assess the extent of the symptoms or problems associated with their gall bladder disease patients were asked to respond either yes or no to a series of questions covering problems with their bowels, vomiting, tolerance of fatty foods, flatulence, abdominal distension, loss of appetite, or pain which had led them to take analgesics. The replies to questions were combined into a single score by simply counting the number of positive symptoms reported; the total scores could therefore range from zero to seven. The follow up forms also included questions about details of any problems patients had had since discharge and whether patients had been readmitted and, if so, why.

In order to examine changes over longer periods a subset of patients, selected as those who had been in the study long enough, were also followed up at 12 months after cholecystectomy. It represented the early patients in the main study admitted up to November 1989. All follow up data on these patients were obtained by post with the same forms as used for the three month follow up. If no response was made within a month a repeat form was sent. One potential problem with this method of collecting data was that scores on admission may be artificially high

because patients adopted a sick role in hospital or were anxious about the operation or simply because they were in an unusual and unsettling environment. Therefore a series of patients were asked to complete the questionnaire before admission as well as once in hospital. In total, 19 patients completed the forms in this way and enabled a comparison of the health status scores before and after admission to hospital.

All data were entered on to a microcomputer using general database software and analysed using the statistical package for the social sciences (SPSS-PC). Non-parametric tests of statistical significance were used as much of the data had appreciably skewed, non-normal distributions.<sup>13</sup> Comparisons between changes in measures of health status over time were based on Wilcoxon's signed ranks test for matched pairs. The Mann-Whitney U test was used to compare health status scores between two categories of re-admission whereas comparisons across the multiple categories based on the numbers of reported symptoms was by the Kruskal-Wallis analysis of variance by ranks. Finally, the degree of association between scores before and after admission to hospital was based on Kendall's  $\tau_b$ .

The responsiveness of a scale, as shown by the effect size, is a reflection of the magnitude of change observed from before to after treatment and expressed relative to the initial variability in the score.<sup>14</sup> More responsive scales show large changes after treatment and have relatively less initial variability in score. The effect size was calculated as: mean score at three months after operation – the mean score before operation/ standard deviation of the score before operation.

### Results

Complete data were obtained preoperatively and at three months' follow up from 154 patients (excluding known deaths); 108(70%) of patients were women (mean age 58.9) and 46(30%) were men (mean age 61.0). Only seven (5%) patients did not respond to the follow up postal questionnaire. There was no evidence within the hospital to suggest that these had died. Of 84 patients who had been in the study for over 12 months after operation, 76(90%) provided additional 12 month follow up data and eight (10%) did not respond.

#### SCORES BEFORE AND AFTER CHOLECYSTECTOMY

Table 1 shows the mean NHP scores before and at three months after cholecystectomy. The distributions of scores were typically

Table 1 Changes in Nottingham health profile scores before cholecystectomy and at three months' follow up

Dimension	Before operation		At 3 months		Significance of difference*	Effect size
	Mean(SE)	Median	Mean(SE)	Median		
Energy	35.34(3.08)	24.00	19.53(2.50)	0.0	p < 0.0001	0.38
Pain	27.38(2.23)	19.74	9.80(1.68)	0.0	p < 0.0001	0.60
Emotional reactions	16.12(1.57)	8.27	7.56(1.39)	0.0	p = 0.001	0.45
Sleep	26.99(2.23)	14.34	17.51(2.05)	0.0	p = 0.0002	0.39
Social isolation	7.69(1.41)	1.46	6.41(1.37)	0.0	NS	0.07
Mobility	14.49(1.76)	0.00	10.35(1.46)	0.0	NS	0.16

\*Wilcoxon's matched pairs signed ranks test.

Table 2 Changes in mean Nottingham health profile scores before and at three and 12 months after cholecystectomy (n=76)

Dimension	Before operation	At 3 months	Significance of difference* <sup>†</sup>	At 12 months	Significance of difference* <sup>‡</sup>
	Mean(SE)	Mean(SE)		Mean(SE)	
Energy	39.4(4.4)	25.0(3.9)	p = 0.001	23.7(3.9)	NS
Pain	26.4(3.2)	12.5(2.7)	p < 0.0001	10.8(2.8)	NS
Emotional reactions	15.8(2.2)	8.3(1.8)	p = 0.001	7.8(2.1)	NS
Sleep	27.0(3.2)	21.2(2.9)	NS	20.7(3.3)	NS
Social isolation	7.7(2.0)	6.0(1.7)	NS	6.0(1.9)	NS
Mobility	13.7(2.5)	11.6(2.2)	NS	13.1(2.5)	NS

\*Wilcoxon's matched pairs signed rank test.

<sup>†</sup>Before operation to three months' follow up.

<sup>‡</sup>Three months' follow up to 12 months' follow up.

Table 3 Mean (SE) Nottingham health profile scores 3 months after cholecystectomy for patients who were and were not readmitted

Dimension	Readmitted (n=8)	Not readmitted (n=146)	Significance*
Energy	59.0(12.02)	18.6(2.57)	p < 0.0008
Pain	45.0(10.50)	7.8(1.51)	p < 0.0001
Emotional reactions	26.3(19.23)	6.3(1.22)	p = 0.0001
Sleep	44.8(13.05)	15.8(1.98)	p = 0.0119
Social isolation	23.2(9.69)	4.8(1.13)	p = 0.0003
Mobility	27.4(9.79)	9.8(1.34)	p = 0.0396

\*Mann-Whitney U test.

skewed, with a high proportion of patients both before and after the operation scoring zero. The median scores for all dimensions were very much lower than the means, and in the case of scores for mobility the median was zero before the operation. Therefore comparisons between the scores were based on the non-parametric Wilcoxon's matched pairs signed ranks test.<sup>13</sup> Significant changes were observed for the dimensions covering energy, pain, sleep, and emotional reactions. There was no significant change for mobility or social isolation.

The actual change in scores for any one dimension covered a wide range. With so many patients scoring zero there was the inevitable tendency for the change in score to be related to the initial score. Patients with higher preoperative scores tended to show the greatest improvements.

When the relative responsiveness of the different dimensions was assessed with the effect size the scores for individual dimensions ranged from 0.07 for social isolation to 0.60 for pain (table 1). Using a rough benchmark of effect sizes of < 0.2 to indicate low responsiveness, around 0.5 moderate responsiveness, and 0.8 high responsiveness<sup>15</sup> disclosed that four of the dimensions (pain, emotional reactions, sleep, and energy) showed moderate responsiveness whereas two (mobility and social isolation) showed little or none.

Table 2 shows the changes in mean score for the 76 patients followed up at three and 12

months. Significant changes were seen between preoperative scores and three month and 12 month follow up scores for three dimensions (energy, pain, and emotional reactions). The scores for all dimensions remained stable after the three month follow up and no significant differences were found between three and 12 months after operation.

#### RELIABILITY: TEST - RETEST SCORES

The check on differences in scores before admission and after admission was based on a comparison of 19 patients. The scores when the form was completed in hospital were not significantly different from those obtained before admission when analysed with Wilcoxon's signed rank matched pairs test. There was high degree of consistency between the two sets of results. Using Kendall's  $\tau_b$  as a measure of association showed values ranging from 0.72 for energy to 0.83 for pain. For all dimensions the probability of observed associations appearing by chance was below 0.01.

#### HEALTH STATUS OF READMITTED PATIENTS

In total, eight patients were readmitted within three months of their operation. The reasons for readmissions were varied and in two patients were unrelated to the original operation. The follow up scores of patients who had been readmitted were significantly higher than those of the rest of the sample. Table 3 illustrates the differences by using mean values. For all dimensions the few patients who had been readmitted showed much worse average NHP scores. These differences were significant with non-parametric tests and suggest that readmission and poorer health status as measured by the NHP were significantly associated.

#### HEALTH STATUS AND REPORTED SYMPTOMS

The number of symptoms or problems reported by patients could range from zero to

Table 4 Preoperative mean (SE) Nottingham health profile scores according to number of reported symptoms or problems

Dimension	No of reported symptoms or problems			Significance*
	<3 (n=57)	3-4 (n=53)	>5 (n=44)	
Energy	22.3(4.22)	36.6(5.37)	56.0(5.80)	p = 0.0001
Pain	13.0(2.85)	29.5(4.05)	44.8(4.34)	p < 0.0001
Emotional reactions	6.6(1.88)	19.3(2.53)	17.2(2.59)	p < 0.0001
Sleep	22.9(3.66)	26.4(3.63)	41.7(4.25)	p = 0.0014
Social isolation	5.3(2.00)	5.0(1.83)	15.5(3.48)	p = 0.0035
Mobility	11.3(2.90)	10.2(1.88)	25.0(2.92)	p = 0.0006

\*Kruskal Wallis analysis of variance by ranks.

seven. Patients with higher initial symptom scores reported significantly higher NHP scores for all dimensions. This relation was also observed between NHP and symptom scores at follow up. Table 4 illustrates the scale of differences by showing the mean preoperative NHP scores according to the number of reported symptoms. The association between NHP score and number of symptoms was highly significant for all dimensions, including those of mobility and social isolation, which had proved less liable to change as a result of cholecystectomy. For these latter two dimensions there was a appreciable difference in scores between the group of patients with five or more symptoms or problems and those with fewer than five reported problems.

### Discussion

Despite growing interest in the use of measures of general health status in assessing patient outcome there is still relatively little evidence of the performance characteristics of the various scales or their responsiveness for different disease or treatment groups. Guyatt *et al*<sup>16</sup> noted the important distinction to be made between the sensitivity of these various tools when used in cross section in contrast to the responsiveness of scales – that is, their ability to show up clinically meaningful change.

This simple study has shown good results with the NHP after cholecystectomy despite the distributions of scores for all dimensions being appreciably skewed (containing a high proportion of patients with a score of zero on individual dimensions). On examination of the distributions and changes in scores it seems that the instrument is better suited to differentiating serious health problems within this patient group and does not give fine gradations of health normally distributed about a mean.

Nevertheless, the aggregate scores did show clear and significant change after cholecystectomy. Moreover, change was in those dimensions which intuitively seem most appropriate for this type of patient. The observed pattern of improvements in pain, energy, and sleep was broadly in line with that expected for patients undergoing cholecystectomy. Mobility and social isolation were thought less likely to be problems before the operation or to show improvement afterwards. Scores for emotional reactions were not expected to change much, yet significant improvements were observed. It seemed therefore that the NHP was capable of detecting not only changes in physical function but also in some aspects of improved psychological wellbeing after acute surgery. Using the effect size as an indicator of the responsiveness of the scale showed reasonable results for four of the six dimensions. Despite the problems the NHP seems to be a workable scale for this patient group. The results also show that improvements in health status obtained at three months were maintained to 12 months after cholecystectomy. Thus the

apparent benefits of the operation were manifested within three months, and subsequent change to 12 months, for better or worse, was less significant.

Evidence of convergent validity in the scales in this setting was provided by the association between the NHP score and a simple count of symptoms reported by patients. Patients reporting more symptoms before cholecystectomy showed significantly higher scores for all dimensions of the profile. Four of the dimensions showed what seemed to be an approximately linear increase in NHP scores with increasing numbers of reported symptoms or problems whereas for two dimensions (mobility and social isolation) the relation seemed stepped. Higher scores for these two dimensions might be associated with some threshold levels of numbers of symptoms rather than as a linear relation. In general, the high degrees of association for all dimensions suggest the type of relation that would be expected from these scales – that is, the greater the number of symptoms the poorer the health state as shown by the NHP.

Similarly, the observed higher follow up scores in patients who had been readmitted after the procedure provides additional support for the validity of the scale in this setting. Though few in number, readmitted patients scored significantly worse for all dimensions, and the fact of readmission, whatever the reason, seemed to be linked to worse health status according to this scale.

Studies of the behaviour of the scale around the time of admission to hospital suggested that results obtained with the NHP were sufficiently reliable and that for a limited sample of patients there was no significant difference in score before and after admission to hospital. This result is important when considering the subsequent change in scores after cholecystectomy in that it suggests that the preoperative scores were not artificially high as a result of the process of admission to hospital or the way in which the scale was administered. We conclude therefore that the subsequent change in NHP scores after operation is a reflection of the benefits of treatment rather than an artefact of the way data were collected.

Perhaps most importantly, high response rates were obtained in this study. The collection of postal follow up data proved far easier than was originally thought, and the administration of postal follow up forms proved one of the simpler ways to collect data. We suspect that much of the high response rates can be attributed to the persuasive powers of the research nurse and the fact that patients were told of the nature of the study and were expecting to receive the postal forms. The addition of prepaid envelopes, explanatory letters, and automatically sending repeat forms to non-responders may also have helped. There was some early scepticism about whether patients would respond to quite detailed postal questionnaires. The results have been unequivocal in this respect and suggest that there is greater potential for using

postal survey techniques in monitoring outcome after acute treatments. Using patients as a source of information for assessing outcome in this context seems to be satisfactory from both theoretical and practical points of view.

Deyo and Patrick identified several barriers to the wider use of health status measures and important issues concerning the scepticism that may exist about them.<sup>17</sup> One of their proposed ways of overcoming such barriers was by promoting comparisons between health status scores and other physiological and clinical measures and through greater efforts to show (and improve) the responsiveness of scales to clinically important changes. The behaviour of the NHP during this study has helped us overcome some of our initial scepticism; we have shown that even with an acute treatment such as cholecystectomy good results can be obtained. As a consequence it has become possible to evaluate the benefits of surgery in a way that more directly reflects improved patient health rather than by relying on physiological variables or proxy values. Moreover, relying on patients to provide data on their own health did not lead to any greater problems in data collection; in many cases this information was easier to obtain than that for traditional measures. There seems ample opportunity for using such measures in increasing our understanding of the effects of medical interventions, in auditing the processes of care, and, eventually, in general systems of quality assurance in the acute sector. In all these settings it is possible to envisage ways in which such measures of health status could yield information that

more directly assesses whether medical care achieves its goals and really benefits patient health.

- 1 Spilker B, Molinek FR, Johnston KA, Simpson RL, Tilson HH. Quality of life bibliography and indexes. *Med Care* 1990;**28**(suppl):DS1-77.
- 2 Fries J. Towards an understanding of patient outcome measurement. *Arthritis Rheum* 1983;**26**:697-704.
- 3 Brook RH, Kamberg CJ. General health status and outcome measurement: a commentary on measuring functional status. *Journal of Chronic Diseases* 1987;**40**(Suppl):131-6.
- 4 Bombardier C, Ware J, Russell LJ, Larson M, Chalmers A, Reid JL. Auranofin therapy and quality of life of patients with rheumatoid arthritis. *Am J Med* 1986;**81**:565-77.
- 5 Hopkins A, Costain D, eds. *Measuring the outcomes of medical care*. London: Royal College of Physicians, 1990.
- 6 Ellwood PM. Shattuck lecture – outcomes management. A technology of patient experience. *N Engl J Med* 1988;**318**:1549-56.
- 7 Hunt SM, McEwen J, McKenna SP. *Measuring health status*. Beckenham: Croom Helm, 1986.
- 8 Hunt SM, McEwen J, McKenna SP. Perceived health: age and sex comparisons in a community. *J Epidemiol Community Health* 1984;**38**:156-60.
- 9 Jenkinson C, Fitzpatrick R, Argyle M. The Nottingham Health Profile: an analysis of its sensitivity in differentiating illness groups. *Soc Sci Med* 1988;**27**:1411-4.
- 10 Kind P, Carr-Hill R. The Nottingham Health Profile: a useful tool for epidemiologist? *Soc Sci Med* 1987;**25**:905-10.
- 11 O'Brien BJ, Banner NR, Gibson S, Yacoub MH. The Nottingham Health Profile as a measure of quality of life following combined heart and lung transplantation. *J Epidemiol Community Health* 1988;**42**:232-4.
- 12 Caine N, Harrison SCW, Sharples LD, Wallwork J. Prospective study of quality of life before and after coronary artery bypass grafting. *BMJ* 1991;**302**:511-6.
- 13 Blalock HM. *Social statistics revised* 2nd ed. Tokyo: McGraw-Hill, 1979.
- 14 Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care* 1989;**27**(Suppl):S178-89.
- 15 Cohen J. *Statistical power analysis for the behavioural sciences*. Vol 8. New York: Academic Press, 1977.
- 16 Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *Journal of Chronic Diseases* 1987;**40**:171-8.
- 17 Deyo RA, Patrick DL. Barriers to the use of health status measures in clinical investigation, patient care and policy research. *Med Care* 1989;**27**(Suppl):S254-68.