The use of the SF36 as an outcome measure for renal replacement therapy

Geographical Area covered: Sheffield
Focus: Case studies focusing on subdistrict variation in health outcome

Contributors:
Authors: Jeremy Wight, Senior Registrar in Public Health Medicine
Address for correspondence: Sheffield Health, 5 Old Fulwood Road, Sheffield, S10 3TG

Editorial comments on how case study is linked to improving health outcomes: (also published in Volume 1)

Wight highlights the problems of obtaining information about the outcomes of different modalities of renal replacement therapy. Over 1% of their hospital budget is spent on these therapies for 0.04% of the population and the numbers and proportion of the budget are expected to rise. Together with other local purchasers, they felt that it was important to have an assessment of the quality of life (QoL) being enjoyed by patients on different modalities of RRT, which have varying costs, in order to inform the debate about choice of modality, and numbers to be taken onto the end stage renal failure programme. They wished to assess the potential for using a QoL outcome measure as a quality measure in contracting for RRT and whether routinely collected data could be used as a proxy measure for QoL.

Abstract (also published in Volume 1)

Objectives: To explore the use of the Short Form 36 (SF36) as an outcome measure for End Stage Renal Failure (ESRF) treatment, and to document the Quality of Life (QoL), as measured by this instrument, of patients treated at the Sheffield Kidney Institute.

Design: A cross sectional survey of 660 patients being treated at the Sheffield Kidney Institute, using the standard UK version of the SF36 supplemented by eight questions relating specifically to symptoms common in renal failure, and demographic questions, as well as clinical and laboratory data taken from patient records. Comparison with age and sex matched normal population. Multiple regression analysis of possible predictor variables.

Setting: Health Authority with a population of 530,000, hosting a Renal Unit which also provides services for surrounding Districts.

Subjects: The entire population of patients treated for ESRF by any modality (haemodialysis, peritoneal dialysis and transplantation) at the Sheffield Kidney Institute in April and May 1995.

Main outcome measures: Individual dimension scores of the SF36 for each modality. Effect sizes. Spearman Rank correlation between the responses to the additional ESRF specific questions and SF36 scores. Multiple Regression model of predictor variables and SF36 scores.

Results: All SF36 dimension scores were significantly worse than those of the age and sex matched sample from the normal population. When individual modalities were examined separately, transplanted patients did not differ from normals in Energy and Mental Health scores, and home haemodialysis from normals in Role Emotional and Mental Health. Significant floor and ceiling effects were seen in the Role Physical and Role Emotional dimensions. The additional ESRF specific questions did give additional independent information about the QoL of dialysis (but not transplanted) patients. Multiple regression analysis revealed successful transplantation, age, gender, comorbidity, and the level of social and emotional support to be significant predictors of SF36 scores. Routinely collected laboratory data is not a good proxy for QoL.

Conclusions: The SF36 is a useful outcome measure for ESRF treatment, though a fuller picture of overall QoL can be gained by adding disease specific questions. The decline in QoL with age is no greater than that seen in the normal population, and there is no cut off in age, above which it would
be inappropriate to offer treatment. Transplanted patients have a significantly better QoL, but there is otherwise little difference between modalities. Further work is needed to determine whether interventions or changing practice by clinicians can lead to improvements in SF36 scores.

**Introduction:**

**Why this clinical area was chosen:**

Sheffield Health Authority serves a population of approximately 530,000. Of these, approximately 200 have end stage renal failure (ESRF) and are receiving renal replacement therapy (RRT), at a cost to the District of £2.25 million. Thus over 1% of the hospital budget is spent on the treatment of 0.04% of the population. Further, even at the current take on rate of new patients (which is accepted to be less than the incidence of ESRF, and so, arguably, ought to be increased), the numbers on RRT, and hence the proportion of the overall budget, will continue to increase for some years (Wight et al. 1996).

There are marked differences in the costs charged by the local provider for different modalities of RRT. In particular, chronic ambulatory peritoneal dialysis (CAPD) is cheaper than haemodialysis (whether carried out at home or in hospital). Maintenance with a functioning transplant is the cheapest modality. There is also a tendency for more patients to be dialysed in satellite dialysis units, outwith the main renal unit.

Little information exists about the outcomes of RRT, other than mortality data. Together with other local purchasers, we felt that it was important to have an assessment of the quality of life (QoL) being enjoyed by patients on different modalities of RRT, in order to inform the debate about choice of modality, and numbers to be taken onto the ESRF programme. We also wished to assess the potential for using a QoL outcome measure as a quality measure in contracting for RRT. We also wished to know whether routinely collected data could be used as a proxy measure for QoL.

**Further information that was required:**

We therefore undertook a study of the QoL of patients being treated for ESRF at the Sheffield Kidney Institute at the Northern General Hospital, Sheffield. The study was a cross-sectional survey of all 660 patients being treated (sample size calculations indicated that this was desirable), including those from outlying districts being treated at that centre. (This also enabled us to examine the effect of distance traveled to treatment on QoL). Data was collected during the months of April, May and June 1995.

In order to be able to make comparisons between the QoL of ESRF patients and patients with other chronic conditions, as well as the normal population, it was important to use a generic health status measure for which norms were available for the general population, and which either had been or could be used in other patient populations. The Short Form 36 (SF36) is one such measure (Ware 1993(a); Ware 1993(b); Brazier et al. 1992). However a generic instrument may not be sensitive to symptoms that are common in ESRF, and which have an impact on QoL. We therefore decided to supplement the SF36 with questions about ESRF specific symptoms, in order to assess the impact of those symptoms on QoL and the extent to which this impact was independent of SF36 scores. This has been advocated elsewhere (Gill and Feinstein 1994).

The eight additional questions asked about symptoms which were thought by clinical staff to be common and important to ESRF patients, but which were not included in the SF36. They also concurred with symptoms found by others to be significant in determining quality of life in ESRF (Laupacis et al. 1992; Hays et al. 1994). The eight symptoms were: dry or itchy skin; cramps; difficulty sleeping; thirst; problems with access site; muscle wasting; nausea; and impairment of sex drive. In order to distinguish between the simple presence of the symptom and the extent to which it affected quality of life, the questions were phrased ‘during the past 4 weeks how much have you been bothered by ...’. Respondents were asked to indicate the extent on a scale from one (not at all) to six (extremely).

Patients on RRT clearly vary in age, extent of comorbidity, as well as other factors which may affect quality of life. We were keen, therefore, to be able to take these into account in any analysis. Factors examined were: age; co-morbidity (using a simple scoring system that has been found to predict survival in ESRF patients) (Wright 1991); duration of treatment; previous transplantation; serum albumin and haemoglobin levels; adequacy of dialysis as measured by urea kinetic modeling; district of residence; distance traveled for treatment; marital status; the number of people living in the same household; and the level of emotional and social support. (This was assessed by asking respondents...
to state the level of emotional and social support received at home by choosing from a six point scale going from 'none at all' to 'an awful lot of support'. The other information was obtained by including further questions in the survey instrument, or from the hospital records.

The correlations between these factors and SF36 scores were explored using multiple regression analysis.

Finally, we were keen to examine the extent of correlation, if any, between physiological variables that are routinely analysed as part of the usual care of patients and QoL. A high degree of correlation would indicate that those variable(s) could be used as a proxy measure of QoL, and so obviate the need for dedicated surveys.

Age and sex matched comparisons with the normal Sheffield population were made using data collected for a previous study (Brazier et al. 1992). Appropriate age (within five year age bands) and sex matched cases were chosen at random.

Comparisons were made with published results of the measurement of QoL using the SF36 in other patient groups (Garratt et al. 1993).

Data were collected onto a personal computer on an Epi-Info (Dean et al. 1990) record file. This was converted into an Excel spreadsheet for data manipulation. Statistical tests were performed using SPSS for Windows. Because of the non-normal distribution of most of the data, non parametric tests (Wilcoxon Matched-Pairs Signed Ranks test or Kruskal-Wallis test, as appropriate) were used.

Data validity studies:

The validity of the SF36 in the general population has been established elsewhere (Brazier et al. 1992; Lyons et al. 1994). One of the purposes of our study was to establish its validity in the ESRF population.

The renal specific questions were written specifically for this study, and so clearly had not been validated elsewhere. Similarly, the question asking respondents to grade the level of social and emotional support received was written for this study and had not been validated elsewhere.

There are limitations inherent in a non-randomised cross sectional study such as this. Firstly, comparison between groups is complicated by clear differences between the groups, i.e. differences in case-mix. This is addressed partly by the multiple regression analysis, but that only considers a few predictor variables, and the possibility of unknown confounders cannot be excluded. Secondly, there is a possibility of response bias (see below). Thirdly, the effect of any change in treatment cannot be assessed. Finally, only correlation, not causation, can be inferred.

Main findings

520 responses were received, giving a response rate of 78.8%. There was no difference between responders and non-responders in terms of age, sex, treatment modality, duration of treatment, district of residence, or distance traveled for treatment. Internal consistency as measured by Cronbach’s Alpha, was high. Significant ‘floor’ and ‘ceiling’ effects were only seen in the ‘Role Limitations - Physical’ and ‘Role Limitations - Emotional’ dimensions. The results had face validity, in that the scores were correlated with age, co-morbidity and modality of treatment (transplanted patients had higher scores). The SF36 appeared therefore to be providing useful measures of QoL in ESRF patients. It also appeared to be acceptable to the large majority of the population studied.

Overall, the quality of life of these patients as measured by this instrument is poor to moderate, especially for the ‘physical’ dimensions. The results for each modality, and for all patients overall, are available from the author. The difference between the dialysis patients and controls varies between modalities and dimensions, but averages over 30 points - greater than the difference between ‘normal’ 65-74 year olds and 25-34 year olds (Brazier et al. 1992). The scores for the ‘mental’ dimensions were higher, and showed less difference between modalities, as well as smaller effect sizes (the difference between the means divided by the standard deviation of the control group). This suggests that overall, the mental/emotional adjustment of ESRF patients was better than the physical rehabilitation.

Transplanted patients scored higher than patients treated by any other modality, and this difference was most marked for the physical dimensions (e.g. mean Physical Functioning score 62.48 for Transplanted patients, 28.29 for Satellite dialysis patients). There was no difference in Vitality and Mental Health scores between the transplanted patients and controls. The effect sizes (comparing transplanted patients with age and sex matched controls) for individual dimensions were less than for
other modalities.

Amongst dialysis treated patients, Home dialysis patients tended to score higher, particularly on the Mental Health, Role Emotional and Social Functioning dimensions, than did the Hospital, Satellite or Peritoneal dialysis patients (e.g. mean Social Functioning score 62.87 for Home dialysis patients, 41.89 for Hospital dialysis patients). There was no significant difference in score on the Mental Health and Role Emotional dimension between the Home dialysis patients and age and sex matched controls. Details are available from the author.

A large majority of responses to the additional questions about ESRF specific symptoms revealed that they did not have an impact on the quality of life of respondents. In each case the distribution of responses was heavily skewed towards the lower scores. As a consequence, these particular questions do not discriminate well between patients.

However, when the responses of dialysis and transplanted patients were analysed separately, it was clear that in dialysis patients the heavy skew towards lower responses is lost for the four questions relating to dry skin, cramps, difficulty sleeping and thirst. This suggests that in these patients, these particular questions do discriminate between some patients for whom these symptoms do interfere with their quality of life, and others for whom this is not the case.

As a test of whether the extra questions were giving useful extra information about the quality of life of these patients, the Spearman Rank correlation coefficients between the SF36 dimension scores and the responses on the extra questions were calculated. When all patients were included, correlations overall were moderate. For dialysis patients only (i.e. excluding transplant patients) the coefficients were lower, the greatest being -0.3876, between Bodily Pain and Difficulty Sleeping. In ten cases there was no significant correlation, using a cut-off value of > 0.01. This indicates that for dialysis patients there was substantially less correlation between the extra questions and the SF36 dimensions, implying that the questions were more useful in detecting extra information for these patients.

Clearly some of the differences in SF36 scores between modalities could be attributable to case mix. Both the Transplanted patients and the Home dialysis patients tended to be younger than the Hospital, Satellite and Peritoneal dialysis patients, as well as suffering from less co-morbidity. Multiple regression analysis established that Transplantation remained a significant predictor of higher scores even when these factors were taken into account. Hospital haemodialysis was a significant predictor of worse Mental Health score. Otherwise differences in scores between modalities were not significant. In particular, the quality of life of patients treated with Chronic Ambulatory Peritoneal Dialysis was no different to that of those treated with Haemodialysis.

The multiple regression analysis also demonstrated that correlations exist between age, co-morbidity, gender and the level of emotional support that are independent of modality of treatment. However the decline in scores with age is no greater than that seen in the normal population, and in particular there was no cutoff in age above which it would be unreasonable to offer dialysis. Summaries of the multiple regression models are available from the author.

Comparisons may also be made with other chronic disease conditions (Garratt et al. 1993). The ESRF patients studied had better scores for Vitality and Mental Health, and similar scores for Physical Functioning, than a group of patients with low back pain. They also scored higher in the Mental Health and Vitality dimensions than patients with menorrhagia and higher in Mental Health only than patients with suspected peptic ulcer. Patients with varicose veins scored higher in all these dimensions.

Correlation between routinely collected laboratory variables and QoL scores was so low that they are unlikely to be useful proxy measures for QoL.

In summary:

• The overall QoL as measured by the SF36 was poor to moderate;
• Transplanted patients had a significantly better QoL than dialysis patients;
• Questions about renal specific symptoms can add useful data about the QoL of patients being treated by dialysis;
• The effect of age on QoL in these patients is no greater than in the normal population;
• Routinely collected laboratory data cannot act as a proxy for QoL measurements;
• Further work is necessary to establish the extent to which interventions can alter the QoL of these patients.
Further detailed results are available in a report for the Trent Regional Research and Development Office, which is available from the author.

**Summary findings from initial work:**

**Changes which were made:**

Although a number of factors were identified that correlate with the SF36 scores, no causal relationship can be inferred from a cross sectional study.

Discussions are underway with the providers about looking at changes in QoL with changes in modality of treatment (in particular transplantation or transplant failure). Furthermore, we are discussing whether there are any interventions that can be undertaken by clinicians that have an effect on the QoL scores. In particular we are exploring the possibility of a randomised controlled trial to establish whether increasing the level of emotional and social support can increase the mental health scores.

The providers have yet to be convinced of the value of routine monitoring of SF36 scores, pending firm evidence that these scores are affected by medical interventions. However, the collaborative working with the provider unit that this project entailed has demonstrated a shared interest in outcomes measurement, thus contributing to a continuing improvement in relations, which enhances the potential for further work in this area.

**How changes will be monitored:**

**Resource Implication:**

The largest single resource input needed for this study was the time of a research nurse, who worked on the project (data collection and entry) for six months. Data analysis was undertaken by a Senior Registrar in Public Health Medicine, and took approximately one month whole time equivalent. Other costs were for printing of the questionnaire, postage etc.

The resource implications of the findings are potentially considerable. Although the overall scores indicated that the QoL of these patients is not good, it is comparable to that of other patients with chronic illness. Furthermore, the decline in SF36 score with age is no greater than that seen in the normal population. This data therefore lends support to the argument that more resources ought to be made available for patients with ESRF to be treated, and also to the argument that there is no clear age cut off above which it would be inappropriate to offer treatment. For Sheffield (with a population of 530,000), if the take on rate onto the programme is held constant at 60 per million, the overall annual costs will increase (because a constant take on rate leads to an increased stock of patients) from £2.36 million to £2.96 million over the next five years. An increase in the take on rate to 80 per million population would lead to a further increase in annual cost of £0.45 million to £3.41 million over the same time period.

**Practical lessons learnt:**

Collecting and analysing large amounts of data is time consuming! QoL is clearly affected by ESRF and RRT. However, of those factors that were shown to correlate with the SF36 scores in these patients, most were outwith the control of clinicians (age, gender, co-morbidity). There may therefore be little that clinicians can do to improve the SF36 scores of patients in their care. (This may be due to the fact that the majority of patients who would benefit from Erythropoietin treatment were already being treated, thus minimising the range of haemoglobin levels, and the close attention that is paid to urea kinetic modeling, ensuring that the majority of patients undergo adequate dialysis).

One exception to this is the self-assessed level of emotional support enjoyed by patients, which correlated with the Mental Health score of the SF36. Further work is planned to see whether interventions designed to increase the level of emotional support received have any effect on Mental Health scores.
**Conclusion:**

The SF36 is a useful outcome measure for ESRF treatment, though a fuller picture of overall QoL can be gained by adding disease specific questions. The decline in QoL with age is no greater than that seen in the normal population, and there is no cut off in age, above which it would be inappropriate to offer treatment. Transplanted patients have a significantly better QoL, but there is otherwise little difference between modalities. Further work is needed to determine whether interventions or changing practice by clinicians can lead to improvements in SF36 scores.

**References:**

**B) The use of the SF36 as an outcome measure for renal replacement therapy**


**Organisational Context:**

The first study described here was undertaken as a result of a special research project funded by non-recurrent funds in the Department of Public Health as part of its continued interest in the impact of deprivation on health in Sheffield. The authors had previous experience in the use of the survey methods described.

The second project was a result of the North Trent Renal Purchasers Group's interest in finding out more about the outcomes of a service which needs to consume increasing amounts of resources each year just to keep acceptance rates static, and still more to extend the treatment to more elderly groups of the population. This latter project was led by a senior registrar in Public Health Medicine and was made possible by a grant from the Trent Research and Development Programme. Again the author had special experience, having worked for some years as a clinician on the renal unit.

Neither, therefore, was part of a systematic programme using health outcome measures in other clinical areas. It is our view that the use of outcome indicators is important, but can be time and resource consuming. Sheffield Health Authority has substantial familiarity with routine data and good access to it. However, while the wider use of more simple or less time consuming indicator
methodology can be of value we believe, like others, that it is important to avoid over-interpreting the results of such work, irrespective of whether it is carried out at local or national level. Thus, for example, the CABG rate per head of population is not a good national indicator given that there is clear evidence that the prevalence of the disease the intervention is designed to treat is so different between different parts of the country. Moreover, at Health Authority level an ‘acceptable’ CABG rate could disguise substantial inequalities in utilisation between different areas within the District. This caution with respect to simplistic measures ought to be an important part of a Health Authority’s approach to the use of population health indicators. Finally, it is important to work for a collaborative approach with clinicians in respect of developing outcome measures and in attempting to change clinical behaviour. There is no magic solution to carrying this out - rather one needs time and commitment to carry out detailed work on specific topics such as, for example, the renal outcome measure described here, or on aspirin prescribing as has been demonstrated in the FACTS project (Framework for Appropriate Care Throughout Sheffield).