REVIEW ARTICLE

# Routine administration of Health Related Quality of Life (HRQoL) and needs assessment instruments to improve psychological outcome – a systematic review

S. M. GILBODY,[1] A. O. HOUSE AND T. SHELDON

*From the NHS Centre for Reviews and Dissemination and Department of Health Studies, University of York; and Academic Unit of Psychiatry and Behavioural Sciences, University of Leeds*

## ABSTRACT

**Background.** Routine administration of Health Related Quality of Life (HRQoL) and needs assessment instruments has been advocated as part of clinical care to aid the recognition of psychosocial problems, to inform clinical decision making, to monitor therapeutic response and to facilitate patient–doctor communication. However, their adoption is not without cost and the benefit of their use is unclear.

**Method.** A systematic review was conducted. We sought experimental studies that examined the addition of routinely administered measures of HRQoL to care in both psychiatric and non-psychiatric settings. We searched the following databases: MEDLINE, EMBASE, CINAHL, PsycLIT and Cochrane Controlled Trials Register (to 2000). Data were extracted independently and a narrative synthesis of results was presented.

**Results.** Nine randomized and quasi-randomized studies conducted in non-psychiatric settings were found. All the instruments used included an assessment of mental well-being, with specific questions relating to depression and anxiety. The routine feedback of these instruments had little impact on the recognition of mental disorders or on longer term psychosocial functioning. While clinicians welcomed the information these instruments imparted, their results were rarely incorporated into routine clinical decision making. No studies were found that examined the value of routine assessment and feedback of HRQoL or patient needs in specialist psychiatric care settings.

**Conclusions.** Routine HRQoL measurement is a costly exercise and there is no robust evidence to suggest that it is of benefit in improving psychosocial outcomes of patients managed in non-psychiatric settings. Major policy initiatives to increase the routine collection and use of outcome measures in psychiatric settings are unevaluated.

## INTRODUCTION

### Routine outcome measurement

The measurement of patient outcome has risen in prominence over the past 30 years (Donabedian, 1966; Ellwood, 1988; Lohr, 1988). Instruments have been developed that measure symptoms of illness, and more recently 'patient based measures' have been developed. The latter assess the impact of illness on the individual (Jenkinson, 1994), and are often referred to as health status, health related quality of life (HRQoL) or functional status measures (Bowling, 1997). They measure more than just symptoms, since they incorporate some combination of the following domains: physical health; mental health; social functioning; role functioning; general perceptions of health and well-being;

[1] Address for correspondence: Dr Simon M. Gilbody, Academic Unit of Psychiatry and Behavioural Sciences, University of Leeds, Leeds LS2 9LT.

cognitive capacity; and, patient satisfaction (Ware, 1995). A related development has been the emergence of standardized needs assessment tools, which measure unmet emotional, physical, social and financial needs of the individual patient – including those with mental illness (National Institute of Mental Health, 1987; Thornicroft *et al*. 1992; Brewin & Wing, 1993; Johnson *et al*. 1996).

Outcomes measures and needs assessment tools are now used for a number of purposes, including: the evaluation of the clinical and cost effectiveness of interventions; the monitoring of population health; clinical audit; service planning; quality improvement; and, as an aid to clinical decision making in routine clinical practice (Faden & Leplege, 1992; Fitzpatrick *et al*. 1992; Fitzpatrick, 1994; Ware, 1995; Johnson *et al*. 1996). It is the last of these uses that will be considered in this review.

## The potential benefits of routine HRQoL and needs assessment

When used as aids to individual patient decision-making in routine care, HRQoL measures and needs assessment tools may potentially improve individual patient care in a number of ways. First, they help in identifying problems that might not otherwise be recognized. For example, clinicians are often unaware of a patient's social and psychological problems (Sprangers & Aaranson, 1992), the identification of which might lead to improvement of the overall quality and outcome of patient care (Kazis *et al*. 1990; Young & Chamberlain, 1987). Secondly, standardized measurement of the patients' progress over time may help the clinician to make informed decisions about treatment and to assess the impact of treatment changes. Thirdly, patients often welcome the opportunity of giving clinicians information about their health status, particularly when they perceive this information is not otherwise comprehensively assessed, thus improving satisfaction with patient–doctor communication (Nelson *et al*. 1990).

In this study we are interested in two areas in which routine HRQoL and needs assessment might be particularly useful. The first is in the recognition and management of psychiatric disorders (such as anxiety and depression) in non-psychiatric settings (such as primary care and the general hospital). The second is in the management of already recognized psychiatric disorders in specialist psychiatric care settings.

Disorders such as anxiety and depression are common in primary care and general hospital settings, and yet often go unrecognized (Goldberg & Huxley, 1980; Feldman *et al*. 1987; van Hemert *et al*. 1993). Psychiatric screening (or 'case finding') questionnaires, such as the General Health Questionnaire, have been advocated as an aid to case detection and clinical decision making (Goldberg, 1986), and we recently reviewed their use (Gilbody *et al*. 2001 *a*). Overall, we found no evidence for their effectiveness in improving quality of care. Measures of health related quality of life, such as the Short Form 36 (SF36) also identify psychosocial problems (Greenfield & Nelson, 1992), since many contain items which measure 'psychological well-being' (Ware & Sherbourne, 1992) and discriminate between groups of patients with and without clinically diagnosed depression (e.g. McHorney *et al*. 1993; McHorney & Ware, 1995).

Health-related quality of life for those with disorders such as depression and schizophrenia is especially poor, and is on a par with chronic medical conditions such as rheumatoid arthritis and ischaemic heart disease (Wells *et al*. 1989; Orley *et al*. 1998). In the case of schizophrenia, impairments in health related quality of life are often unrelated to the number or severity of symptoms such as delusions and hallucinatory experiences (Becker *et al*. 1993; Anthony & Rogers, 1995). Clinicians often underestimate the health related quality of life of patients, when compared to ratings made by patients themselves (Becker *et al*. 1993; Lehman, 1983 *a*, *b*; Sainfort *et al*. 1996). Empirical evidence shows that clinicians do not routinely measure HRQoL and needs using standardized measures (Gilbody *et al*. 2002).

For these reasons, recent mental health policy documents in England place great emphasis on the routine measurement of outcome and of patient needs (Secretary of State for Health, 1999). However, the routine measurement of outcome has not been without its critics (Crombie & Davies, 1997), and concerns have been raised that 'outcomes measures' (including HRQoL and needs assessments) are un-interpretable, unwieldy and a bureaucratic hindrance to successful patient care. The measurement of

outcome in the context of individual patient care is also not without cost, since instruments must be developed, administered (often by clinicians), coded, stored and retrieved. Similarly, there is a danger that outcome measurement triggers investigation or additional treatments, which are either of no benefit or harm to patients. There is a danger that the use of outcome measurement becomes a marketing ploy, in which measurement is used to demonstrate an institution's 'customer orientation', while the results do not in fact inform the provision of care (Fitzpatrick, 1994).

One way in which the usefulness of outcome measures in everyday care might be judged is by evaluating the degree to which their administration and feedback improves the outcome and quality of care. Previous (non-systematic) reviews of this question have been equivocal (Fitzpatrick, 1994).

## Aims and objectives

This review assesses systematically the best available evidence on the value of routine HRQoL and needs assessment in: (1) improving the psychological care and outcome of people being managed in non-psychiatric settings (such as primary care and the general hospital); and, (2) improving the quality of care and outcome of those with common mental disorders such as anxiety, depression and schizophrenia, being managed in specialist psychiatric settings.

## METHOD

### Search strategy

The following bibliographic databases were searched: MEDLINE (1966–2000); EMBASE (1981–2000); CINAHL (1982–2000); PsycLIT (to 2000); Cochrane Controlled Trials Register (to 2000); Cochrane schizophrenia and depression, anxiety and neurosis group specialist registers (to 2001). The search strategy combined Medical Subject Heading (MeSH) terms relating to all forms of mental illness and utilized a comprehensive strategy for identifying research that relates to outcome and needs assessment (Brettle *et al*. 1998) (full search terms available from the authors).

Titles and abstracts from electronic searches were scrutinized and all potentially relevant articles were obtained. Reference lists were scrutinized for additional studies. In addition the following journals were hand searched: *British Journal of Psychiatry* (1976–2000); *American Journal of Psychiatry* (1976–2000); *Archives of General Psychiatry* (1976–2000); *Psychological Medicine* (1976–2000); *Quality of Life Research* (1993–2000); *Journal of Psychosomatic Research* (1980–2000); and, *Medical Care* (1976–2000).

### Study inclusion criteria

#### Patients

Studies were included if the subjects were: (1) patients in non-psychiatric settings, such as general hospitals or general practice; or, (2) patients with psychiatric disorders being managed by specialist psychiatric services.

Studies relating to the following patient groups were excluded: (1) patients whose primary problem was one of substance abuse or who are managed in specialist substance abuse services; (2) children and adolescents; and, (3) those with learning disabilities or dementia.

#### Interventions

We included studies comparing the introduction and feedback of a routine form of HRQoL or needs assessment, with routine care. Routine care (the control/comparator condition) involved usual patient–doctor interaction, with non-standardized history taking, investigation, referral, intervention and follow-up.

The active intervention involved the addition of a standardized HRQoL or needs assessment instrument to routine care; with the information from the outcome assessment being fed back to the clinician or being incorporated into routine care procedures (such as out-patient assessment, hospital admission or routine discharge planning). Outcome could be assessed in both intervention and control conditions, but the active component in an intervention involved the feeding back of this information to the clinician.

Potentially relevant outcome assessment instruments included all measures of health related quality of life (HRQoL) or patient need, when either used in a psychiatric population or applied in a non-psychiatric setting with the aim of detecting and monitoring emotional problems. To be included as a measure of HRQoL or need, the instrument must address mental well-being,

and at least two of the following domains outlined in an operational definition offered by Ware (1995): physical health; social functioning; role functioning; general perceptions of health and well-being; cognitive capacity; patient satisfaction.

### Design

Randomized controlled clinical trials and quasi-randomized trials were included.

### Quality assessment

First, studies were judged according to accepted quality assessment criteria, using the Jadad scale (Jadad *et al.* 1996), the criteria of Schulz *et al.* (1995) and Cochrane criteria (Mulrow & Oxman, 1999). Particular attention was paid to the method of randomization, such that those studies that described themselves as randomized but did not describe an adequate method of randomization and concealment of allocation were distinguished from those that did.

Secondly, the unit of randomization was established. Cluster randomized studies were considered to be superior to non-cluster based studies, since the former avoid the inherent problem of cross contamination between patients seen by individual clinicians (Gilbody & Whitty, 2002). For those studies in which the unit of randomization was by clinician or clinical population, rather than individual patients, evidence was sought that clustering had been incorporated into the analysis of the study by the authors (Ukoumunne *et al.* 1999). Cluster-based studies that fail to incorporate the effect of intra-cluster correlation in their analysis are prone to a unit of analysis error, with a high probability of spurious positive results (type 1 errors) (Divine *et al.* 1992; Gilbody & Whitty, 2002).

### Outcomes

We sought data on primary psychological outcomes, including: the detection of psychiatric disorders, such as depression or anxiety; initiation of treatment or referral for psychiatric disorders; and, the outcome of psychiatric disorders and changes in aspects of health related quality of life. In addition, we sought data on: consulting behaviour and service use (both psychiatric and non-psychiatric); hospital status (e.g. discharge, readmission or length of stay);

patient satisfaction with care and patient–doctor communication; and, cost (direct and indirect).

We also identified process data on: clinician and patient perceptions of the usefulness or acceptability of measurement instruments; and, self-reports of the use of outcome information in changing patient management.

**Data synthesis**

Interventions, settings and outcomes were too heterogeneous to apply formal meta-analytical pooling. Individual studies were reported separately, with their specific design features and results, in accordance with accepted guidelines (NHS Centre for Reviews and Dissemination, 2001).

## RESULTS

Nine studies using HRQoL instruments in non-psychiatric settings were identified. Details of these studies are provided in Table 1. We did not find any trials of the routine use of HRQoL and needs assessment measures in psychiatric settings.

**Study design and quality**

Seven studies (Goldsmith & Brodwick, 1989; Rubenstein *et al.* 1989, 1995; Wasson *et al.* 1992*a*; Calkins *et al.* 1994; Street *et al.* 1994; Wagner *et al.* 1997) investigated the use of generic health status measures: the SF36 (Ware *et al.* 1993), two studies; the functional status questionnaire (FSQ) (Jette *et al.* 1986), three studies; the Dartmouth COOP (Wasson *et al.* 1992*b*), one study; and, the Sickness Impact Profile (SIP) (Bergner *et al.* 1981), one study. One further study (Mazonson *et al.* 1994) combined an anxiety questionnaire – the anxiety components of the Symptom Check List-90 (SCL-90) (Derogatis, 1994; Fifer *et al.* 1994), with a generic health status questionnaire (the SF36) (Ware *et al.* 1993). One study (Kazis *et al.* 1990) used the disease specific Arthritis Impact Measurement Scale (AIMS) (Meenan, 1982), which includes a series of depression and anxiety items. Instruments were generally administered in the waiting room by research assistants prior to consultation.

All studies were described as randomized, although method of randomization was rarely described in adequate detail. One study (Street

Table 1.  *Studies that evaluate the use of routine administration and feedback of HRQoL instruments*

| Authors | Design | Population, setting (N) | Routine outcome measure used | Intervention and control conditions | Length of follow-up and outcomes studied | Results |
|---|---|---|---|---|---|---|
| Calkins *et al.* (1994) | RCT, physicians randomized | 60 US general hospital physicians, (497) | FSQ | Int.: physicians given a seminar on the importance of FSQ test results. FSQ administered and results included in the patients record<br>Cont.: FSQ administered as above, with no physician training and no report feedback | Six summary scales of the FSQ (activities of daily living; mental health; work performance; social activity; quality of interaction), measured at 4, 8 and 12 months | No significant difference on any subscale, including mental health |
| Goldsmith & Brodwick (1989) | RCT, clinicians randomized, stratified by clinical experience | Sequential US family practice attenders, paid $5 to participate (62) | SIP | Int.: physicians given instruction in the SIP. SIP administered by research assistant and fed back prior to consultation<br>Cont.: SIP administered, but results not fed back | Use of rehabilitative services, and follow-up by the physician for rehabilitative problems<br>Physicians and patients' perceptions of the value of the SIP | No effect on patient care for the following: return visits to the family physician; referrals to other physicians; use of rehabilitative services<br>All physicians and patients gave some indication that the SIP was potentially of use, but that the SIP was too long to assimilate into the clinical encounter |
| Kazis *et al.* (1990) | RCT, individual patients randomized | US, out-patients with rheumatoid arthritis (1920) | AIMS, which includes a battery of questions relating to anxiety and depression | Int.: AIMS administered and fed back to the clinician, at least four times over a 12-month period. Substantial change scores and scores outside of population norms were highlighted<br>Cont.: AIMS administered, but not fed back | Patient satisfaction with care and health status scores at 12 months<br>Process measures of physician impressions of the usefulness of the questionnaires also reported | No significant difference in patient satisfaction<br>No significant difference in endpoint depression or anxiety scores on the AIMS |
| Mathias *et al.* (1994); Mazonson *et al.* (1994) | RCT, primary care group practices randomized | US, primary care patients with hitherto unrecognized anxiety | SCL-90 (anxiety sub-scales only SF36) | Int.: physicians ($N = 40$) educated on the importance of anxiety problems. Received structured feedback of anxiety scores (SCL-90) and functional status (SF36) scores from patients ($N = 357$) | Recognition and treatment for anxiety problems<br>Change in anxiety scores at 3 and 5 months<br>Changes in SF36 scores at 3 and 5 months<br>Self-reported global improvement in anxiety and functional status | Increased recognition and treatment for anxiety symptoms (35·6% *v.* 20·8%, $P < 0.001$). Increased referral to mental health sector (9·5% *v.* 3·2%, $P < 0.001$) |

## Table 1 (*cont.*)

| Authors | Design | Population, setting (*N*) | Routine outcome measure used | Intervention and control conditions | Length of follow-up and outcomes studied | Results |
|---|---|---|---|---|---|---|
| | | | | Cont.: physicians (*N* = 35) received no feedback from patients (*N* = 216) who had completed the SCL and SF36 questionnaires | | No differences in change for anxiety scores (*P* = 0·89) No differences in change for SF36 (total and mental health scores) Self-reported global anxiety and functional status both improved in intervention group (46·3% *v.* 37·0% report improvement for anxiety) |
| Rubenstein *et al.* (1989) | RCT, physicians randomized | US internists in community internal medicine practices (76), and their patients (510) | FSQ, includes a five item mental health scale | Int.: FSQ administered to patients (*N* = 253) and fed back to clinicians (*N* = 39) with major deficits on domains highlighted. Clinicians encouraged to integrate FSQ results through problem identification Cont.: FSQ administered to patients (*N* = 257), with no feedback to clinicians (*N* = 37) | Clinicians perception of usefulness of FSQ results Scores on FSQ items at 4, 8 and 12 months | 48% of clinicians in the experimental group reported using the questionnaire to change therapy No differences for any subscale of the FSQ (including mental health) at 12 months |
| Rubenstein *et al.* (1995) | RCT, individual clinicians randomized | US adult internal medicine out-patient attenders | FSQ, includes a five item mental health scale | Int.: physicians (*N* = 40) given an educational package on functional status deficits (including depression). Received structured feedback of FSQ scores from 309 patients Cont.: physicians (*N* = 33) received no feedback from their 248 patients who had completed the FSQ | Patient willingness to complete FSQ instruments Case note review of recognition of and interventions for identified functional status deficits (including depression or anxiety) FSQ scores at 6 months | Non-significant increase in recognition of depression (int. *v.* cont.: 23% *v.* 20%) Significant increase in the recognition of anxiety (13% *v.* 4% *P* < 0·001) Mental health scores improved in the feedback group (endpoint mean change difference = 4·5 points (95% CI, 0·5–8·3) on a a 100-point scale) |
| Street *et al.* (1994) | Quasi-RCT individual clinicians allocated to intervention or control | Pregnant women attending obstetric out-patients in USA (53) | SF36, including subscales on role limitations due to emotional problems, and mental health | Int.: SF36 administered over the phone by researcher & summary scores included in medical charts at next attendance Cont.: SF36 administered as above, but not fed back to clinicians | Patient expectation of the clinical encounter Patient satisfaction with care | Patients were keen to be asked about the dimensions of care included on the SF36 (incl mental health). The provision of summary scores did not influence the pattern of consultation or coverage of these items |

Table 1 (*cont.*)

| Authors | Design | Population, setting (*N*) | Routine outcome measure used | Intervention and control conditions | Length of follow-up and outcomes studied | Results |
|---|---|---|---|---|---|---|
| Wagner *et al.* (1997) | RCT, individual patients randomized | Routine patients with epilepsy, being treated by two US neurologists (163) | SF36, including subscales on role limitations due to emotional problems, and mental health | Int.: clinicians received a training on SF36 scores. SF36 administered and summary scores, presented as individualized profile. Cont.: Clinician training programme given and SF36 administered, but SF36 scores not fed back | Physician perceptions of usefulness of scores. Patient satisfaction with care. No follow-up beyond the study index encounter | Physicians generally felt the data to be useful. No change in patient satisfaction between intervention and control groups (46% *v.* 50% NS) |
| Wasson *et al.* (1992*a*) | RCT, individual clinicians randomized in blocks according to patient demographics | US HMO in internal medicine specialists (56) and their patients (1522) | Dartmouth COOP, includes items on physical and emotional condition; daily work; social activities; health change; overall condition | Int.: clinicians educated about COOP charts. COOP administered prior to consultation, and taken into the consultation by the patients. Cont.: COOP administered, but not given to the clinician | Clinician self-reported use of the charts. Process of care, including test ordering; new medications; patient advice and referral. Patient satisfaction with care | Clinicians reported that charts provided new information on 15–30% of patients. No difference in process of care and patient satisfaction measures. Nothing could be established about the specific role of the COOP in affecting the management of mental health problems |

RCT, Randomized controlled trials; Int., intervention; Cont., control conditions; FSQ, Functional Status Questionnaire; SIP, Sickness Impact Profile; AIMS, Arthritis Impact Measurement Scales; SCL-90, Symptom Checklist-90; SF36, Short Form 36.

*et al.* 1994) used a quasi-randomized (alternate odd/even) method of allocation. Seven of the nine studies used individual clinicians or practices as the unit of randomization (Goldsmith & Brodwick, 1989; Rubenstein *et al.* 1989, 1995; Wasson *et al.* 1992*a*; Calkins *et al.* 1994; Mazonson *et al.* 1994; Street *et al.* 1994). None of these studies accounted for their clustering in their analysis of results.

The settings of the studies were: general medical/internal medicine out-patients (Rubenstein *et al.* 1989, 1995; Wasson *et al.* 1992*a*; Calkins *et al.* 1994); general practice/family medicine (Goldsmith & Brodwick, 1989; Mazonson *et al.* 1994); rheumatology out-patients (Kazis *et al.* 1990); antenatal clinic (Street *et al.* 1994); neurology out-patients (Wagner *et al.* 1997).

The active interventions broadly involved the feedback of instrument test results to the clinician – generally in the form of a sheet containing summary scores. Instruments were generally administered only once in each of the studies. In only four studies (Rubenstein *et al.* 1989; Kazis *et al.* 1990; Calkins *et al.* 1994; Mazonson *et al.* 1994) were assessments administered sequentially during the course of care or follow-up. In each case this was done at fixed points by research assistants, rather than at each clinical encounter.

In some studies (e.g. Mazonson *et al.* 1994; Rubenstein *et al.* 1995), feedback of outcome results was combined with an active educational programme and the provision of standardized best practice guidelines on the management. For example, in the study by Mazonson *et al.* (1994), the active educational programme involved sessions on the importance of deficits in health related quality of life and untreated anxiety, together with a description of the psychometric instruments and their interpretations. Results of profiles from three of their own patients were then discussed in detail and educational materials on the management of anxiety were provided in the form of audiotapes and articles. Additionally, a toll free telephone number was provided so that further questions could be answered by a study team physician.

## Recognition of emotional problems and minor psychiatric disorders

Of the eight studies that employ broader measures of health related quality of life as their principle outcome measure (Goldsmith & Brodwick, 1989; Rubenstein *et al.* 1989, 1995; Kazis *et al.* 1990; Wasson *et al.* 1992*a*; Calkins *et al.* 1994; Street *et al.* 1994; Wagner *et al.* 1997), four reported the effect of these measures in improving the overall rate of recognition of emotional problems (Rubenstein *et al.* 1989; Kazis *et al.* 1990; Calkins *et al.* 1994). Three of these four studies (Rubenstein *et al.* 1989; Kazis *et al.* 1990; Calkins *et al.* 1994) showed no differences in the rate of recognition of mental disorders at assessment or for any subscale of the FSQ or AIMS (including mental health) at 12 months.

In contrast, a later study by Rubenstein *et al.* (1995) reports that feedback of the FSQ increases both the rate of recognition of depression and anxiety. Symptoms of anxiety or depression were recorded by physicians in 30% of case notes by clinicians receiving feedback, compared to 21% among those not receiving feedback of results. This difference was of borderline significance (relative risk of detecting anxiety or depression following feedback = 1·42; 95% CI, 0·98–2·08). The rate of recognition of anxiety problems was increased by the largest magnitude (13% *v.* 4%, relative risk of recognition of anxiety following feedback = 3·33; 95% CI, 1·40–7·92), while the rate of recognition of depression was subject to a non-significant increase in recognition (23% *v.* 20%, relative risk of recognition of depression following feedback = 1·17; 95% CI, 0·78–1·77). The major limitation of this study is, however, the fact that while it is a cluster randomized trial (clinicians are the unit of randomization), it is analysed according to individual patients without reference to intra-class correlation coefficients.

The study by Mazonson *et al.* (1994) specifically employed routine administration of the SCL-90 and the SF36, in combination with a physician education programme in order to increase the rate of recognition and improve the outcome of anxiety in primary care. This combined intervention increased the rate of recognition of anxiety disorders (defined as 'chart notations') from 19% in the control arm to 32% in the intervention arm (relative risk of recognition of an anxiety disorder = 1·72; 95% CI, 1·25–2·37). There was a marked increase in the rate of mental health referrals (10% *v.* 3%, relative risk of outside referral for an anxiety

problem = 2·94; 95% CI, 1·33–6·51). This increased rate of intervention was not accompanied by an increased rate of initiation of psychotropic prescriptions (13% *v.* 13%).

### Subsequent outcome of emotional disorders

Of the two studies that showed a positive effect of routine outcomes measurement on the detection of mental disorders, the study by Mazonson *et al.* (1994) found no overall improvement in either total scores on the anxiety components of the SCL-90, or the mental health component of the SF36. The only positive effect reported was using an unpublished self-report scale of anxiety, used in conjunction with the SF36 and the SCL-90. The other positive study (Rubenstein *et al.* 1995) found a small, but statistically significant change in the mental health component of the FSQ (endpoint mean change difference = 4·5 points; 95% CI, 0·5–8·3, on a 100-point scale). Of the four component scales of the FSQ (activities of daily living; mental health; social activities; work performance), mental health was the only scale to show a between group difference at the end of a 6-month study period.

Three further studies (Rubenstein *et al.* 1989; Kazis *et al.* 1990; Calkins *et al.* 1994) showed no overall between group difference in any score on the disease specific AIMS or generic FSQ, including anxiety and depression scores at 12-month follow-up.

### Consulting behaviour

Only two studies (Goldsmith *et al.* 1989; Wasson *et al.* 1992) examined the effect of feedback on consulting behaviour, and found no effect on number of physician visits or resource use.

### Patient satisfaction with care and patient–doctor communication

The study by Street *et al.* (1994) examined the effect of the administration and feedback of the SF36 on patient satisfaction and communication in the antenatal clinic. Their survey showed that patients generally wanted to be asked about 'health status overall' and listed the components of the health status about which they wanted to be asked. All patients wanted to be asked about 'pain' and 'perceptions of health', fewer expressed a preference to be asked about 'social functioning' and 'mental health'. The administration of the SF36 increased the patients' satisfaction with care, but feedback of these instruments did not affect the degree to which physicians were perceived as having asked about 'health status overall'. No data were presented on the degree to which feedback of SF36 results increased the detection or discussion of mental health problems.

The study by Mazonson *et al.* (1994) included a patient interview of those who received treatment for anxiety. Feedback seemed to encourage clinicians to be more proactive in raising the problem of anxiety and need for treatment. Among those who had their scores fed back and received treatment, 67% reported that their physicians had been proactive in initiating treatment, whereas among those whose scores were not fed back, only 33% reported that the physicians had taken the first step in suggesting treatment.

The study by Wagner *et al.* (1997) showed that physicians generally felt that data from the SF36 were useful in guiding clinical practice, but this was not reflected in any between group differences in patient satisfaction with their care. Interestingly, the study by Goldsmith *et al.* (1989) showed that the administration of the SIP was felt to be unhelpful in the context of routine consultations, since the instrument was perceived as being over long and difficult to assimilate into routine decision making. The results of the SIP were discussed in less than one-third of consultations.

### Costs

No study examined the costs and resource use associated with routine outcome measurement.

## DISCUSSION
### Main findings of the review

The main findings of this systematic review are two-fold. First, the evidence to support the routine use of HRQoL instruments in improving the quality of care and mental well-being of non-psychiatric populations is largely negative. Secondly, there is no randomized evidence to support the routine use of HRQoL or needs assessment instruments in improving the quality of care or outcome of patients with recognized psychiatric disorders being managed in specialist

mental healthcare settings. These two findings will now be discussed in turn.

### HRQoL instruments to improve the psychological care and outcome in non-psychiatric settings

Mental well-being forms a core component of health related quality of life, as defined by many authors (Bergner & Rothman, 1987; Walker & Rosser, 1993; Ware, 1995; Bowling, 1997). The application of instruments designed to measure HRQoL, which include component scales of depression and anxiety, has not generally been shown to improve mental well-being or to increase the rate at which clinical disorders are recognized. The only two positive studies in the present review involved an intensive clinician educational component, targeting the importance, recognition and treatment of anxiety (Mazonson *et al.* 1994), or were of borderline statistical significance in an inappropriately analysed study (Rubenstein *et al.* 1995).

The results of the present review should be considered alongside the results of our previous review (Gilbody *et al.* 2001*a*), which found that specific mood questionnaires (such as the General Health Questionnaire and Zung Depression Inventory) have little impact on the rate of detection of mood disorders when administered routinely to all patients. The major reasons for the ineffectiveness of these instruments on routine clinical care are likely to relate to: the impracticality of the instruments; clinicians lack of familiarity with their results; and the difficulty of integrating the results of such instruments into routine clinical decision making (Deyo & Patrick, 1989).

It is possible that benefit cannot and will never be demonstrated for the routine use of HRQoL measures in individual patient decision making, since this is a purpose for which the instruments were not designed. In particular, some critics have suggested that the psychometric properties of generic health status measures are such that their scores are un-interpretable at an individual patient level (McHorney & Tarlov, 1994). Generic outcomes measures are essentially designed to evaluate healthcare and to identify need at a population level (Ware, 1995), and extrapolation of use beyond this may be unwise (Dunn, 1996).

### Routine measurement of outcome in psychiatric settings

Given the centrality of routine outcome measurement, including the measurement of HRQoL and patient needs, in recent mental health policy, it is perhaps surprising that there is no randomized evidence to support its adoption.

There are a number of candidate measures and instruments such as the Health of the Nation Scale (HoNOS) (Wing, 1994) and the Camberwell Assessment of Need (CAN) (Phelan *et al.* 1995), which were developed for use in routine care settings, and HRQoL measures such as the SF36, which have been shown to be both reliable and valid in depression (Coulehan *et al.* 1997) and in schizophrenia (Russo *et al.* 1998).

Outcome measurement and needs assessment among those cared for by specialist psychiatric services represent a health technology, which consumes resources – including clinician time and administrative support. In order to justify the collection of these data on a routine basis some benefit in terms of patient care and outcome must be demonstrated. National mental health research and policy initiatives, such as the development and adoption of the HoNOS (Wing, 1994) are dependent upon individual clinicians collecting data in routine practice (Stein, 1999). For clinicians to be willing to collect such data for each and every patient, there must be some value in terms of improving the management of the individual patient. A case study of the implementation of the HoNOS shows that clinicians find it to be of limited value in care planning and day-to-day clinical practice (Sharma *et al.* 1999); the authors noted that the HoNOS fell from use after research funding had ceased.

This observation would have come as no surprise to Alvan Feinstein (1967) who more than 30 years ago wrote:

The care of the patient is the ultimate specific act that characterises the clinician, and any classificatory system that cannot help in that will fail to gain acceptance.

Given these findings, the effectiveness of the routine use of outcomes measures in specialist psychiatric settings cannot be assumed. There remains an important gap in the research

literature, and randomized evaluations of the effectiveness of routinely administered HRQoL and needs assessment tools should precede their widespread introduction.

This review will also be published and updated in line with emerging evidence on the Cochrane Electronic Library. S.G. was supported by the UK Medical Research Council Health Services Research Training Fellowship programme.

## REFERENCES

Anthony, W. & Rogers, S. (1995). Relationship between psychiatric symptomotology, work skills, and future vocational performance. *Psychiatric Services* **46**, 353–358.

Becker, M., Diamond, R. & Sainfort, F. (1993). A new patient focussed index for measuring quality of life in persons with severe and persistent mental illness. *Quality of Life Research* **2**, 239–251.

Bergner, M. & Rothman, M. L. (1987). Health status measures: an overview and guide for selection. *Annual Review of Public Health* **8**, 191–210.

Bergner, M., Bobbitt, R. A., Carter, W. B. & Gilson, B. S. (1981). The Sickness Impact Profile: development and final revision of a health status measure. *Medical Care* **19**, 787–805.

Bowling, A. (1997). *Measuring Health: A Review of Quality of Life Measurement Scales, Vol. 2.* Open University Press: Milton Keynes.

Brettle, A. J., Long, A. F., Grant, M. J. & Greenhalgh, J. (1998). Searching for information on outcomes: do you need to be comprehensive? *Quality in Health Care* **7**, 163–167.

Brewin, C. R. & Wing, J. K. (1993). The MRC Needs for Care Assessment: progress and controversies. *Psychological Medicine* **23**, 837–841.

Calkins, D. R., Rubenstein, L. V. & Cleary, P. D. (1994). Functional disability screening of ambulatory patients: a randomised controlled trial in a hospital based group practice. *Journal of General Internal Medicine* **9**, 590–592.

Coulehan, J. L., Schulberg, H. C., Block, M. R., Madonia, M. J. & Rodriguez, E. (1997). Treating depressed primary care patients improves their physical, mental, and social functioning. *Archives of Internal Medicine* **157**, 1113–1120.

Crombie, I. K. & Davies, H. T. O. (1997). Beyond health outcomes: the advantages of measuring process. *Journal of Evaluation in Clinical Practice* **4**, 31–38.

Derogatis, L. R. (1994). *Symptom Checklist-90-R (SCL-90-R) Administration, Scoring and Procedures Manual, 3rd edn.* National Computer Systems: Minneapolis.

Deyo, R. A. & Patrick, D. L. (1989). Barriers to the use of health status measures in clinical investigation, patient care, and policy research. *Medical Care* **27** (suppl. 3), S254–268.

Divine, G. W., Brown, J. T. & Frazer, L. M. (1992). The unit of analysis error in studies about physicians' patient care behavior. *Journal of General Internal Medicine* **7**, 623–629.

Donabedian, A. (1966). Evaluating the quality of medical care. *Milbank Memorial Fund Quarterly* **44** (suppl.), 166–206.

Dunn, G. (1996). Statistical methods for measuring outcomes. In *Mental Health Outcome Measures* (ed. G. Thornicroft and M. Tansella), pp. 3–15. Springer Verlag: Berlin.

Ellwood, P. M. (1988). Shattuck lecture – outcomes management. A technology of patient experience. *New England Journal of Medicine* **318**, 1549–1556.

Faden, R. & Leplege, A. (1992). Assessing quality of life. Moral implications for clinical practice. *Medical Care* **30** (suppl. 5), Ms166–175.

Feinstein, A. (1967). *Clinical Judgement.* Williams and Wilkins: Baltimore, MD.

Feldman, E., Mayou, R., Hawton, K., Ardern, M. & Smith, E. B. (1987). Psychiatric disorders in medical in-patients. *Quarterly Journal of Medicine* **63**, 405–412.

Fifer, S. K., Mathias, S. D., Patrick, D. L., Mazonson, P. D., Lubeck, D. P. & Bensching, D. P. (1994). Untreated anxiety among adult primary care patients in a health maintenance organisation. *Archives of General Psychiatry* **51**, 740–750.

Fitzpatrick, R. (1994). Applications of health status measures. In *Measuring Health and Medical Outcomes* (ed. C. Jenkinson), pp. 27–41. UCL Press: London.

Fitzpatrick, R., Fletcher, A., Gore, S., Jones, D., Spiegelhalter, D. & Cox, D. (1992). Quality of life measures in health care. I: Applications and issues in assessment. *British Medical Journal* **305**, 1074–1077.

Gilbody, S. M. & Whitty, P. A. (2002). Improving the delivery and organisation of mental health services: beyond the conventional RCT. *British Journal of Psychiatry* **177**, 13–18.

Gilbody, S. M., House, A. O. & Sheldon, T. A. (2001). Routinely administered questionnaires for depression and anxiety: a systematic review. *British Medical Journal* **322**, 406–409.

Gilbody, S. M., House, A. O. & Sheldon, T. A. (2002). UK psychiatrists do not use outcomes measures – a national survey. *British Journal of Psychiatry* **180**, 101–103.

Goldberg, D. (1986). The use of the general health questionnaire in clinical work. *British Medical Journal* **293**, 1188–1189.

Goldberg, D. & Huxley, P. (1980). *Mental Illness in the Community.* Tavistock: London.

Goldsmith, G. & Brodwick, M. (1989). Assessing the functional status of older patients with chronic illness. *Family Medicine* **21**, 38–41.

Greenfield, S. & Nelson, E. C. (1992). Recent developments and future issues in the use of health status assessment measures in clinical settings. *Medical Care* **30** (suppl. 5), Ms23–41.

Jadad, A. R., Moore, R. A. & Carroll, D. (1996). Assessing the quality of reports of randomized clinical trails: is blinding necessary? *Controlled Clinical Trials* **17**, 1–12.

Jenkinson, C. (1994). *Measuring Health and Medical Outcomes.* UCL Press: London.

Jette, A. M., Davies, A. R. & Calkins, D. R. (1986). The Functional Status Questionnaire: its reliability and validity when used in primary care. *Journal of General Internal Medicine* **1**, 143–149.

Johnson, S., Thornicroft, G., Phelan, M. & Slade, M. (1996). Assessing needs for mental health services. In *Mental Health Outcome Measures* (ed. G. Thornicroft and M. Tansella), pp. 217–226. Springer Verlag: Berlin.

Kazis, L. E., Callahan, L. F., Meenan, R. F. & Pincus, T. S. O. (1990). Health status reports in the care of patients with rheumatoid arthritis. *Journal of Clinical Epidemiology* **43**, 1243–1253.

Lehman, A. F. (1983*a*). The effect of psychiatric symptoms on quality of life assessments among the chronically mentally ill. *Evaluating Programme Planning* **6**, 143–151.

Lehman, A. F. (1983*b*). The well being of chronic mental patients: assessing their quality of life. *Archive of General Psychiatry* **40**, 369–373.

Lohr, K. N. (1988). Outcome measurement: concepts and questions. *Inquiry* **25**, 37–50.

McHorney, C. A. & Tarlov, A. R. (1994). The use of health status measures for individual patient level applications: problems and prospects. *Quality of Life Research* **3**, 43–44.

McHorney, C. A. & Ware, J. E. Jr. (1995). Construction and validation of an alternate from general mental health scale for the Medical Outcomes Study Short-Form 36-Item Health Survey. *Medical Care* **33**, 15–28.

McHorney, C. A., Ware, J. E. Jr. & Raczek, A. E. (1993). The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Medical Care* **31**, 247–263.

Mathias, S. D., Fifer, S. K., Mazonson, P. D., Lubeck, D. P., Beusching, D. P. & Patrick, D. L. (1994). Necessary but not

sufficient: the effect of screening and feedback on outcomes of primary care patients with untreated anxiety. *Journal of Internal Medicine* **9**, 606–615.

Mazonson, P. D., Mathias, S. D., Fifer, S. K., Beusching, D. P. & Patrick, D. L. (1994). The mental health patient profile: does it change primary care physicians practice patterns? *Journal of the American Board of Family Practice* **9**, 336–345.

Meenan, R. F. (1982). The AIMS approach to health status measurement: conceptual background and measurement properties. *Journal of Rheumatology* **9**, 785–788.

Mulrow, C. D. & Oxman, A. D. (1999). *Cochrane Collaboration Handbook* [updated June 1999]. In *The Cochrane Library*. Update Software: Oxford.

National Institute of Mental Health (1987). *Toward a Model Plan for a Comprehensive Community Based Mental Health System*. National Institute of Mental Health: Washington, DC.

Nelson, E. C., Landgraf, J. M., Hays, R. D., Wasson, J. H. & Kirk, J. W. (1990). The functional status of patients: how can it be measured in physician's offices? *Medical Care* **28**, 1111–1126.

NHS Centre for Reviews and Dissemination (2001). *Undertaking Systematic Reviews of Research on Effectiveness: CRD Report 4*, 2nd edn. University of York: York.

Orley, J., Saxena, S. & Herman, H. (1998). Quality of life and mental illness. *British Journal of Psychiatry* **172**, 291–293.

Phelan, M., Slade, M., Thornicroft, G., Dunn, D., Holloway, F. & Wykes, T. (1995). The Camberwell Assessment of Need (CAN): the validity and reliability of an instrument to assess the needs of people with severe mental illness. *British Journal of Psychiatry* **167**, 589–595.

Rubenstein, L. V., Calkins, D. R. & Young, R. T. (1989). Improving patient functioning: a randomised trial of functional disability screening. *Annals of Internal Medicine* **111**, 836–842.

Rubenstein, L. V., McCoy, J. M., Cope, D. W., Barrett, P. A., Hirsch, S. H. & Messer, K. S. (1995). Improving patient quality of life with feedback to physicians about functional status. *Journal of General Internal Medicine* **10**, 607–614.

Russo, J., Trujillo, C. A., Wingerson, D., Decker, K., Ries, R., Wetzler, H. & Byrne, R-P. (1998). The MOS 36-Item Short Form Health Survey: reliability, validity, and preliminary findings in schizophrenic outpatients. *Medical Care* **36**, 752–756.

Sainfort, F., Becker, M. & Diamond, R. (1996). Judgements of quality of life of individuals with severe mental disorders: patient self report versus provider perspectives. *American Journal of Psychiatry* **153**, 497–502.

Schulz, K. F., Chalmers, I., Hayes, R. J. & Altman, D. G. (1995). Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Journal of the American Medical Association* **273**, 408–412.

Secretary of State for Health (1999). *National Service Framework – Mental Health*. HMSO: London.

Sharma, V. K., Wilkinson, G. & Fear, S. (1999). Health of the Nation Outcome Scales: a case study in general psychiatry. *British Journal of Psychiatry* **174**, 395–398.

Sprangers, M. A. & Aaranson, N. K. (1992). The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease: a review. *Journal of Clinical Epidemiology* **45**, 743–760.

Stein, G. S. (1999). Usefulness of the Health of the Nation Outcome Scales. *British Journal of Psychiatry* **174**, 375–377.

Street, R. L. Jr., Gold, W. R. & McDowell, T. (1994). Using health status surveys in medical consultations. *Medical Care* **32**, 732–744.

Thornicroft, G., Brewin, C. & Wing, J. (1992). *Measuring Mental Health Needs*. Royal College of Psychiatrists: London.

Ukoumunne, O. C., Gulliford, M. C., Chin, S., Sterne, A. C. & Burney, P. G. J. (1999). Methods for evaluating area-wide and organisation based interventions in health and health care: a systematic review. *Health Technology Assessment* **3**, 1–98.

van Hemert, A. M., Hengeveld, M. W., Bolk, J. H., Rooijmans, H. G. & Vanden Broucke, J. P. S. O. (1993). Psychiatric disorders in relation to medical illness among patients of a general medical out-patient clinic. *Psychological Medicine* **23**, 167–173.

Wagner, A. K., Ehrenberg, B. L., Tran, T. A., Bungay, K. M., Cynn, D. J. & Rodgers, W. H. (1997). Patient based health status measurement in clinical practice: a study of its impact in epilepsy patients. *Quality of Life Research* **6**, 329–341.

Walker, S. R. & Rosser, R. M. (eds.). (1993). *Quality of Life Assessment: The Key Issues of the 1990s*. Kluwer Academic Press: Lancaster.

Ware, J. E. (1995). The status of health assessment in 1994. *Annual Review of Public Health* **16**, 327–354.

Ware, J. E. Jr. & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care* **30**, 473–483.

Ware, J. E., Snoww, K. K., Kosinski, M. & Gandek, B. (1993). *SF-36 Health Survey: Manual and Interpretation Guide*. The Health Institute, New England Medical Centre: Boston, MA.

Wasson, J., Hays, R., Rubenstein, L., Nelson, E., Leaning, J., Johnson, D., Keller, A., Landgraf, J. & Rosenkrans, C. (1992*a*). The short-term effect of patient health status assessment in a health maintenance organization. *Quality of Life Research* **1**, 99–106.

Wasson, J., Keller, A., Rubenstein, L., Hays, R., Nelson, E. & Johnson, D. (1992*b*). Benefits and obstacles of health status assessment in ambulatory settings. The clinician's point of view. The Dartmouth Primary Care COOP Project. *Medical Care* **30** (suppl. 5), Ms42–49.

Wells, K. B., Stewart, A., Hays, R. D., Burnam, M. A., Rogers, W., Daniels, M., Berry, S., Greenfield, S. & Ware, J. (1989). The functioning and well-being of depressed patients. Results from the Medical Outcomes Study. *Journal of the American Medical Association* **262**, 914–919.

Wing, J. (1994). Measuring mental health outcomes: a perspective from the Royal College of Psychiatrists. In *Outcomes into Clinical Practice* (ed. T. Delamonthe), pp. 147–152. BMJ Publishing: London.

Young, J. B. & Chamberlain, M. A. (1987). The contribution of the Stanford Health Assessment questionnaire in rheumatology clinics. *Clinical Rehabilitation* **1**, 97–100.