Cambridge University Press & Assessment
978-1-904-67192-3 — Mental Health Outcome Measures
Edited by Graham Thornicroft, Michelle Tansella
Excerpt
More Information

Part I
Methodological issues

CHAPTER 1

Measures of outcomes that are valued by service users

Thomas Kabir and Til Wykes

This book has many chapters on outcome assessments, from the global functioning scales to more specific and detailed measures for use in service evaluation as well as in measuring treatment outcome. So we suspect that the reader will be asking a number of questions before dipping into this first chapter, specifically on measures of outcome that are valued by service users. These we are sure will include:

- Will this chapter examine new measures of outcome that are not covered by other chapters?
- Will any new measures look radically different from those already described in other chapters?
- Why is there a chapter on outcomes that are valued by service users?

The answers to the first two questions are: yes, but not many; and probably not. But it is the answer to the last question which is key. A chapter has been specifically allocated to this topic because such measures are important in evaluations of mental healthcare. Many professionals will consider that this is just a focus on what is politically correct and that service users' experiences are not generally that helpful or, more often, that their clinical carers usually know best. We have some sympathy with the view that when healthcare resources are scarce there is a need to measure outcomes that reflect the performance of services and that this is sometimes at variance with service users' views. Service users may not be interested in symptom remission, patient throughput or even the assessment of their global functioning. They may be more interested in their definition of recovery or simply their happiness and sometimes these aspirations may lie outside the remit of mental health service provision.

But even with this understanding of different perspectives, we believe that the pendulum has swung too far and for too long in the direction of outcomes which may not be relevant to service users, may not be appropriately measured even if they are relevant, or may just be the opposite of what service users would expect of a mental health service. In this chapter we try to rebalance the field by introducing or describing outcome measures that are accepted by service users, and show how methods involving service users can be used to develop new, relevant and acceptable measures of outcome that reflect the purpose of services. It is also important that we have measures of what service users expect from services, as their models of appropriate care will undoubtedly affect the take-up of services, especially novel ones. Assessments compared from different perspectives can also provide new insights into how services might improve as well as how new treatments or types of care can be organised. This chapter may overlap with some others in this volume in its description of measures but, crucially, it highlights the need for good measures of satisfaction, service outcome and quality of life that reflect the mental health service user's experience.

What constitutes a user-valued measure?

The answer to this question is simple. A user-valued measure is one that is recognised as such by a majority of service users. A user-valued measure is one which reflects the values and experiences of a majority of service users. To use an analogy, to be a user-valued outcome measure, it would have to pass a 'Turing Test'. A majority of service users would have to be able to say from looking at the measure that it did indeed make sense and evaluate human factors that were important to them. In other words, the majority of service users that the measure applies to have to believe that the scale has been constructed *by people* who have an understanding of their situation.

But this is not the only criterion that we are going to use to judge a measure as worthy of being called a user-valued measure. Clearly, the instrument also needs to be useful in understanding services and therefore we want to know whether it can be used in different settings, with different groups of service users and in different types of mental health organisation. We would like to see evidence not only that the measure is being used in the academic literature but also that it is being used in mental health service assessment. In other words, our aspiration is to see user-valued measures used in routine practice.

Fitzpatrick *et al* (1998) propose eight essential properties of a patient-based outcome measure. These are shown in Table 1.1. One kind of measure which is clearly user valued is the method designed by Diana Rose and called user-focused monitoring (UFM) (Rose *et al*, 1998). This is a general method of involving service users in producing a schedule of assessment that is grounded in their views within a particular service setting. The questions therefore vary between services rather than being generally appropriate. So although this method is one of the best for producing a measure with user value, it does not fulfil the criteria we have set out, as each UFM measure will be service specific and therefore it would not be 'appropriate' (see Table 1.1) for the measure to be generalised.

This book concentrates on outcome measures and these are normally used in evaluations of treatments or services. We therefore need to be certain

Cambridge University Press & Assessment
978-1-904-67192-3 — Mental Health Outcome Measures
Edited by Graham Thornicroft, Michelle Tansella
Excerpt
More Information

Table 1.1 Essential properties of an outcome measure

Property	
Appropriateness	Is the content of the instrument appropriate to the question that the clinical trial is intended to address?
Reliability	Does the instrument produce results that are reproducible and internally consistent?
Validity	Does the instrument measure what it claims to measure?
Responsiveness	Does the instrument detect changes over time that matter to patients?
Precision	How precise are the scores of the instrument?
Interpretability	How interpretable are the scores of the instrument?
Acceptability	Is the instrument acceptable to the patients?
Feasibility	Is the instrument easy to administer and process?

Adapted from Fitzpatrick *et al* (1998) and Gilbody *et al* (2003).

that the information is accurate; that is, if we repeated the evaluation we would gain similar results. These psychometric qualities of any instrument therefore also need to be available for a measure to begin to be used in routine practice. The usual list includes: validity, reliability and, crucially, sensitivity to change over time. The reason for the importance of psychometric data is highlighted in a report of the outcome measures used in 300 trials randomly selected from the Cochrane database of trials in schizophrenia (Marshall *et al*, 2000). The review showed that outcome measures that were developed for a specific study and did not have adequately published reliability and validity data were more likely to conclude that the new treatment was superior to a control treatment. When pharmacological and non-pharmacological studies were independently assessed, the non-pharmacological treatments fared the worst, with nearly twice the chance of a report of clinical significance if an unpublished rather than a published scale had been used. Overall, the authors concluded that in more than one-third of all studies in this area the claims for clinical significance could not have been made if published scales had been used. Although this was a report of studies of schizophrenia, we assume that similar results would have been found for other disorders. The fact that the problems seemed to be amplified in the non-pharmacological area is of great importance to service users, who tend to prefer these sorts of treatments to medication alone. For service users, the diversion of resources into therapies which would not have been found to be useful if rigorously assessed scales had been used is a further reason to ensure that all studies use such scales.

It might be expected that this chapter would describe a set of outcomes with a specific user value attached to them. However, this is not possible, as there have been few investigations of what types of outcome measure are acceptable and have value. The approach that we have taken is to investigate those measures to which service users have contributed in any way, as these are likely to have value; we also point out ones that users may value but where they have not been involved in their development. One clear research agenda would be to investigate the value expert service users (i.e. those with experience of the disorder or service) would consider as an outcome worth measuring.

It should be noted, however, that a measure need not be constructed exclusively by people with experience of using mental health services for it to be a user-valued outcome measure. It is not good scientific practice for the validity of any measure to be dependent on the background of those who constructed it. It is only necessary that the majority of service users are able to say that the outcome measure does indeed measure factors that are important to them.

In what types of study would user-valued measures be relevant?

It is not controversial to suggest that service users' views are important in assessing the value of services. Hence, most user-valued measures of outcome have come from the field of service satisfaction. But we believe that user-valued measures could be used to understand treatment outcome, including side-effects in randomised controlled trials.

Which treatments are included in guidance for healthcare in the UK is determined by the National Institute for Health and Clinical Excellence (NICE), which makes its judgements on the basis of the evidence from high-quality controlled trials. These are nearly always studies where the primary outcome is symptoms and the secondary outcomes are those of interest to the clinicians. Often the only information in the evidence base about whether service users value the treatment is their decision to opt out of the study, that is the drop-out rate. Given that most decisions about treatments are based on the balance of outcomes, surely it would be helpful to allow service users to state their choice? For instance, although efficacy is based on symptom reduction, the severity of side-effects may make it unlikely that the treatment will be prolonged. An outcome measure which shows this balance might provide clearer information to the healthcare professionals and service providers about the likely effectiveness of treatments.

There is a further reason for introducing user-valued measures, and that is the acceptance of the term 'recovery' rather than 'cure' in the mental health literature (Davidson *et al*, 2005). The acceptance of new terms will change the measurement field to include issues of balance in the service user's life between a number of competing choices; one specific measure is unlikely to reflect this complex world view. For example, consider someone who is on antipsychotic medication which has reduced their clinical symptoms to the point where they might clinically be considered to have recovered but this person has experienced significant medication-related side-effects, including weight gain, which has led to a negative self-image and social withdrawal, and drowsiness, which has impaired their ability to carry out his usual occupation. It is unlikely that a single outcome measure would capture all the information that is pertinent to this service user's situation.

The production of user-valued measures

The majority of outcome measures are produced through a process of item generation, item testing, scale feasibility testing and further work on the psychometrics. What need to be added to this list are those extra items that service users would most notice about a scale, regardless of the content. Some of these have been described by Fitzpatrick *et al* (1998). Of prime importance is the acceptability of the scale to service users, which includes the length of time needed to complete it and whether there is a chance for them to add their own context to their answers. Scales that are administered within an interview do give some time for service users to reflect on their answers (although this does depend on the skills of the interviewer). Scales with fixed items that are completed by self-report are unlikely to reflect the key issues for an individual. It is vital in this latter case to ensure that the items chosen and the categories for the answers allow service users to give an accurate rendition of their perspective. For this, they need to be involved in the production of the scale items and category assembly. Below we provide a brief description of the continuum of involvement in the outcome measures we have reviewed. This is based on a continuum adopted by a UK Department of Health funded patient involvement group called INVOLVE (www.invo.org.uk). However, we recognise that this system is too simple to distinguish some subtleties in service user involvement. In our opinion the best possible measures would be those:

- where service users had been closely involved in the development as full partners
- which have been rigorously tested for their psychometric properties (especially sensitivity to change)
- which have been tested in a variety of service settings in a variety of countries.

The few such measures on the market are described below. First, though, we describe the methods used to develop user-valued measures, as this will allow researchers to carry out similar exercises where there are gaps in the field.

Derived without user involvement

Even if a measure is derived without their involvement, the end result may still be valued by service users. It does seem unlikely that this could happen but there are some measures that have been produced with no consultation with service users which are valued. These include quality-of-life measures and some service-satisfaction measures. These scales were produced through the traditional route, by clinicians working with academics. The feasibility of the scales was tested with service users but they were not involved in item generation. An example of such a scale is the Working Alliance Inventory (WAI; Hovrath & Greenberg, 1989). The scale can be said to be user-valued partly because of its subsequent usage and because 'some of the definitions used include the client's experience of the relationship with the therapist as being helpful in achieving goals' (Tracey *et al*, 1989). There are two versions of this scale, one for the patient and one for the counsellor. The 36 items contained in both versions of the scale were generated by consulting seven experts in the field of working alliance and 21 randomly selected psychologists. The WAI was piloted using 29 graduate students in a 'counselling psychology program engaged in a peer counselling task' (Hovrath & Greenberg, 1989). This outcome measure was specifically chosen as an outcome in one study of motivational interviewing (Hayward *et al*, 2009).

Consultation

Consultation can take various forms, from representation on steering groups to surveys of service users. This can further be refined by looking at the level of representation, through service users' voluntary organisations or through representation of service user experts. Consultation does not imply that the research team (where the power lies) will accept the advice given. Outcome measures derived in this way may or may not be valued by service users – usually this has not been tested – but do show some involvement and so are more likely to be valued by service users than those measures developed with no user involvement.

Representatives of services users were involved in the development of the Carers' and Users' Expectation of Service – User Version (CUES–U; Lelliott *et al*, 2001; Blenkiron *et al*, 2003). This scale was the result of a collaboration between two Royal Colleges, the University of East Anglia, and an organisation representing service users and carers (then called National Schizophrenia Fellowship, now called Rethink). Domains that were important to service users were identified in two ways. The first was by carrying out a comprehensive literature search, including a search of the 'grey literature' and unpublished material. The second was by carrying out focus group interviews with service users. The development of the scale was also informed by the establishment of an advisory group of service users. Once the large number of domains that were important to service users was identified, then the domains were 'grouped into the smallest number of items without losing definition or meaning' (Lelliott *et al*, 2001). The scale was field tested using nearly 450 service users from across the UK.

The final scale is self-rated and consists of 17 items, split into three sections. The items concern: where you live; money; help with finances;

Cambridge University Press & Assessment
978-1-904-67192-3 — Mental Health Outcome Measures
Edited by Graham Thornicroft, Michelle Tansella
Excerpt
More Information

how you spend your day; family and friends; social life; information and advice; access to mental health services; choice of mental health services; relationships with mental health workers; consultation and control; advocacy; stigma and discrimination; medication; access to physical health services; relationships with physical health workers; and other issues (anything else that the service user would like to raise which has not been covered by the other 16 items). The paper reporting the development of the scale notes that service users 'appear to place less emphasis on symptom reduction than they do on improvements in other area of their lives' (Lelliott *et al*, 2001). A slightly different level of consultation took place to produce the Camberwell Assessment of Need (CAN) (Phelan *et al*, 1995, see http://www.iop.kcl.ac.uk/virtual/?path=/hsr/prism/can/).

A slightly different level of consultation took place to produce the Camberwell Assessment of Need (CAN)9 (see http://www.iop.kcl.ac.uk/virtual/?path=/hsr/prism/can/). This measure reflects the needs (both met and unmet) of people in touch with high-support services. Although the development of the items was based on consultation between clinicians and researchers, the development team also included a survey of service users about the importance of the items that were proposed to make up the scale. All the proposed items were judged to be at least moderately important but there was no opportunity for service users to contribute new items or to suggest ways in which the scale should be promoted, including the anchoring of the scale. However, at least we can be clear that this measure does include moderately important items and so has some value to service users.

Service users were also involved in a consultative way in the development Patient Reported Outcome Measures (PROMS; see http://phi.uhce.ox.ac.uk/home.php). Continuity-of-care measures where service users were certainly consulted include the CONNECT (Ware et al, 2003) and ACSS–MH (Durbin *et al*, 2004) measures.

Contribution

This term is used to denote that service users have been consulted and that their views have been taken into account. This differs from consultation, where views may be sought but need not have had any effect on the research. This terminology is taken from Sweeney & Morgan (2009), who specifically point out that 'contribution' is where some commitment to service user involvement has been agreed (unlike in consultation) but that the role is limited. For instance, it may mean only that a service user has been employed as a researcher on a project which has already been designed.

The CUES–U measure is likely to be one in which service users and carers were both consulted and made a contribution. The measure itself was in part constructed using focus groups of service users. The scale also contains a number of items, such as 'stigma and discrimination' and 'medication', which are commonly reported by service users as being important to them.

It can therefore be said that, to a reasonable degree, the CUES–U measure does indeed take into account the views of service users and carers.

A similar process was undertaken in the development of the Accommodation and Enabling Scale for Eating Disorders (AESED; Sepulveda *et al*, 2009). Two expert carers were involved in screening items for use in this questionnaire.

Partnership

Again, there is a continuum of involvement in measures that have been developed in partnership with service users. In some partnerships the academic researcher takes the lead but in others service users lead the research and develop the measures through new methodologies or participatory research.

The first measures we highlight here are those where service users took part as either paid or voluntary researchers. One example of this sort of development is a measure of the outcome of cognitive–behavioural therapy for people with psychosis (CBTp), called Choice of Outcome in CBT for psychoses, or CHOICE (Greenwood *et al*, 2010). The study team included a paid service user researcher, who was involved in the design of the study. The first step in the measure development was that topics for exploration were identified by CBTp experts and these were then discussed in service user focus groups, which were jointly led and analysed by a service user and a clinical researcher. Focus group participants then took part in a ‘Delphi exercise’ to work towards a consensus on the items that needed to be included. This produced a total of 26 items. An acceptability study with 15 further service users provided information on whether the questionnaire was understandable; these users also commented on different possible formats. The measure was then given to 152 further service users before and after they had received CBTp as part of a standard clinic. Reliability, validity and sensitivity were measured. The final measure covers both recovery-oriented items, such as general distress and quality of life, as well as CBTp-specific ones, such as psychosis-specific distress. There are data relating to its psychometric properties and therefore this is a measure which holds great promise for what is now a burgeoning area, not only in research studies but also in the provision of CBTp in routine care.

Another method of including service users has been developed from a health economics method called conjoint analysis, which allows service users specifically to influence the weighting of items in a scale. One measure that utilises this approach is the Older Person’s Utility Scale for Social Care (OPUS; Netten *et al*, 2002). This scale measures key outcomes for social care. The research team included service users at different stages. At the first stage, ‘experts’ in social care, from social care policy-makers to care managers, were consulted about the topics that might be included. At the second stage, 356 older people were consulted about their preferences in relation to 27 different scenarios covering five different domains; these had been piloted in consultation with older people. In this part of the project, older people were given two descriptions which differed on two key domains, for example where personal safety was maximised in one but social participation reduced, compared with another where personal safety was reduced but social participation was increased. Each of the scenarios reflected different levels of unmet need in each of the five domains. This part of the development also included 58 people who indicated additional preferences on the basis of how much money they would pay for their choice. The results of these preference decisions were analysed using conjoint analysis, which produced weights for the different domains. There was a clear ranking by both the older people and those who were using services of the domains from least to most important: personal safety, food, and control over daily life, social participation and personal care. This method, by giving weights to the items, allows a score to be produced which can reflect, adequately, the importance of each item rather than each item contributing the same amount to the total score. Further testing showed that the scale was both reliable and valid.

Service user researchers have also led teams in the development of measures. One of these is a new participatory model based on UFM methodology and constructed by the Service User Research Enterprise (SURE) at the Institute of Psychiatry, King’s College London. In this method, the service user leads the research project. The process falls into five stages:

- production of a topic guide
- pilot and then full focus groups to generate specific domains
- expert panels to comment on the item form and completeness
- feasibility study to test comprehensibility and acceptability of the measure
- formal psychometric testing including reliability, validity and sensitivity to change.

Two measures of value to service users have been developed using this method. The first is a measure of continuity of care which is thought to be important by service planners and policy-makers (CONTINU–UM; Rose *et al*, 2009). This is a measure of the perceptions of continuity of service users (which differ from service providers’ and policy-makers’ perceptions of continuity). The second measure developed using this method records service users’ perceptions of the in-patient environment and is called VOICE. Both measures have been or will be used as a main outcome in studies investigating service and treatment changes.

The Empowerment Scale is another example of an outcome measure which has been designed using a participation action research method. The scale was developed by a defined partnership between researchers and a consumer research advisory board comprising service users and one of the researchers (Rogers *et al*, 1997).