

Meaningful & Measurable

A Collaborative Action Research Project

Developing Approaches to the Analysis & Use of Personal Outcomes Data

Measuring Personal Outcomes in Service Settings: Collected Briefings from the Meaningful and Measurable

Karen Barrie & Emma Miller

May 2015



Measuring Personal Outcomes in Service Settings



Background

Meaningful and Measurable was an ESRC funded collaborative action research project that ran from November 2013 until March 2015, with an explicit focus on using personal outcomes data in service settings. Developing approaches to the recording, analysis and use of personal outcomes data had been found to pose challenges at all levels of relevant organisations. It was already established that measuring outcomes itself presents a range of challenges, including the question of attribution and the need to balance measurement with meaning (Miller 2011). With regard to the latter point, it has been a longstanding concern to ensure that the predominant emphasis on and drive to measure outcomes does not undermine the quality of interaction (Cook and Miller 2012) and this remained a priority through the lifespan of the project.

Broadly, the aspiration is to use both qualitative and quantitative data to inform decision-making. However, standard practice in most organisations has focused almost exclusively on gathering, analysing and reporting quantitative information. During the project it became apparent that this focus meant that understandings of what you can and cannot do with qualitative data remain limited. We published briefings in a separate resource on the use of qualitative data in response to issues arising in the project ([Barrie and Miller 2015](#)).

This resource has a more quantitative focus. While ‘measurement’ was not the explicit focus of any of the eight action research projects carried out by project partners, the use of quantitative methods within encounters in practice featured extensively in partner discussions throughout. These discussions shaped this resource and we draw on previous evidence about measuring outcomes. The resource summarises learning on the following themes in the form of four briefings:

1. The limitations of establishing causality in addressing social issues, and the implications for measurement
2. What are we measuring: How personal outcomes have been conceptualised, with reference to approaches emphasising wellbeing as compared to personal outcomes
3. What conclusions can be drawn from the Meaningful and Measurable project with regard to the categorisation of outcomes
4. What conclusions can be drawn from the Meaningful and Measurable project with regard to scale measures?

Finally, it should be noted that distinctions between ‘*measurement for judgement*’ and ‘*measurement for improvement*’ have been examined previously in relation to an outcomes approach (Miller, 2012) and are not repeated within this resource. However, these distinctions call attention to the importance of clarifying ‘purpose’ and are implicit throughout.



Briefing 1: Limitations of Establishing ‘Causality’ in Addressing Social Issues, and Implications for Measurement

“Measurement in the public sector is less about precision and more about increasing understanding and knowledge. It is about increasing what we know about what works in an area and thereby reducing uncertainty”.

Mayne 1999, p5

Overview

A significant element of international public sector reform over many years has been the shift from a management regime focused on rules and procedures toward an approach that pays greater attention to the results that matter to the public. This involves a focus on results and impact, raising challenges about how to effectively measure outcomes, including the problem of attribution (Mayne, 1999). The extent to which outcomes can definitively be attributed to a particular programme will always be limited.

In developing and embedding outcomes based working in Scotland, personal outcomes have been described as the ‘missing piece of the jigsaw’ (Cook and Miller, 2012) alongside other data sources required for decision-making, such as those concerned with throughput and cost-effectiveness. There was a decisive shift in thinking from ‘attribution’ to ‘contribution’ during the initial outcomes data retreat (Meaningful and Measurable, 2012). This included recognition that outcomes are co-produced, often involving networks of support, and bound up with diverse influences. The preferred concept of contribution is helpful in the context of integration, whereby different agencies can contribute towards the same outcomes, and importantly, the role of the person can be acknowledged as a contributor to their own outcomes (Cook and Miller, 2012). While the concept of contribution is helpful, the tendency to rely on quantitative data remains a limitation (Mayne, 1999) as will be described broadly in the Scottish context below. We intend to report in detail on this theme in the near future.

The early Meaningful and Measurable project findings reported here, with wider literature, facilitate a critique of the real world applicability of many established principles of **reliability** (concerned with consistency, particularly over time) and **measurement validity** (the extent to which we are measuring what we claim to be measuring) when working with personal outcomes data (Bryman, 2012). Part of the challenge involves finding credible means of demonstrating contribution towards results, in unpredictable human service contexts. While this means that cause and effect cannot be definitively ‘proven’, a key challenge is to maximise data quality, while recognising their limits, in order to be facilitate data use.

Limitations and Possibilities of Quantitative Outcomes Data

Early in 2014, several project partners directly compared their quantitative and qualitative outcomes data, leading to two sets of concerns to be explored in the second phases of the projects. The first concerns related to how narrative data about personal outcomes were recorded, leading to a separate guide on recording outcomes (Miller and Barrie 2015). The second set of concerns related to the identification of limitations in the scale measure data, raising questions about the validity of quantitative modes of inquiry into subjective



experiences. While analysis of the data is ongoing, and will be reported in more detail in the near future, a few of these concerns are briefly considered here.

A key concern raised in early discussions was that percentage outcomes scores can give an artificial sense of accuracy and mask important differences. In particular, partners recorded scale ratings differed between cases where supporting qualitative accounts seem very similar. Conversely, identical ratings were associated with seemingly very different accounts of progress towards outcomes. Alongside this, fixed response options were found to inadequately capture the diverse and conflicting experiences found within qualitative accounts, where typically a mid-range response option is selected to cancel out experiences at both extremes. This lack of texture identified in data analysis was subsequently identified in practitioner interviews, which highlighted the inability of fixed response options to adequately cater for what one project partner described as the ‘shades of grey’ experienced in choosing between response options.

Lack of equivalence of meaning is a familiar issue when working with the type of indefinite response categories used by project partners, such as ‘Big Improvement, Moderate Improvement, Small Improvement’; ‘Met, Partially Met, Unmet’ or ‘Sometimes, Often’ (Mallinson, 2002). One person’s threshold between options may be very different from another’s. This gives rise to the consequence that individuals with seemingly similar progress may not have identical data. Generally, this does not mean that the data based on indefinite response categories are incorrect as the variations are expected to even out across sufficiently large populations or random samples. One project partner has reported practical benefits at the organisational level from their approach to outcomes measurement (Monger et al, 2012). However, caution is warranted in comparing different groups in different contexts, with significant implications for comparability of data.

In addition, project partners identified apparent conflicts between qualitative accounts and the associated scale ratings. Such inconsistencies are not entirely unfamiliar within the wider literature, with more advanced cognitive interviewing techniques for instance uncovering the very diverse, highly context-specific and often unanticipated processes that individuals use to choose between fixed response options (Rapkin and Schwartz, 2004). Current measurement theory views these individual response variations as sources of error and tries to eliminate them. However, it has been argued that these variations are intrinsic to the appraisal of subjective phenomena such as quality of life and, rather than trying to eliminate them, effort should be expended in securing a better understanding (Schwartz and Rapkin, 2004). This argument would appear to hold when working with personal outcomes data to inform service improvement, redesign and commissioning decisions.

A further concern highlighted by some project partners, relevant to notions of attribution/contribution, relates to the fact that reported changes in personal outcomes sometimes appear to be the consequence of altered perceptions rather than to specific service inputs or supports. This is known as ‘response shift’, whereby changes reported may be due to a natural process of adaptation, resulting in recalibration of internal rating scales, the relative importance of different outcome types, or even what constitutes a good quality of life (Schwartz and Rapkin, 2004).



The limitations of considering quantitative data only were further highlighted by the Meaningful and Measurable partners through the opportunity to directly engage with qualitative data in individual assessment and review records while conducting local action research projects. Specifically, they reported heightened awareness of the complexity of people's lives, the difficulties practitioners face in addressing this complexity, and the limitations of simplistic notions of causality when contemplating social issues. The effects at times were profound.

Causality and '**internal validity**' are key considerations in quantitative research, broadly concerned with how confident we can be that an independent variable (e.g. a service) is responsible for the variation identified in the dependent variable (e.g. personal outcomes). The concept of internal validity does not directly translate into qualitative research. However, it is important for qualitative researchers to be able to demonstrate that they are accurately reflecting the phenomena under study, as perceived by the study population, studies and it has been argued that alternative criterion of a '**credibility**' (or 'believability') be used instead (Lincoln and Guba, 1985). Assessing credibility raises subsidiary questions concerning the strength of the study design and conduct, and the quality of analysis and interpretation that takes place. The insights from the project partners seem consistent with the proposed use of the alternative concept of 'credibility' when working with personal outcomes data.



Briefing 2: ‘What’ are We Measuring: How ‘Personal Outcomes’ Have Been Conceptualised

A key question raised by the Meaningful and Measurable project partners related to the conceptualisation of ‘personal outcomes.’ A particular concern here was whether we are measuring ‘*outcomes important to people using services and unpaid carers*’ or measuring ‘*wellbeing*’. It is important to understand these different conceptualisations as they have significant implications for how personal outcomes are measured.

Outcomes Important to People Who Use Services and Unpaid Carers

Six project partners (five statutory sector and one third sector provider) are concerned with the measurement of ‘*outcomes important to people who use health and social care services or unpaid carers*.’ These partners draw to varying extents upon the Talking Points Personal Outcomes Approach. The conceptual underpinnings of Talking Points are described in full in the Practical Guide (Cook and Miller, 2012), but in short, each framework is founded upon extensive research into the views of people using services / unpaid carers about the outcomes important to them in life and the impact that services have on outcomes. Each framework sets out in very broad terms the high-level categories of outcomes that matter which, for adult service users living in the community, fall into three main types:

Process Outcomes are the outcomes that individuals experience through seeking, obtaining and using services and supports. [While many outcomes frameworks purposefully exclude process outcomes in order to maintain a separation between the impact of services and the way in which they are provided, outcomes such as being listened to, respected and treated as an individual have consistently been found to have intrinsic as well as instrumental value and are therefore included].

Change Outcomes relate to the typically time-limited improvements in functioning and wellbeing that individuals are seeking, often from a particular service or support.

Quality of Life Outcomes are the aspects of a person’s whole life that they are working to achieve or maintain in partnership with services and other forms of support.

Additional frameworks have been developed for older people living in care homes and for unpaid carers. Each framework provides a structure for thinking about the outcomes that people might be working towards through their engagement with services, for organising outcomes data, and reporting in a language that is understood by people using services.

The specific categories within each framework have been found to be sufficiently high level to capture most outcomes important to most people. This is not to say that all outcome categories will be applicable in all service settings. Equally, in some cases, experience of using the framework has led to the identification of specific additional outcome categories that, while not universally applicable, are relevant for inclusion by particular services (Cook and Miller, 2012). Provider organisations such as VOCAL report that the ability to start with limited core outcome categories and then review and expand them to accommodate issues raised by unpaid carers and practitioners has been critical.



Wellbeing

The remaining two project partners, Penumbra and Angus Council Children and Families Services, in contrast, understand personal outcomes as '*aspects of wellbeing*'. Angus Council developed the Wellbeing Web based upon the Scottish Government's national practice model 'Getting it Right for Every Child' (GIRFEC) and the SHANARRI model of Wellbeing, which comprises eight indicators. Penumbra developed the individual Recovery Outcomes Counter (I.ROC) in-house, based on its HOPE model of wellbeing, which comprises 12 indicators.

Both organisations understand wellbeing as a multi-dimensional concept. As is the case for the different categories of outcomes described above, the inter-relationship between the dimensions of wellbeing is critical, non-linear and points to inter-dependencies between various aspects of an individual's life and the lives of others.

However, wellbeing differs from understandings of 'outcomes important to people who use services or unpaid carers' in a number of critical respects:

- First, wellbeing is important to everyone, whether or not they use services. Although quality of life outcomes correspond quite closely with aspects of wellbeing, 'process outcomes' and time-limited 'change outcomes', directly associated with service use, do not feature within the wellbeing dimensions.
- Second, wellbeing dimensions are incommensurable and irreducible, and each has intrinsic value (Alkire, 2002). They also exist in balance with each other. This underscores the importance of attending to and indeed visualising all wellbeing dimensions in the round. With personal outcomes, it is possible to focus on selected outcomes relevant to the individual.
- Thirdly, the notion of different aspects of wellbeing existing in balance also suggests the possibility of imbalance. This calls into question the implicit measurement assumption that that 'more' or a 'higher score' is necessarily better and that all aspects of wellbeing should improve (or be maintained). This is illustrated poignantly through the Angus Council project and the reported experiences of young carers who had assumed too much responsibility for their age and whose 'Responsible scores' would have to decrease in order for improvements in other aspects of wellbeing to be achieved. This has significant implications for the approach to measurement, highlighting the need to contextualise scale ratings through supporting text.
- Finally, as aspects of human existence, wellbeing dimensions cannot be 'achieved' once-and-for-all (Alkire, 2002). (Equally, 'recovery' is commonly described as a personal 'journey' with 'no fixed endpoint' and 'no pre-defined route', in the course of which individuals are enabled to live meaningful and satisfying lives, whether or not they continue to experience symptoms of mental illness). It is therefore more appropriate to speak of 'pursuing' or 'improving' wellbeing, or 'realising' an aspect of a dimension of wellbeing. This again has significant implications for the approach to measurement, with scales such as 'unmet / met' simply not making sense, and use of scale points based on fixed assumptions rejected.

How Are We Measuring Personal Outcomes?

The two different conceptualisations of 'personal outcomes' held by the project partners have given rise to two broadly different ways of 'measuring personal outcomes':



1. Standardised Measures Developed from Collective Views of People Using Services

Both Angus Council and Penumbra measure wellbeing using a traditional quantitative approach whereby wellbeing is operationally defined using multiple indicators. In both cases the indicators (which take the form of standardised questions and accompanying prompts) have been developed from service user views, strengthened by wider literature review. The indicator question items are combined with numeric (interval) scales which support statistical manipulation and are capable of visual representation at individual, service or organisational level in the form of a 'web'.

Measurement is a key preoccupation in quantitative inquiry and as with any 'measure', a measure of wellbeing is expected to provide a consistent device or yardstick for making comparisons and gauging differences in reliable and valid ways. Psychometric tests have been developed to assess the reliability and validity of measures. The reliability and validity of the wellbeing measures developed by the two project partners is discussed subsequently in relation to the operationalisation of the wellbeing dimensions and the associated measurement scales.

2. Aggregation and Counting of 'User-defined' Responses to Open-ended Questions

Most of the remaining project partners 'measure' outcomes important to people by gathering responses to open ended questions about what matters and why during assessment and planning, which are later reviewed using 'improvement' or 'attainment' (ordinal) scales. These scales can be more easily embedded within the flow of conversations, but present fewer statistical possibilities. As part of this process, five partners classify the user-defined outcomes into pre-defined outcomes categories, again based on collective views and wider literature, as set out within an overarching outcomes framework. Two partners, East Renfrewshire and Moray collate percentages of 'met' or 'unmet' outcomes within each outcome category, which can later be reported. VOCAL uses a simple 5-point improvement scale, which is again embedded in conversations, supported by a tool which also includes space for narrative. At the start of the project, Edinburgh, Bridgend and Stirling had not yet embedded scale measures, but were focusing on outcomes focused engagement and the recording of relevant narrative data.

For each of these partners, the overarching approach is qualitative in nature, grounded in good conversations, and the reliability of the measurement component has tended to be a lesser concern. In terms of 'measurement validity', there is a wealth of research and practice evidence to support the use of the Talking Points Outcomes Frameworks, albeit taking a very different form and drawing upon different assumptions than the psychometric tests used for validating standardised measures, such as those developed to measure wellbeing.

Within and cutting across these two seemingly fundamentally different measurement approaches, a number of debates have arisen concerning:

- The 'categorisation' of personal outcomes and the point at and means by which this happens
- The choice of measurement scale

These two areas of debate are the subject of the last two briefings in this paper, and will be considered in turn.



Briefing 3: Concerns that Relate to the Categorisation of Outcomes

The two distinctive conceptualisations of personal outcomes give rise to a range of approaches to the categorisation of outcomes for measurement purposes.

Distinctions in the Approaches used to Operationalise the Dimensions of Wellbeing and the Relationship with Personal Outcomes

Penumbra operationalise wellbeing using a traditional quantitative measurement approach, namely the development of 12 standardised indicators, informed by service user perspectives and literature review. Each indicator question asks the person *how often* they have *felt* e.g. mentally and emotionally happy or have *been* e.g. involved in group activities within the past three months. The specification of a time period of three months coincides with review frequency and ensures a common reference point. The additional prompts serving to explore each indicator further tend to be quite specific. The 12 indicators that comprise the measure are completed in a set order, and ideally in a single session. This operationalisation of wellbeing therefore engages with frequency and with lived experience in a consistent way that readily supports comparison. I.ROC has been psychometrically tested and validated against other measures of recovery, with extensive investment in this process deemed critical in establishing credibility in the measurement-rich field of mental health and in demonstrating improved outcomes to funders.

However, in taking this approach, personal outcomes may, but need not, directly relate to the standardised indicators that comprise the measure, and are explored and recorded separately. A concern arising in discussions was that the early introduction of standardised indicators into discussions with people using services might shape individual goals or personal outcomes, rather than using the dimensions as loose containers for holding self-determined concerns, and that the person's voice might be thus muted. The counter argument was that the indicators relate to aspects of wellbeing that people consistently find important in the context of recovery, bringing focus and purpose to personalised support planning discussions. On the one hand, this has been found to help to manage the complexities or chaos of the lives of people living with mental illness by breaking things down into manageable areas, while the requirement to attend to all indicators of wellbeing can support people to look at the bigger picture rather than getting bogged down in the latest issue.

Angus Council in contrast do not stipulate a particular time frame in operationalising wellbeing and the eight indicators can be completed in any order and over multiple sessions. For children, each SHANARRI indicator is concerned with how the child *feels* at that point in time [e.g. *I feel safe*]. Moreover, the series of prompts are quite generic and intended simply to open up discussion about what e.g. *feeling safe* means to the child and the extent to which this feeling holds true. The interest is thus not in the '*frequency*', but in the '*truth*'. This approach engages more directly with subjectivity and ostensibly allows for greater interpretation and self-determination, but makes consistent comparison challenging. For parents, each indicator is concerned with the extent to which they are supporting the aspect of the child's wellbeing. The same visual scale is used, allowing comparisons to be made



between the perceptions of the parent, the child, and also the practitioner at any point. The Wellbeing Web has not been psychometrically tested- the tool's capacity to open up dialogue is the key concern, rather than its reliability or validity.

However, the expectation that this approach would open up discussions about personal outcomes and what needs to happen to progress them remains largely unrealised. Practitioners have also indicated that the statutory nature of their work often results in outcomes for the child being prescribed during multi-disciplinary team discussions, rather than negotiated with the family using the Wellbeing Web. While this approach has opened up discussions about the SHANARRI principles and individual feelings about aspects of wellbeing, full integration with outcomes focused support planning has not happened.

Distinctions in the Approaches used to Categorise Personal Outcomes using the Outcomes Framework

There are also important distinctions in the way that organisations using the Talking Points approach identify and map personal outcomes to the high-level outcomes categories within the overarching framework. Talking Points was developed on the basis of an understanding that outcomes focused assessment and planning should flow from semi-structured conversations, loosely based around the framework of outcomes important to people. Not every outcome would necessarily be important to each individual at a given point in time. Following from previous use of the highly structured single shared assessment format, some early versions of locally developed Talking Points tools set an expectation that personal outcomes would be identified by making use of the pre-defined outcome categories within the framework to structure discussions in a manner not dissimilar to that envisaged for the Wellbeing Web. However, as organisations have increasingly prioritised good conversations, there has been a shift from tool-led to more open approaches. This shift has resulted in a requirement to map diverse user-defined outcomes upward to the overarching outcomes categories.

Project partner concerns centre on who should complete the mapping and when, with the need to avoid premature location of outcomes within specific categories being important to practitioners. This has been accompanied by recognition that what matters to the person can impact upon multiple outcomes categories. While some partners find this problematic, others perceive it positively. Partner discussions highlighted distinctions between 'fixing personal outcomes within a pre-defined category', 'mapping personal outcomes to a specific pre-defined category retrospectively' and 'locating personal outcomes within pre-defined categories through dialogue and negotiation'. The contextual factors shaping these distinctions are the focus of ongoing analysis.

Categorisation of Outcomes: Conclusions

Seven project partners use pre-defined outcomes categories to help make sense of and utilise personal outcomes information and to support identification of personal outcomes, although in different ways. Distinctions can be located on a continuum from:

- The 'top down' use of pre-defined but locally owned and valued categories to develop standardised questions or indicators



- The fluid use of pre-defined categories by practitioners to frame personal outcomes questions
- The retrospective use of pre-defined categories to map personal outcomes
- The ‘bottom up’ use of pre-defined categories to jointly negotiate the location of personal outcomes within the practice encounter

Applications listed towards the top of this continuum afford greater reliability, as understood within the quantitative research tradition, than those toward the bottom and thus generate greater confidence on the part of organisational decision makers who value consistency when making comparisons. However, this comes at the potential expense of privileging the person’s voice in determining personal outcomes and creates a separation between outcomes measurement and outcomes focused support planning. What can be said is that aggregated outcomes, whether recorded initially as personal outcomes within or subsequently mapped to pre-defined categories, can be grouped into a relatively small number of reasonably universal goals, upon which there is a good deal of agreement. Whether or not the value and composition of the aggregated outcome data differs depending upon the collection mechanism is unknown. This merits further investigation.



Briefing 4: Concerns that Relate to Choice of Measurement Scale

The two broad approaches to conceptualising and measuring personal outcomes are also associated with different measurement scales:

- Both standardised measurement tools favour a numeric scale that lends itself to visual representation in the form of a web and greater statistical manipulation.
- The user-defined outcomes categorisation approach favours measurement scales that can be readily embedded within the flow of a conversation.

Standardised Measurement Tools: Numeric Scale Measures

Despite both favouring numeric scale measures, supported by a text box for recording contextual descriptions, there are important differences in the approaches used by Penumbra and Angus Council.

In Angus Council, for each indicator, the child is invited to place themselves on a visual continuum between 1 and 10, where 1 = not at all true of me, and 10 = very true of me. The scale points in between are not defined. The scores for the different dimensions can be joined up to form a Wellbeing Web, which holds appeal for various reasons including being person-centred, non-threatening, affording different degrees of ownership, giving the person something to do, making wellbeing tangible by giving it a physical form, and ease of use and acceptability across a wide age range.

The actual numeric score assigned to any given indicator is not deemed important and it is recognised that the scale points will be subject to varying interpretation. For practitioners, it is the ability to interpret and respond to individual changes in perception, or to compare the perspectives of child, parent and practitioner at a given point in time that is deemed critical, rather than the actual changes recorded by individuals over time, underscoring the importance of supporting contextual information. While documentation states that the outcomes information gathered through the tool can be used to report on overall service outcomes, current practices mean that the scores themselves cannot be readily used beyond the level of the individual family.

In Penumbra, I.ROC comprises a six point Likert-style response scale, where each scale point is defined, with 1 = never, 6 = all of the time. Again the scores for different indicators can be joined to form a web. Penumbra initially invested time and energy in trying to reach agreement about the meaning of individual numeric scores, but found that it doesn't work. Asserting that everyone has or will experience the same six stages in the recovery journey was found to be flawed and unhelpful, as everyone is different, uses different calibration mechanisms, and is on a different pathway. This varies markedly from tools such as the *Outcomes Star*, which is founded upon a ladder of change that assumes a single, common pathway to recovery, with each stage based on fixed, normative expectations about attitudes and behaviours.

What is important however is how the change is perceived or experienced by the individual. For this change to be both meaningful to the person and support comparison, the record has



to provide something to reflect back on and a fixed frame of reference: “*Compared to where I was 3 months ago I am now feeling ‘X’ more or less often*”.

I.ROC scores are being aggregated and used for external and internal reporting purposes, with external reports showing changes in the mean scores (made possible through the choice of interval scale) for each indicator between baseline and latest I.ROC. Internally, analysis has generated important insights for service improvement and planning purposes. Key learning includes a developing understanding of both the size of the service and number of individual I.ROCs that need to be completed for analysis to be statistically meaningful, and the realisation that generating service data more frequently than annually is not sensible for the mental health service user population. The ability to aggregate data has resulted in requests from external agencies for figures detailing the number of individuals making improvements against I.ROC indicators. However, the organisation is resisting, asserting that only changes in mean scores actually make sense statistically and that supporting contextual information is also needed to be able to make deeper sense of summary numeric patterns.

User-Defined Personal Outcomes Measurement: Conversational Scale Measures

A number of different scale measures have been incorporated within the ‘user-defined’ approach to measuring outcomes, with the two most frequently used being ‘met/partially met/unmet’ and ‘improvement’ scales at review, with variants making use of an ‘importance’ scale during the initial assessment.

Project partner findings indicate that despite initially appearing to offer advantages over ‘improvement’ scales, particularly where maintaining quality of life may be a more appropriate goal, there is a growing sense that the ‘met/partially met/unmet’ scale is problematic for a number of conceptual, ethical and practical reasons:

- There is a sense that this scale is a legacy from the days of needs based assessment and the associated concept of ‘unmet need’
- Conceptually the idea of an outcome being ‘fully met’ suggests that outcomes can be achieved once-and-for-all. The reality is often more complex and varied, with fluctuations evident over time
- In a climate of resource constraints, practitioners and people using services can be reluctant to identify an outcome as ‘fully met’ for fear of the service being withdrawn
- Practitioners have identified that recording an outcome as ‘unmet’ can be discouraging for people using services, particularly where effort has been invested in making change
- Together these issues render ‘partially met’ the default option

The combination of ‘importance’ and ‘improvement’ scales has proven less problematic and has been used to good effect, particularly in settings where the choice of scale has been permitted to evolve in response to local feedback.

In VOCAL, numeric scales proved particularly difficult to embed within telephone conversations with carers and were first replaced by improvement scales, with the scale point ‘no improvement’ later replaced with ‘no deterioration’ to offer a more accurate and encouraging description of carers’ realities. The ‘importance’ scale in assessment has enabled



the organisation to determine which categories of outcomes reflect carer priorities, while the ‘improvement’ scale has highlighted outcomes categories that are most or less difficult to progress, together with categories of outcomes where an initial dip in scores is likely within some service settings (due to enhanced carer awareness). This approach supports reporting by outcomes category at service level, with quantitative information included within reports to commissioners and in funding proposals, with supporting contextual descriptions.

Conclusions

Together the various project partner findings highlight the impact of several related issues resulting from an overarching pressure to measure. They are helpful in strengthening the case for making greater use of qualitative outcomes data for a variety of purposes.

- The key measurement issues that project partners are facing relate to:
 - The pressure to measure and its impact on understandings and assumptions about the role of qualitative data (Barrie and Miller 2015)
 - The need to reconcile different conceptualisations of personal outcomes
 - Concept operationalisation and outcomes categorisation
 - Choice of scale measure
- The extent to which it is feasible to ‘measure’ is contested and subject to varying conceptualisations of personal outcomes.
- Current measurement approaches can be located on a continuum based upon different approaches to the categorisation of personal outcomes, notably the point at and means by which this happens.
- There is no perfect scale measure: what is important is that the choice of measurement scale fits with the conceptualisation of personal outcomes, with practice, and is acceptable to people using the service:
 - Numeric scales and importance/improvement scales have both been used to good effect in different contexts
 - There is a growing sense that the ‘met/partially met/unmet’ scale is problematic for conceptual, ethical and practical reasons.
 - A simple improvement scale appears to be less problematic
- It is possible to develop a valid and reliable measure of wellbeing as understood within a specific context e.g. I.ROC
- Each current approach should be used with caveats and not in isolation.



References

- Alkire, S. (2002), Dimensions of Human Development, *World Development*, 30(2), 181-205
- Barrie, K. and Miller, E. (2015) *Supporting the use of qualitative data in service settings*, Meaningful and Measurable
- Bryman, A (2012) *Social Science Research Methods*. Sage: London
- Cook, A. & Miller, E. (2012) *Talking Points Personal Outcomes Approach: A Practical Guide*, JIT: Edinburgh
- Lincoln, Y. S. & Guba, E. G. (1985) *Naturalistic Inquiry*. Beverley Hills, CA: Sage
- Mallinson, S. 2002, "Listening to respondents: a qualitative assessment of the Short-Form 36 Health Status Questionnaire", *Social Science & Medicine*, vol. 54, no. 1, pp. 11-21
- Mayne, J. (1999) Addressing Attribution Through Contribution Analysis: Using Performance Measures Sensibly, Office of the Auditor General of Canada
- Meaningful and Measurable (2012) Exploring the use of personal outcomes information in health and social care: Data retreat report <https://meaningfulandmeasurable.wordpress.com/>
- Miller, E. (2012) *Measuring Personal Outcomes: Issues and Challenges*, IRISS: Glasgow
<http://www.iriss.org.uk/resources/measuring-personal-outcomes-challenges-and-strategies>
- Monger, B., Ion, R., Henderson, N., Cumming, J., and Hardie, S. (20) Outcome Measurement in a Scottish Mental Health Charity, *Mental Health Today*, Mar-April 24-27
- Rapkin B D & Schwartz C E (2004) Toward a theoretical model of quality-of-life appraisal: Implications of findings from studies of response shift, *Health Quality of Life Outcomes* (2)
- Schwartz, C. E & Rapkin, B. D. (2004) Reconsidering the psychometrics of quality of life assessment in light of response shift and appraisal, *Health Quality of Life Outcomes* (2)



