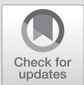

Doing Quantitative Research with Outcome Measures

Charlie Duncan and Barry McInnes

Contents

Introduction – 197

What Are Outcome Measures and why Are they Used? – 197

Two Incentives for Outcome Measurement – 198

To What Extent Do OMs Have a Place in Routine Practice Settings, and Why? – 199

What Is Effect Size? – 199

What Does Research Tell us about Variations in Outcomes across Therapists, Services and Settings? – 200

Can we Use Outcome Measures to Determine our Effectiveness? – 202

Cycle of Excellence – 203

© The Author(s) 2020
S. Bager-Charleson, A. McBeath (eds.),
Enjoying Research in Counselling and Psychotherapy,
https://doi.org/10.1007/978-3-030-55127-8_11

How Can I Choose an OM that Is Suitable for my Purposes and Setting? – 204
On Outcomes and OMs: an Overview – 204

What Are the Key Features and Qualities that a Robust Measure Should Possess, and where Can I Find out More? – 204
Nomothetic Measures – 204
Idiographic Measures – 205

Reliability and Validity – 206

Applicability, Acceptability, Practicality and Ethical Considerations – 206

Practicality – 207

Creating your Own Measure – 208

How Are OMs Used in a Range of Research and Routine Practice Settings? – 208

Naturalistic Study – 209

Get Involved! – 210

References – 211

Learning Goals

After reading this chapter you should be able to:

- Outline the wider research and social policy contexts within which therapy outcome measurement sits;
- Describe the key incentives for implementing outcome measurement and how they shape its implementation;
- Appreciate the phenomenon of outcome variance across service and practitioners;
- Define the positive contribution that conscious application of feedback from sessional measures makes to improved outcomes for clients;
- Apply a range of key principles to the process of choosing an appropriate outcome measure;
- Describe how outcome measures are used in a range of research and routine practice settings.

Introduction

The use of outcome measures (OMs) in therapy, and arguments for and against their use, sits within a wider context of research evidence. Attitudes to the use of outcome measures range along a continuum. At one end of that continuum sit those who believe and argue passionately that measures have no place in the therapy room. At the other end are those who believe, equally passionately, that the use of measures provides valuable additional client feedback that can help us to deliver therapy more effectively and efficiently.

As practitioners, our attitudes may be shaped by a range of factors; research approaches will vary depending on our training and our philosophical stances. They might also sometimes depend on anxieties about having our impact or 'performance' measured.

In this chapter, we will explore a wider psychotherapy research context with an interest in a range of 'evidence' to support you through the process of making an informed choice about outcome measures in terms of questions such as the following:

1. What are outcome measures and why are they used?
2. To what extent may OMs have a place in routine practice settings, and why?
3. How can you choose an OM that is suitable for your purposes and setting?
4. How can you use OMs optimally to achieve your aims?
5. How are OMs used in a range of research and routine practice settings?

What Are Outcome Measures and why Are they Used?

Outcome research is often characterised by the use of outcome *measures*, designed to identify the *changes* that take place during therapy. These contrast with **process** measures, which aim to identify the variables that *cause* these changes. One example

of an outcome measure is the PHQ-9, which is a self-report measure of depressive symptoms. This can be completed at various intervals throughout the therapeutic journey and allows for the 'tracking' of clients' symptoms. In contrast, the Working Alliance Inventory is a process measure which aims to quantify the 'strength' of the therapeutic alliance between practitioner and client. In an ideal world, outcome research would encompass both types of measure to allow us to say not only *what* changes as a result of therapy, but also *how* these changes come about.

One of the main assumptions of outcome and process measures is that the constructs they are attempting to measure (e.g. depression or the therapeutic alliance) are phenomena which can be *measured*. Inherently, this relies on there being a shared understanding of what practitioners, clients and society collectively mean by these concepts. Clinically, this can be challenging when working across disciplines where understandings of the nature and meaning of such concepts can vary (see Marsella 2003 for a more in-depth discussion of cultural differences in depression).

Two Incentives for Outcome Measurement

In our experience there tend to be two main incentives, or drivers, for implementing outcome measurement. Each has a different focus. The first focuses on demonstrating the impact of a service to external stakeholders, for example to funders or a board of governors. The second sees the use of measurement as a form of feedback to inform service and practitioner development. While both are perfectly valid incentives, the practical implementation of each is likely to take a very different form. Consider the two following scenarios.

► Example

Scenario 1.

Service A uses pre- and post-therapy outcome measures to determine the proportion of clients that show improvement in their levels of distress. Paper measures are administered by reception staff prior to their first appointment, and in the last session by the therapist. Therapists are not provided with training in the use and interpretation of measures and they are not routinely reflected on with clients in sessions. Data is collected and collated for quarterly and annual reports to funders. Feedback on individual improvement rates is not given to the therapist team.

Scenario 2.

Service B provides training for its practitioner team in the use of sessional measures of outcome with clients. Clients complete a brief measure at the start of each session and their responses form part of a discussion about the client's progress and their experience of the helpfulness of therapy. Clients also complete a brief measure of the working alliance at the conclusion of each session. The feedback from these measures is used collaboratively between therapist and client to monitor progress and adjust focus as necessary. Data about clients' progress also forms part of clinical supervision and practitioner development.

These two examples illustrate two very different sets of intentions. Service A's use of measurement is primarily to satisfy the requirements of funders and other stakeholders. Service B's approach is predicated (based on research evidence) on the assumption that the service and its practitioners can use the feedback generated from measures to reduce the likelihood of premature termination, enhance outcomes and create the best experience of therapy for every client. We will return to these themes later in this chapter. ◄

Activity
With which of these two drivers for outcome evaluations do you feel the greatest affinity? What, in your experience to date, has informed this view? To what extent, other than demonstrating the effectiveness of therapy to lay people, do you feel measures have a valid place in the therapy room? How familiar are you with the body of research showing that when used collaboratively with clients, feedback from measures can improve the outcomes of therapy?

To What Extent Do OMs Have a Place in Routine Practice Settings, and Why?

What is, first, known about the overall efficacy of therapy? Summarising the findings from a range of meta-analyses of the efficacy of psychotherapy, Wampold and Imel (2015) conclude that a reasonable estimate of the effect size of therapy would be *d*=0.8 (ref adjacent panel).

What Is Effect Size?

Effect size is an expression of the strength of the relationship between two variables. For example, we want to know the effect of using a particular therapy (**variable A**) for treating anxiety (**variable B).** The effect size value will show whether that therapy had a small, medium or large effect (or indeed no effect). Cohen's d is commonly used to express the strength (or size) of that effect. Cohen suggested that d=0.2 be considered a 'small' effect size, 0.5 a 'medium' effect size and 0.8 a 'large' effect size.

If we were to compare the effects of treatment with therapy to no treatment, a small effect size of d=0.2 would mean clients receiving therapy would be better off, in outcome terms, than 58% of people who did not receive therapy. A large effect size (d=0.8 or above) would mean clients receiving therapy would be better off than 79% of people not receiving therapy. In social sciences research, this is a large effect size. From the various meta-analyses conducted over the years, the aggregate effect size related to absolute efficacy is remarkably consistent and appears to fall within the range 0.75 to 0.85.

What Does Research Tell us about Variations in Outcomes across Therapists, Services and Settings?

The fact that therapy 'works', as suggested by Wampold and Imel (2015), and that there is broad equivalence among different models, does not mean that all services, or the practitioners within them, are equally effective. Research which considers the therapist as a variable which may impact the outcomes of therapy has found that therapist effects, as they are known, make a vastly greater contribution to therapy outcome than therapy models and techniques. There are different ways of seeking to understand the factors that lie behind these variations and their impact on outcomes. Chow et al. (2015) write, for instance, that 'Evidence has consistently shown that therapist effects dwarf the contribution made by the perennially popular treatment models and techniques, accounting for 5–9 times more variance in outcome'.

Given that variations in outcome exist between practitioners, and between services, where do we imagine we fit on this range of effectiveness? Equally importantly, what informs our view? A study of mental health professionals that included psychiatrists, psychologists and psychotherapists, published in 2012 (Lambert, 2013), refers to 'self-assessment bias' in noting that we are highly likely to rate our level of skill and performance as above average for our profession, and also to overestimate the actual impact of our work with clients. You'll be able to test your own level of self-assessment bias in the reflective questions that follow.

It may help both ourselves and our clients, then, if we can find some 'objective' measure of the true impact of our work with clients, and using percentage can help to discuss this in 'measurable' terms.

Activity

These are the same questions that were put to the subjects in the study of self-assessment bias highlighted above. Answers from those respondents are provided later in this chapter against which you may compare your own responses.

Compared to other mental health professionals within your field (with similar qualifications), how would you rate your overall clinical skills and performance in terms of a percentile (out of 0–100%: e.g. 25% = below average, 50% = average, 75% = above average)?

1. What is percentile?

A percentile is a number, between 1 and 100, where a certain percentage of scores fall below that number. Imagine that you are the fourth tallest person in a group of 20. This means that 80% of people in the group are shorter than you; you are in the 80th percentile. If you imagine that you imagine that you are more proficient in a particular skill than 75% of similarly qualified peers, that would put you in the 75th percentile.

2. What percentage (0–100%) of your clients get better (i.e. experience significant symptom reduction during treatment)? What percentage stay the same? What percentage get worse?

From the study above, in answer to question 1, respondents rated themselves on average in the 80th percentile–in other words, more highly than 79% of their peers. Just 8.4% rated themselves below the 75th percentile. None rated themselves below the 50th percentile, that is, below average.

In response to question 2 respondents believed, on average, that 77% of their clients improved significantly as a result of therapy. Fifty-eight percent believed that 80% or more of their clients improved, and just over one in five (21%) that 90% or more of their clients showed improvement. Almost half of practitioners (47.7%) believed that none of their clients deteriorated. In essence, they believed that their outcomes were far in excess of the rates of improvement shown by the evidence from both controlled and naturalistic settings.

Evidence of outcome *variability* has been demonstrated across several studies. Okiishi et al. (2003) compared the outcome data of 56 therapists in a university counselling service in the US. They found that those whose clients showed the fastest rate of improvement had a rate of change 10 times greater than the average among their colleagues in the same service. The clients of the therapists evidencing the slowest rates of improvement, on average, deteriorated. A more recent UK-based study (Firth, Saxon, Stiles & Barkham, 2019) included data for nearly 27,000 clients, seen by 462 therapists in 30 services. There was a wide range of recovery rates across therapists and services, with an 'average' recovery rate of 58% for therapists and 55.7% for services. However, significant variations in average recovery rates existed, ranging from 48.5% to 69.7% between services and 41.4% to 77.2% between therapists.

Our outcomes also don't appear to improve with experience. A longitudinal study which examined the outcomes of 6591 patients seen by 170 therapists (Goldberg et al. 2016) found that, on average–with some exceptions–therapists tended to obtain slightly poorer outcomes as their experience increased. It also appears that we may be poor at predicting which clients will reach a positive conclusion to therapy and which will not. Hannan et al. (2005) used session-by-session tracking of progress for over 11,000 clients and devised a test to make early predictions about which clients might be at risk of 'treatment failure'. They then compared its reliability with the prediction (based on clinical judgement) by the centre's therapists. Of 550 clients that attended at least one session, three were predicted by the therapists to deteriorate. Outcome data, however, showed that 40 clients had deteriorated by the end of therapy, though only one of these scenarios had been predicted by the therapists. The test tended to over-predict treatment failure, but overall it was far more accurate than therapists' predictions.

Can we Use Outcome Measures to Determine our Effectiveness?

Two key factors that have been shown to be strong and early indicators of a successful outcome are signs of improvement early in therapy and the client's rating of the therapeutic alliance. Numerous studies have shown that, in general, progress in therapy follows a relatively predictable trajectory, with most improvement occurring in the early stages. A study by Howard et al. (1986) found, for example, that up to 40% of clients show significant improvement in the first three sessions, 65% within 7 sessions, 75% within 6 months and 85% within 12 months. They also found that clients who don't display this pattern of early improvement are significantly less likely to improve later on. In another study, Brown et al. (1999) found that clients who showed no improvement by the third session did not, on average, improve over the entire course of therapy. Furthermore, those that showed deterioration by the third session were twice as likely to drop out as they were to progress. From these and other studies, we can conclude that if improvement is going to happen, there are likely to be early signs of it, and that early deterioration or lack of early progress is a potential predictor of drop-out.

Revisiting our earlier points about clarifying the basic 'why' of using measures of outcome, two questions emerge. The first of these is 'Can we use outcome measures to determine our effectiveness?' Whether we're using measures at the first or last sessions of therapy, or in the case of sessional use of measures the first and most recent, we are able to measure the degree of difference between the two.

The second question is 'Can we use outcome measures to track the progress of clients in a way that helps us to identify early those clients who are not "on track" and are therefore at risk of a poor outcome, including premature drop-out?' If we can identify these clients can we then intervene in such a way as to improve their chances of a beneficial outcome? A considerable and growing body of research evidence suggests that the answer to this question is also 'yes'. Lambert et al. (2005) studied the effects of four feedback conditions on clients at risk of treatment failure. The active feedback conditions improved the proportion of clients who clinically and reliably improved; no feedback or treatment as usual (21%); feedback to therapists about 'not on track' clients (35%); feedback to therapists with additional clinical support tools, for example measures of the working alliance (49%); and feedback to both therapist and client about the client's not on track status (56%).

They concluded that "It seems likely that therapists become more attentive to a patient when they receive a signal that the patient is not progressing. Evidence across studies suggests that therapists tend to keep 'not on track' cases in treatment for more sessions when they receive feedback, further reinforcing the notion that feedback increases interest and investment in a patient" (p.168).

Whipple et al. (2003) found that clients at risk of a negative outcome were less likely to deteriorate, more likely to stay in treatment longer and twice as likely to achieve clinically significant change when their therapists had access to information on outcome and alliance. Another study (Miller, Duncan, Brown, Sorrell & Chalk, 2006) examined the impact of introducing short measures of outcome and working alliance into an international employee assistance programme. In the early

phase of the study, 20% of clients at intake had outcome measure but not alliance data. These clients were three times less likely to return for a second session and had significantly poorer outcomes. Improving a poorly rated alliance early in therapy was correlated with significantly better outcomes by the end of therapy.

Using outcome measures in a way that supports practitioners to improve the outcomes of clients at risk of a poor outcome requires something of a conceptual shift. It involves moving from using measures simply to determine outcome, to seeing them as a further way in which we can elicit feedback about client progress and build that feedback into our shared discussion. It needs to be part of a conscious and deliberate process.

Cycle of Excellence

Miller and colleagues (Miller, Hubble & Duncan, undated) propose a framework for the development of professional competence they call the "cycle of excellence". This comprises three principal components:

1. Determining a baseline level of effectiveness;
2. Obtaining systematic, ongoing, formal feedback and
3. Engaging in deliberate practice.

They argue that the establishment of our individual levels of effectiveness is a first basic step in identifying our learning and development needs. We will argue later in this chapter that the use of routine measures of outcome is a cornerstone in the process of gaining some objective measure of just how effective we are.

Activity

Anecdotally at least, much of the resistance towards using outcome measures is a result of practitioners believing that their clients won't like them or won't benefit from them–but is this really true? To our knowledge, there's been little research undertaken in this area from the client perspective. However, a public perceptions survey that was commissioned by the British Association for Counselling and Psychotherapy (BACP) in 2019–which surveyed over 5000 UK adults–found that just over half that clients who had had counselling or psychotherapy had completed outcome measures, and of these 80% said that they were happy to do so. Not only this but two-thirds felt that outcome measures helped both them and their therapist to track their progress and only 21% felt that they got in the way of the therapy. Another recent study which used a much smaller sample (Börjesson and Boström 2019) found that it's particularly important to make sure that clients are aware of the purpose and use of their outcome data and that it's used as part of therapy to increase awareness of inner states. Hence, whilst this shows that outcome measures might not be well-received by *all* clients, it appears that they're not quite so averse to them as people think.

How Can I Choose an OM that Is Suitable for my Purposes and Setting?

On Outcomes and OMs: an Overview

We have mentioned outcomes research and people often use the term. But what does this really mean? Jefford, Stockler and Tattersall (2003) describes it as

> a broad umbrella term without a consistent definition. However, it tends to describe research that is concerned with the effectiveness of public-health interventions and health services; that is, the outcomes of these services. Attention is frequently focused on the affected individual – with measures such as quality of life and preferences – but outcomes research may also refer to effectiveness of healthcare delivery, with measures such as cost-effectiveness, health status and disease burden (p. 110).

Whilst this is a somewhat medicalised definition, essentially outcome research is asking: *What changes for a client or service as a result of therapy?*

This might be individual changes in terms of psychological distress, self-esteem, depressive symptoms and so on, or it might be changes in a service, for example 'How has the number of clients ending therapy prematurely changed as a result of this alteration I've made to my practice?'

> **Activity**
> Stop and consider the term 'outcome research'. What does it mean to you? What thoughts, feelings and emotions does it stir up in you? Just sit with that for a moment and think about why you feel like this. What, from your experience, has led to you feeling like this?

What Are the Key Features and Qualities that a Robust Measure Should Possess, and where Can I Find out More?

Choosing an outcome measure can, in turn, be a minefield as there are just *so* many different measures available. GAD-7, PHQ-9, CORE, WEMWEBS, GHQ, IES, HADS, BDI, SRS, OQ-45, Goal Based Outcomes (GBOs)—there's a measure for every condition and every setting. So how can you choose one that's right for you, your client and your service? Broadly speaking, there are two main types of outcome measure: nomothetic and idiographic.

Nomothetic Measures

The term 'nomothetic' has been referred to in earlier chapters. Nomothetic measures are, as mentioned, designed to establish general principles or assumptions by asking large groups of people a set of pre-determined questions and then making

generalisations about them based on their answers. They are often quantitative in nature, that is, relating to 'numbers' or 'amounts' that can be measured. An example of a nomothetic measure is the PHQ-9 which asks questions like 'over the last two weeks, how often have you been bothered by any of the following problems: little interest or pleasure in doing things' and clients can choose 'not at all', 'several days', 'more than half the days' or 'nearly every day'. As you can see, there's no option for clients to change any of the items or response options, which can make these measures restrictive if you want to incorporate the client voice more. However, one of the benefits of using nomothetic measures is that their results can be compared with other services who are using the same measure as a 'benchmark'.

Idiographic Measures

Idiographic measures, on the other hand, are, as also described in earlier chapters, more able to focus on individual feelings and experiences, by collecting some qualitative data (typically text or words) about the individual. In the field of outcome measures, an example of an idiographic measure is the Goal Based Outcomes (GBOs) tool (Law 2018). This asks clients to state a goal for therapy in their own words–so no predefined question–and then rate their progress on that goal from 0 (not met at all) to 10 (fully met). Nomothetic measures can appeal to therapists because they allow clients the opportunity to set their own definition of what an 'effective' or 'desirable' outcome might be, rather than having it set for them. On the other hand, these types of measures can be criticised for not being generalisable across all clients because of their individualised nature.

Activity

Return to your ideas about outcome research, asking yourself, or someone else:

- What type of measure better fits my beliefs? Am I more interested in being able to provide a general overview of all my clients or do I want to tailor my therapy (and therefore what I measure) to my clients?
- What type of measures do my clients prefer? Would they struggle to come up with a goal because they don't know what they want from therapy yet? Do they want more direction from me as a therapist?
- Does my service need me to collect a particular measure for the funder or commissioner?
- Do I want to be able to benchmark client outcomes from my practice with a similar service so that I can make comparisons?
- Do I want to collect more than one measure with clients and use a mix of nomothetic and idiographic measures so that I'm able to capture the client voice but also make generalisations?

> Ultimately, the decision around which type of outcome measure you use should be based on what works for you, your clients and your service. You even have the option to create your own bespoke measure if you don't think there's one out there which meets your or your clients' needs. If you're interested in creating your own measure, you might find the paper by Boynton and Greenhalgh (2004) helpful. However, the next section on reliability and validity may also help you decide what's right for you.

Reliability and Validity

The terms 'reliability' and 'validity' are often conflated, but there is a slight difference between them. We have looked at this in earlier chapters too, for instance in ► Chap. 2 in the context of qualitative research. Reliability refers to the consistency of a measure–its ability to return similar results from the same respondent (when used in the same circumstances) each time it is completed, while validity refers to the ability of a measure. A simple example of quantitative research given by Heale and Twycross (2015) is of an alarm clock that should ring at 7 am each morning but is set for 6:30 am. It is reliable in that it consistently rings at the same time every morning, but it isn't valid because it's not ringing at the time you want it to.

Understandably, there's a great deal to be said for choosing a reliable and valid outcome measure for your practice, not least because you can be fairly confident that there's some robust evidence underpinning its use. From the perspective of quantitative research, it is easier to determine the reliability and validity of *nomothetic* measures than for *idiographic* measures because of the former's focus on generalisations and standardisations. As explored elsewhere in this book, idiographic studies have other criteria for their reliability and validity.

Applicability, Acceptability, Practicality and Ethical Considerations

Whatever measures you choose to use, it is important that they are applicable, acceptable to those using them, practical and used ethically. You can have all the reliable and valid measures you want but if they don't meet these objectives then they probably aren't going to be appropriate for your work. Starting with **applicability**: the measure needs to be appropriate and relevant to the client group and setting where you work. If you see a variety of clients with a range of presenting issues then you might be more inclined to collect a global measure of distress such as the CORE-OM or CORE-10, which measure psychological distress more broadly, rather than the PHQ-9, which measures depression specifically. Or, you might pick from a selection of tools depending on the issues your client brings to therapy and choose to take a more tailored approach to outcome monitoring.

In addition, is it acceptable to both you and your clients, and do you both get something out of using it? If the answer to either of these questions is no, then the measure might not be acceptable. Best practice, and common sense, would tell us that both practitioners and clients should be clear on why a measure is being collected and what it will be used for. It's your responsibility as a practitioner to be clear on this yourself and to explain it to your client. If one or both of you don't know why you're collecting it then how ethical is it to be asking them to complete it?

Practicality

Sometimes practicality can trump other factors. If you are looking for a session-by-session measure, it will not be practical to use a measure which has 200 items and takes half an hour to score–there will be no time left for any therapeutic work. Another thing to consider in terms of practicality is whether there are any copyright factors you need to be aware of. Not all measures are free to use, and some can only be used in a certain format at the discretion of the author–so make sure you check! A good place to start is the Child Outcomes Research Consortium website ► https://www.corc.uk.net/outcome-experience-measures/ as they list this information for numerous measures, including those that are appropriate for adult and younger clients.

Finally, there are always going to be ethical issues to consider in your work with clients and using outcome measures is no different. This might include obtaining informed consent from your clients to use the measures in the first instance, which you might choose to include as part of your contracting. There's also the issue of data storage. Secure storage of the data, such as a locked filing cabinet or a secure online system, is paramount. Can you realistically collect it without compromising your data protection responsibilities? For more guidance on practice and research ethics, see BACP's Ethical Framework for the Counselling Professions (BACP 2018a) and its Ethical Guidelines for Research in the Counselling Professions (BACP 2018b).

It's important to consider these issues prior to the collection of measures, so we've put together the following checklist to help you:

- *Does the measure have good evidence of reliability and validity, consistent with the nomothetic research approach?*
- *Do you and your client understand why you are collecting this measure?*
- *Do you, as a practitioner, get something out of using this measure with your clients?*
- *Do your clients get something out of using this measure?*
- *Is it feasible to collect this measure at the timepoints that you have determined?*
- *Do you know how to interpret the measure which you have chosen to use?*
- *Are you collecting and using the data ethically?*
- *Do you have somewhere safe and secure to store the data you're collecting?*

Our suggested further reading materials may also help you with some of the practicalities of using outcome measures in your practice.

Creating your Own Measure

If you are interested in creating your own measure, you might find the paper by Boynton and Greenhalgh (2004) helpful. For any measure that you're considering we would recommend that you first test it on yourself, and other colleagues if possible, and are familiar with the measure's construction, scoring and clinical cut-offs. There is a rationale behind the construction of every properly validated measure. The CORE-OM, for example, is a 34-item measure that spans four key domains (Wellbeing, Problems/symptoms, Functioning and Risk) and contains high- and low-intensity items that relate to problems such as anxiety, depression, trauma, and aspects of life and social functioning. It is important that your choice of measure is based on a clear rationale and that it suits your purposes.

How Are OMs Used in a Range of Research and Routine Practice Settings?

Outcome measures can be used across many different research methods, from randomised controlled trials to case studies. Here, we'll provide some examples of how outcome measures have been used in some real-life research projects, with feedback from some of the practitioners involved.

► **Example**

Randomised controlled trial (RCT)

An RCT is a study where people are randomly assigned to two or more conditions to test a specific intervention or treatment, without any similarities or differences between the people in the groups being taken into account. This is often described as the 'gold standard' for research.

RCT will be explored in more depth in ► Chap. 13, by Megan Stafford. But let's look at an example from Stafford's research with Judith, a school counsellor who has taken part in a real-life RCT (Stafford et al. 2018):

> Judith says: 'As a school counsellor, I was excited to take part in an RCT both to participate in gathering evidence and to extend my own experience. I quickly realised that being part of a research study - of course - involves measurement; far more measurement than I was accustomed to. As a counsellor in the study, I used the Outcome Rating Scale (ORS) with clients in each session, whilst also being measured myself (for adherence to the research protocol). As an assessor I met with young people who were interested in participating and administered a battery of measures to screen them for the study.
>
> I was apprehensive; would using the measures feel clunky or like minimising or marginalising client's experience? Sometimes this felt true, but often I found the opposite. In counselling, the ORS helped focus our joint attention on what was going on inside and outside sessions and often empowered clients to be able to quickly communicate more of this. As an assessor I only met the young per-

son once but even in these paperwork intense meetings, it was possible to have a human and helpful interaction. The richness of the resulting information and the ease with which most of the young people communicated it via the measures surprised me. [One disclosed serious risk that he hadn't been able to voice before, and I was able to help him get the immediate support he needed.]

Since the study, I have incorporated measures into all my work. Now that I have become practised and familiar with using them collaboratively, I see them as an additional resource; more to do with input than outcome, another way of hearing clients, and often helpful for young people'. ◄

Naturalistic Study

A naturalistic study is one where the researcher observes or records a behaviour or phenomenon in its natural setting, whilst interfering as little as possible.

► Example

In counselling and psychotherapy research this might be similar to a service evaluation where the intervention and measures being collected don't change, but the researcher analyses the data collected to say something about the clients using the service.

Let's take a look at this example of a naturalistic study:

» Alicia is a counsellor working in a community counselling service for children and young people up to the age of 25. At every session, she asks her clients to complete either the YP-CORE or the CORE-10, depending on their age, and she also collects the Strengths and Difficulties Questionnaire (SDQ) at the first and last session with those aged 16 and under. Alicia uses the measures as a talking point during each session but does not score them and passes them on to her service manager. This is also how other counsellors in the service work.

Over the last few years, she's noticed that the clients coming to see her are increasingly distressed and many are on the waiting list for, or have been rejected from, a Child and Adolescent Mental Health Service (CAMHS). When Alicia raises this with her manager, her manager says that she has also become aware of this and has been having conversations about this with the commissioners in their local area. However, the commissioners believe that the interventions being provided in the community setting are for 'less distressed' clients and ask them what evidence they have that what they are providing 'works'.

Alicia and her manager decide that with the YP-CORE, CORE-10 and SDQ data that they collect as a service, they may be able to provide some evidence to back up what they're saying. When they analyse the data, they notice that 80% of the clients coming into the service are moderately to severely distressed, similar to those accessing CAMHS. They also find that 60% of the clients coming to their service 'recover', which again is similar to the recovery rate in CAMHS. They take this evidence back to the commissioners, who agree that they're providing a vital service which can operate alongside CAMHS. They agree to provide the

service with some funding, allowing them to employ two additional full-time counsellors each week.

This is a very basic example and it might not be as easy as this in 'real life', but it's one way in which data can be used to evidence what it is that you're already seeing in your service and how that evidence might then be able to make a case for increased funding. ◄

Get Involved!

The British Association for Counselling and Psychotherapy (BACP) is a registered charity and membership organisation for counsellors and psychotherapists. They support practitioners and services to collect routine outcome data and can provide guidance and support in data analysis and interpretation. If this is something that you, or your service, would be interested in, please email research@bacp.co.uk.

If you would like to develop your knowledge about outcome research, we are hoping that you will find the following list of links helpful:

- How to choose a therapy outcome measure: ► http://therapymeetsnumbers.com/how-to-choose-a-therapy-outcome-measure/
- Introducing measures into working with clients: ► http://therapymeetsnumbers.com/introducing-measures-into-working-with-clients/
- How do I use the feedback from measures to reflect on work with clients? ► http://therapymeetsnumbers.com/every-picture-tells-a-story/
- Using sessional measures to deliver effective and efficient therapy–an example:► http://therapymeetsnumbers.com/deliver-effective-therapy-efficiently-at-reduced-cost/
- For an accessible and in-depth exploration into the development of methods to use for evaluating our own practice, please also see Biljana Van Rijn (2020).

Summary

Measurement of outcome in therapy settings, while not new, has until recently been an activity restricted mainly to research and selected practice settings. More recently, demands for evidence of effective use of public funds, and the accumulation of very large datasets in settings such as the UK's Improving Access to Psychological Therapies Programme, have moved the issue of routine measurement of outcomes centre stage. This chapter explores some of those contextual factors and the key drivers shaping this movement. We looked at the underlying philosophies behind two key drivers and how they differentially shape the way in which outcome measurement may be implemented. Moving on, we explored the body of research which demonstrates that while different therapeutic approaches are broadly similar in their outcomes, at a service and practitioner level there is considerable variance. Finally, we provide examples of outcome measurement from research and practice settings, and guidance for practitioners in the selection and implementation of measures appropriate to their practice.

References

BACP. (2018a). *Ethical Framework for the Counselling Professions*. Available at: https://www.bacp.co.uk/events-and-resources/ethics-and-standards/ethical-framework-for-the-counselling-professions/.

BACP. (2018b). *Ethical Guidelines for Research in the Counselling Professions*. Available at: https://www.bacp.co.uk/events-and-resources/research/publications/ethical-guidelines-for-research-in-the-counselling-professions/.

Börjesson, S., & Boström, P. K. (2019). "I want to know what it is used for": Clients' perspectives on completing a routine outcome measure (ROM) while undergoing psychotherapy. *Psychotherapy Research*. https://doi.org/10.1080/10503307.2019.1630780.

Boynton, P. M., & Greenhalgh, T. (2004). Selecting, designing, and developing your questionnaire. *BMJ, 328*(7451), 1312–1315. https://doi.org/10.1136/bmj.328.7451.1312.

Brown, J., Dreis, S., & Nace, D. K. (1999). What really makes a difference in psychotherapy outcome? Why does managed care want to know? In M. A. Hubble, B. L. Duncan, & S. D. Miller (Eds.), *The heart and soul of change: What works in therapy* (pp. 389–406). Washington, DC, US: American Psychological Association.

Chow, D., Miller, S. D., Seidel, J. A., Kane, R. T., Thornton, J., & Andrews, W. P. (2015). The role of deliberate practice in the development of highly effective psychotherapists. *Psychotherapy*, *52*(3), 337–345. http://dx.doi.org/10.1037/pst0000015

Firth, N., Saxon, D., Stiles, W. B., & Barkham, M. (2019) Therapist and clinic effects in psychotherapy: a three-level model of outcome variability. *Journal of Consulting and Clinical Psychology*, 87(4), 345–356.

Goldberg, S. B., Rousmaniere, T., Miller, S. D., Whipple, J., Nielsen, S. L., et al. (2016). Do psychotherapists improve with time and experience? A longitudinal analysis of outcomes in a clinical setting. *Journal of Counseling Psychology, 63*(1), 1–11.

Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., et al. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology, 61*(2), 155–163.

Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence-Based Nursing, 18*, 66–67.

Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose-effect relationship in psychotherapy. *American Psychologist, 41*(2), 159–164.

Jefford, M., Stockler, M.R., Tattersa, M. (2003) Outcomes research: what is it and why does it matter? *Internal Medicine Journal*. 33(3):110–8. doi: 10.1046/j.1445-5994.2003.00302.x.

Lambert, M. J., Harmon, C., Slade, K., Whipple, J. L., & Hawkins, E. J. (2005). Providing feedback to psychotherapists on their patients' progress: Clinical results and practice suggestions. *Journal of Clinical Psychology, 61*(2), 165–174.

Lambert, M. J. (2013). Outcome in psychotherapy: The past and important advances. *Psychotherapy, 50*(1), 42–51. https://doi.org/10.1037/a0030682

Law, D. (2018). Goals and goal-based outcomes (GBOs): Goal progress chart. Available at: https://goalsintherapycom.files.wordpress.com/2018/03/gbo-version-2-march-2018-final.pdf

Marsella, A. J. (2003). Cultural Aspects of Depressive Experience and Disorders. *Online Readings in Psychology and Culture, 10*(2). https://doi.org/10.9707/2307-0919.1081.

Miller, S. D., Duncan, B. L., Sorrell, R., Brown, G. S., & Chalk, M. B. (2006). Using outcome to inform therapy practice. *Journal of Brief Therapy*, 5(1), 5–22.

Okiishi, J., Lambert, M., Nielsen, S. L., & Ogles, B. M. (2003). Waiting for supershrink: An empirical analysis of therapist effects. *Clinical Psychology and Psychotherapy, 10*(6), 361–373.

Stafford, M. R., Cooper, M., Barkham, M., Beecham, J., Bower, P., Cromarty, K., et al. (2018). Effectiveness and cost-effectiveness of humanistic counselling in schools for young people with emotional distress (ETHOS): study protocol for a randomised controlled trial. *Trials, 19*, 175. https://doi.org/10.1186/s13063-018-2538-2.

Van Rijn, B. (2020). Evaluating Our Practice. In S. Bager-Charleson (Ed.), *Reflective Practice and Personal Development in the field of Therapy*. London: Sage.

Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work (Second Edition)*. New York, New York: Routledge. https://doi.org/10.4324/9780203582015.

Whipple, J. L., Lambert, M. J., Vermeersch, D. A., Smart, D. W., Nielsen, S. L., & Hawkins, E. J. (2003). Improving the effects of psychotherapy: The use of early identification of treatment and problem-solving strategies in routine practice. *Journal of Counseling Psychology, 50*(1), 59–68.