Appendix B - Published manuscripts

A guide to outcome measurement in psychiatry

ARTICLE

Howard Ryland, Jane Carlile & David Kingdon

SUMMARY

Measuring outcomes is becoming an increasingly standard (and highly complex) part of what mental health services are expected to do. Practising psychiatrists will need to have a good understanding of approaches to outcome measurement: used well, they have the potential to amplify the patient voice, promote good-quality services and facilitate research. We discuss what constitutes an outcome measure, the different ways that such measures can be obtained and the mechanisms for assessing the quality and appropriateness of an outcome measure. We outline the rapidly evolving research and policy context regarding outcome measurement, with particular reference to the UK's National Health Service. We also consider the potential pitfalls to outcome measurement, such as added clinical burden, inappropriate incentivisation of behaviour and incorrect interpretation of results. We discuss ways that such difficulties can be avoided or their effects mitigated.

LEARNING OBJECTIVES

After reading this article you will be able to:

- understand how to evaluate and interpret outcome measures in psychiatry
- be aware of the role of organisations in developing and selecting outcomes, including the Royal College of Psychiatrists
- consider the opportunities and challenges of implementing outcome measurement in clinical practice

KEYWORDS

Outcome measures; clinical governance; quality improvement; information technologies; psychometrics.

This article builds on a previous article in this journal (Lewis 2014), which covered many relevant areas of outcome measurement, including:

- the difference between outcome and process measures
- the categories of outcome that can be measured, such as the changes in symptoms and quality of life
- different types of measure, such as patient-reported outcome measures (PROMs), patient-reported experience measures (PREMs) and clinician-reported outcome measures (CROMs)
- psychometric properties, such as validity and reliability.

We expand on these elements and provide an updated view of the current state of the art in outcome measurement.

Current policy context

National Health Service (NHS) policy in this area moves swiftly and the landscape is continuously changing. Outcome measurement has become a hot topic in mental health services, with increasing emphasis on demonstrating that services are achieving good results. Under the auspices of the *Five Year Forward View for Mental Health*, NHS England and NHS Improvement published guidelines highlighting the need for a framework approach (NHS England and NHS Improvement 2016a). This document articulates the benefits that such measurement can offer both to individual patients and to the system as a whole.

A key aim for many policy makers has been to produce a system for costing mental health services that links quality and outcome to funding. The payment-by-results (PbR) model for mental health systems attempted unsuccessfully to achieve this, but there are continuing projects seeking ways to do so (Kingdon 2019). NHS RightCare (NHS England 2019a) highlights using information from across the whole patient pathway to help identify maximal improvements in both spend and outcomes.

Many commissioners have responded by incorporating the reporting of outcome measurement within key performance indicators (KPIs). This can be done at the local level, as has been implemented in Oxford Health NHS Foundation Trust, where a proportion of funding is attached to a series of outcomes (NHS England and NHS Improvement 2016b). Alternatively, it can be done at a national level, such as the dashboards of NHS England's specialist commissioning, whereby providers are asked to supply data on key outcome measures (NHS England 2020).

The Royal College of Psychiatrists recommends the use of routine outcome measures in psychiatric practice and will shortly publish a mental health

Howard Ryland, MA (Cantab), MBBS, MRCPsych, MSc, is a National Institute for Health Research (NIHR) Doctoral Research Fellow in the Department of Psychiatry at the University of Oxford, UK, where his research focuses on outcome measurement in forensic mental health services. He works as an honorary consultant forensic psychiatrist in the prison service. He is also the lead for the research working group of the Clinical Reference Group for Adult Secure Services at NHS England and the deputy editor of CPD Online and Trainees Online at the Royal College of Psychiatrists, London. His previous roles have included NICE scholar, president of the European Federation of Psychiatric Trainees and clinical fellow at the Academy of Medical Royal Colleges. **Jane Carlile**, MBBS, MRCPsych, MBA, is a consultant psychiatrist and former Group Medical Director for Cumbria, Northumberland, Tyne & Wear NHS Foundation Trust, UK. She is the Lead for Outcomes in Mental Health and Specialist Advisor for Mental Health Payment Systems for the Royal College of Psychiatrists, London. She has contributed to the CTO working group for the Mental Health Act Independent Review and has published in this field. **David Kingdon**, MD, FRCPsych, is Emeritus Professor of Mental Health Care Delivery at the University of Southampton, UK, and has worked as consultant adult psychiatrist, and medical and clinical director with Nottingham Healthcare and Southern Health NHS Foundation Trusts and been involved in policy developments over the past 40 years.
Correspondence Dr Howard Ryland. Email: howard.ryland@nhs.net

First received 28 May 2020
Accepted 15 Jul 2020

Copyright and usage
© The Authors 2020

Downloaded from https://www.cambridge.org/core. 14 Jun 2021 at 15:28:34, subject to the Cambridge Core terms of use.

outcomes paper, with detailed contributions from each of its faculties. This seeks to provide consensus on the way forward for measuring outcomes across the full spectrum of psychiatric services. This will be available in due course from the Royal College of Psychiatrists' website.

Ways to measure outcomes

A range of definitions exist to describe what outcomes are in the context of healthcare. One of the simplest and most apposite is: 'Outcomes are the results that matter most to patients' (International Consortium for Health Outcomes Measurement 2020a). This rightly puts the emphasis on the patient, rather than the clinician, as the primary arbiter of what constitutes a good outcome.

This patient-centred approach has the advantage of harnessing individuals' subjective experience. Many other stakeholders may also have a relevant perspective on which outcomes are important. For example, family members and clinicians are frequently involved. Repeated measures can provide a longitudinal understanding of a patient's trajectory over time.

Stakeholders' views can be quantified through the use of reporting tools such as patient-reported outcome measures (PROMs) or clinician-reported outcome measures (CROMs). These usually take the form of a questionnaire, which can contain a mixture of question styles. Potential limitations of such an approach include the interpretability of responses, 'gaming' of answers and the reliability of scores between respondents or across time.

Another approach is to measure more objective 'hard' outcomes, such as hospital admission, relapse or death. These measures have been extensively used for research purposes, as they are often routinely collected in standardised databases. These have the advantage of being easy to measure and highly reliable but can be of limited practical use because they occur infrequently and fail to capture the nuances of the patient's own experience. Furthermore, they may still be subject to interpretation, for example what constitutes a relapse or requires hospital admission may vary between local services.

Performance measures focus on an individual's performance on a prescribed task, with calibrated measurement of their resultant efforts. Such tests are commonly utilised to determine physical functioning, but can also be used to measure psychological concepts such as cognitive functioning.

Biomarkers can also be used as outcome measures, such as blood tests or radiographic images. These are much less frequently employed in psychiatric illnesses, as they often give little information about results that matter to patients.

Other types of measurement in psychiatry

There can be confusion about the purpose of measurement in mental health services. This can lead to the use of measurement instruments for different purposes than those for which they were designed, which in turn can result in them being used in ways that are not supported by the evidence.

Common uses of measurement in psychiatry include for diagnosis and prediction of likely future events, such as disease prognosis, risk or side-effects. In routine practice, diagnosis is usually made on the basis of clinical judgement, supported by categorical or dimensional nosologies such DSM-5 or ICD-10. In research and in some specialist clinical areas, the use of diagnostic instruments is more common, such as the Structured Clinical Interview for DSM-5 – Research Version (SCID-5-RV) (First 2015) or the Diagnostic Interview for Social and Communication Disorders (DISCO) (Leekam 2013). Such structured clinical judgement tools offer more robust and defensible diagnosis, while still incorporating the perspective of clinicians. Whether a person continues to meet criteria for a diagnosis after intervention can be used as an outcome measure in research studies. However, such diagnostic instruments often lack responsiveness to change and can therefore be unsuitable for use as dynamic outcome measures.

Prediction of future events using identified risk factors was pioneered in medicine with the use of risk tools such as the Framingham Risk Score for future cardiovascular events (Wilson 1998). The Framingham Risk Score is a gender-specific algorithm used to estimate the 10-year cardiovascular risk of an individual. It is based on data from the Framingham Heart Study, started in 1948 (Dawber 1951). Similar approaches are now being developed in psychiatry to look at a range of risks, such as suicide, recidivism and violence (Fazel 2017). In practice, risk assessment tools may be used as outcome measures, especially in areas where risk reduction is perceived as a desirable result of treatment, such as forensic mental health services. Tracking risk over time in this way may offer a useful way of measuring this aspect of care, but clinicians should be cautious because the qualities that make a good risk prediction model do not necessarily coincide with those of a good outcome measure.

What is a good outcome measure?

The spotlight on outcome measurement has seen a corresponding rise in the number of different available measures (Kilbourne 2018). So how can you choose the best measurement instrument from those on offer? The science behind the development and assessment of outcome measures has advanced

BJPsych Advances (2020), page 1 of 9 doi: 10.1192/bja.2020.58
Downloaded from https://www.cambridge.org/core. 14 Jun 2021 at 15:28:34, subject to the Cambridge Core terms of use.

considerably in recent years. Leading this process is the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) group, which has created a suite of research tools for evaluating the measurement properties of the growing number of measures (Prinsen 2018). Although these are mainly designed for PROMs, many of the principles translate to measures completed by other raters, such as clinicians.

Recently their emphasis has changed to focus on the content validity of measures, trumping all other parameters of validity and reliability. Assessing the content validity essentially attempts to answer the question of whether the concept of interest is truly being measured – i.e. does this measure capture what is of real importance? Closely related to this is considering the way that a measure has been developed. Complex criteria exist for determining whether a PROM meets the expected standard, revolving around the quantity and quality of patient involvement in the development and testing of the measure (Terwee 2018). COSMIN takes the view that if there is not adequate evidence that this process has been performed satisfactorily then the measure should be rejected and there is no need to further assess its other psychometric properties.

If a measure has adequate content validity and has been well developed, then a range of other properties should be considered. COSMIN identifies structural validity as the next most important property. This relies on sophisticated statistical techniques, such as factor analysis, to determine whether the structure of the measurement instrument reflects the concepts it purports to measure. For example, if an instrument is designed to measure psychotic symptoms and health-related quality of life in two separate subscales, does the statistical structure of the test population's scores on these subscales reflect two coherent and distinct underlying factors?

Another example of validity is whether the measure under consideration shows good correlation with the scores of other measures that capture similar concepts. A subscale of a measure that is designed to measure social functioning in people with psychosis could reasonably be expected to produce broadly similar scores to a well-established general measure of functioning, such as the Global Assessment of Functioning (GAF) (American Psychiatric Association 2000). This relies on the assumption that the comparison measure also measures the concept of interest. One type of reliability is the test–retest property, which determines whether the same person gives similar scores over time, when no or minimal change has taken place.

COSMIN recommends considering both the quality of the evidence and the performance within each measurement property when determining the overall quality of an outcome measure. They go on to make recommendations about the selection and interpretation of outcome measures. Ultimately it is the intended purpose of the outcome measurement that will determine the most appropriate measures to use are. For example, a measure used for research could be much more comprehensive than one selected for routine clinical practice, as researchers may have more time to administer the instrument and require more detailed information.

Core outcome sets

Increasingly, individual organisations and wider initiatives are seeking to define which outcomes should be measured in a particular context (Webbe 2018). These so-called core outcome sets (COS) aim to capture sufficient information to understand the most important aspects of a particular condition or population. Originating in research, their purpose was to define the minimum set of outcomes that should be measured in any clinical trial in a specific area. These could include a range of different types of outcome, including biomarkers, performance measures and participant or observer-reported measures. The purpose is to enable more meaningful comparisons between trials and facilitate the process of systematic review and meta-analysis. A rigorous standardised process for developing a COS for research purposes has been established by the Core Outcome Measurement in Effectiveness Trials (COMET) group (COMET Initiative 2020).

Core outcome sets differentiate between 'what' to measure and 'how' to measure. Qualitative methods, such as in depth interviews and focus groups, are often combined with consensus-building activities, such as Delphi surveys and stakeholder meetings, to determine what should be measured (Williamson 2017). Systematic reviews can then be used to map out relevant outcome measures in the field and assess their quality, allowing experts to select the most appropriate and robust measures for their needs (Gargon 2019).

This concept has now been extrapolated to clinical practice, where there is increasing interest in what should be measured in services treating different medical conditions. Spearheading this process internationally is the International Consortium for Health Outcomes Measurement (ICHOM). Its approach is described in more detail below.

Outcome measurement in psychiatry in context

A wide range of organisations, each with a different focus, have an interest in the development and evaluation of outcome measures in mental health. We present an overview of these perspectives, so that

Downloaded from https://www.cambridge.org/core. 14 Jun 2021 at 15:28:34, subject to the Cambridge Core terms of use.

the reader is aware of the policy context and to signpost a number of useful resources that are available.

Royal College of Psychiatrists

The Royal College of Psychiatrists (RCPsych) has long been interested in good outcome measurement. It will soon publish an occasional paper setting out the College's current position with regard to outcome measurement and that of its constituent faculties. This work has involved extensive consultation with and coordination between the 13 faculties to establish an outcome framework that covers the full range of psychiatric illness and therapeutic approaches. The paper is an expert resource to facilitate conversations about outcome measures between service providers, commissioners of services and those with lived experience using these data to drive service improvement.

The RCPsych endorses the routine use of outcome measures in clinical practice and has previously set out a series of principles to inform the development and selection of outcome measures (Hampson 2011). These principles are outlined in Box 1.

The recommendations set out by Hampson et al stress that the choice of outcome measure is dependent on the purpose for which it will be used. This may differ at the individual versus the service level. The authors emphasise that outcome measurement is a way of safeguarding the interests of patients, by ensuring that they are receiving treatments of established effectiveness, tailored to their specific needs. The paper also highlights the need to record potentially negative outcomes, such as the side-effects of treatments, alongside desired improvements. The role of outcome measures in the improvement of clinical services is also underlined, with the caveat that many external factors must be adequately accounted for, such as the challenges of providing services in areas of socioeconomic deprivation. The paper concludes that there was insufficient evidence to support the linking of funding to service outcomes.

Each individual faculty of the RCPsych has adopted a different approach to outcome measurement, with some setting out frameworks recommending particular outcome measures. Other faculties have defined overarching principles relevant to their specialty, without recommending specific measures.

One example of a highly developed approach is that of the Faculty of Liaison Psychiatry, which launched its Framework for Routine Outcome Measurement in Liaison Psychiatry (FROM-LP) (Trigwell 2015). This built on work partly commissioned by the RCPsych that led to the publication of the Centre for Mental Health's report *Outcomes and Performance in Liaison Psychiatry: Developing a Measurement Framework* (Fossey 2014). The authors identify several important sources of heterogeneity within liaison psychiatry that provide particular challenges to coherent outcome measurement (Trigwell 2016). These include variations in the setting of liaison work, the wide range of conditions treated and the considerable differences in service models. They also discuss the difficulty of measuring longer-term improvements in mental health within the context of acute hospital admissions, which are often very brief, and the difficulty of attributing outcomes specifically to the input of liaison psychiatry services, as opposed to any other intervention.

The Liaison Faculty responded to these challenges by identifying the need for a 'balanced scorecard' that includes measures of structure, process and outcome (Donabedian 2005). The FROM-LP (Trigwell 2015) includes recommended and optional measures of process, including waiting time, clinician-reported measures such as the Clinical Global Impression scale (Busner 2007), PROMs such as Recovering Quality of Life (ReQoL) (Keetharuth 2018), PREMs using patient- and carer-satisfaction scales, and also referrer satisfaction, measured using a dedicated scale. The impact of the framework is evaluated through feedback from faculty members and surveys of its use.

International Consortium for Health Outcomes Measurement (ICHOM)

ICHOM is grounded in the value-based healthcare delivery (VBHCD) approach developed by Porter & Teisberg (Porter 2006). This approach hinges

BOX 1 The Royal College of Psychiatrists' principles for outcome measurement

- Focus should be on what is important to patients and carers
- Measures should be relevant to patients and clinicians
- Measures should be simple and easy to use
- Measures should be clear and unambiguous
- Measures should allow comparisons between teams and services
- Measures should be validated for the purpose for which they are used
- IT support should simplify data collection and analysis, and ensure maximum use of data already collected
- Data should be checked for reliability
- Data should be used at the clinical, team and organisational level
- There should be immediate feedback of the data to patients, carers and clinicians so that outcomes can influence the treatment process

(Hampson 2011)

BJPsych Advances (2020), page 1 of 9 doi: 10.1192/bja.2020.58
Downloaded from https://www.cambridge.org/core. 14 Jun 2021 at 15:28:34, subject to the Cambridge Core terms of use.

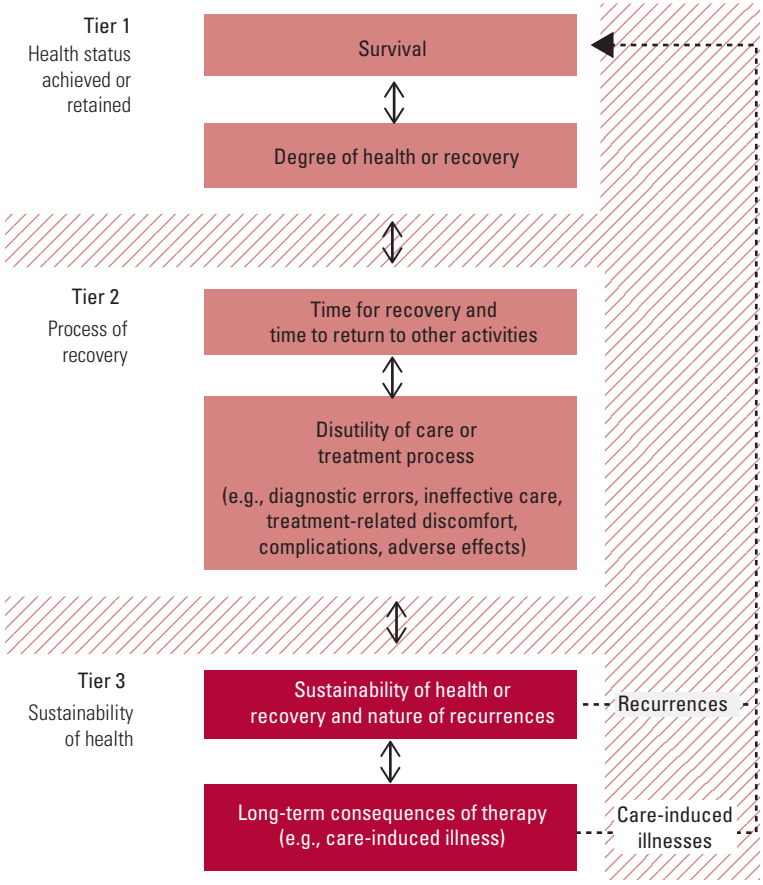


FIG 1 The outcome measures hierarchy (Institute for Strategy & Competitiveness 2020). Reproduced with kind permission of Professor Michael Porter, Harvard Business School.

on a hierarchy of outcomes that matter most to patients, linked to bundled payments made for whole episodes of care. It initially focused on medical and surgical conditions, advocating for treatment to be organised into integrated practice units (IPU), where all interventions relevant to a particular condition can be delivered in a coordinated fashion. The hierarchy of outcomes conceptualised by VBHCD is divided into three tiers, ranging from survival to the long-term consequences of therapy (Fig. 1). This philosophy has subsequently been expanded and adapted to the needs of psychiatric conditions.

ICHOM aims to develop internationally relevant core outcome sets for different conditions. It has developed an iterative methodology to create these sets that includes panels of experts and patient partners, supported by administrative and research staff. The sequence of tasks is (1) to prioritise the outcome domains, (2) to select appropriate outcome measures, (3) to prioritise the case-mix domains and (4) to select the case-mix definitions. ICHOM has initially focused on those disorders that are responsible for some of the greatest burden of disease globally (International Consortium for Health Outcomes Measurement 2020b).

The first published ICHOM mental health outcome set covers anxiety and depression (Obbarius 2017). The outcome domains selected were (a) symptom burden, (b) functioning, (c) disease progression and treatment sustainability and (d) potential side-effects of treatment. The recommendations advocate tracking symptoms via a variety of well-documented general and condition-specific scales, such as the Patient Health Questionnaire (PHQ-9), Generalized Anxiety Disorder scale (GAD-7), Social Phobia Inventory (SPIN) and Inventory for Agoraphobia (IFA). Functioning is tracked via the World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0), as well as work status and disease-related absenteeism. Other parameters include time to recovery, overall success of treatment, recurrence of disease and medication side-effects.

The guidelines also recommend recording a series of factors that define the particular characteristics of individuals within a population, such as age, gender, educational level, work status, comorbidities and

Downloaded from https://www.cambridge.org/core. 14 Jun 2021 at 15:28:34, subject to the Cambridge Core terms of use.

duration of symptoms. These factors influence the likelihood that a particular individual will achieve a good outcome and therefore offer a way of interpreting aggregate outcomes in the context of that particular population. This process of case-mix adjustment enables more accurate comparison between diverse populations. A predominantly elderly population, with more comorbidities, longer duration of symptoms and less social support may be expected to respond differently to an intervention than a younger population without these factors.

The recommendations finish by setting out a proposed schedule for the collection of data, including both the outcome measures and the relevant case-mix factors. This includes a set of recommended measures at baseline, at annual assessment and during treatment monitoring, designed to capture the impact of specific interventions.

ICHOM aims to publish more outcome sets on mental health conditions, including addictions, depression and anxiety in young people, psychotic disorders and personality disorders.

Outcome measurement in commissioning and quality assurance

Outcome measurement is beginning to be used more routinely in commissioning processes and for the quality assurance of existing services. Some providers of mental health services have started to link payments to outcomes. In 2016 the Oxfordshire Clinical Commissioning Group worked with Oxford Health NHS Foundation Trust to develop an outcome-based commissioning model (NHS England and NHS Improvement 2016b). This consisted of a core payment that was fixed, linked with an outcome-based incentive payment and a mechanism that shared the associated risks between commissioners and providers. The agreed structure linked 20% of the funding to outcomes. Of this, 2.5% was in Commissioning for Quality and Innovation (CQUIN) payments, while the rest was linked to seven outcomes that were co-produced between the service and experts by experience (Box 2).

At a national level, outcomes have been used by commissioners within specialised services to monitor and incentivise improved quality of care. One example is the quality dashboards used by the Clinical Reference Groups that support NHS England. These include a range of key descriptors of quality that include measures of both process and outcomes. For example, medium and low secure services for adults are expected to report the proportion of patients with an improved score on the Health of the Nation Outcomes Scales for secure and forensic patients (HoNOS-Secure) on discharge (NHS England 2020).

National Institute for Health and Care Excellence (NICE)

NICE aims to improve outcomes for people using the NHS and other public health and social care services, while reducing variations in quality of care. NICE produces evidence-based guidance and quality standards for health practitioners. It also develops quality standards and performance metrics for providers and commissioners.

NICE has developed a sophisticated methodology for assessing particular interventions, which includes a health economic assessment of the cost-effectiveness of interventions (National Institute for Health and Care Excellence 2014). To do this it uses the impact that conditions and treatments have on quality of life to calculate quality-adjusted life-years (QALYs), which can then be compared with the relevant costs. If the incremental cost-effectiveness ratio (ICER) of an intervention is greater than £30 000, then NICE is unlikely to approve its use without strong additional evidence or rationale. To do this consistently, NICE prefers to use a single outcome measure, the EQ-5D, which contains the five domains of mobility, self-care, usual activities, pain/discomfort and anxiety/depression. The respondent is asked to tick a corresponding statement within each domain that they believe correlates with their current status. Despite containing only one domain that is directly related to mental

BOX 2 Outcomes used for linked payments in Oxford Health NHS Foundation Trust

Outcome 1: people will live longer

Outcome 2: people will improve their level of functioning

Outcome 3: people will receive timely access to assessment and support

Outcome 4: carers feel supported in their caring role

Outcome 5: people will maintain a role that is meaningful to them

Outcome 6: people will continue to live in stable accommodation

Outcome 7: people will have fewer physical health problems related to their mental health

(NHS Improvement 2016)

BJPsych Advances (2020), page 1 of 9 doi: 10.1192/bja.2020.58

Downloaded from https://www.cambridge.org/core. 14 Jun 2021 at 15:28:34, subject to the Cambridge Core terms of use.

health, the EQ-5D is still used by NICE for assessing interventions for mental health conditions.

A range of documents endorsed by NICE that are relevant to mental health services advocate the use of specific outcome measurement approaches to monitor response to treatment. These are often produced collaboratively with other stakeholders. One example is the implementation of the early intervention in psychosis access and waiting time standard (NHS England, NCCMH and NICE 2016). This sets out the measures and timescale for collection expected for outcomes within early intervention in psychosis services.

Patient-Reported Outcomes Measurement Information System (PROMIS)

PROMIS is an initiative funded by the National Institutes of Health (NIH) in the USA. It is part of the overarching health measures programme, which also includes measures for people with neurological illness and sickle cell disease. The NIH toolbox encompasses a series of performance measures of cognitive, motor and sensory function (Northwestern University 2020). PROMIS is specifically designed for people with chronic health conditions and includes over 300 instruments to measure patients' mental health, as well as measuring physical health, social health and overall global wellbeing. Mental health profile domains cover anxiety and depression, with additional domains for adults concerning alcohol use, smoking, substance use, anger, cognitive function, life satisfaction, meaning and purpose, positive affect, psychosocial illness impact and self-efficacy for managing chronic conditions. As well as using conventional 'short-form' questionnaires, PROMIS also uses computer adaptive testing (CAT).

Computer adaptive testing ranks items related to a particular concept in a hierarchy of severity (Gibbons 2016). For example, in depression more people with mild depression would be expected to say that they felt unhappy in the past week than would say that they felt that they wanted to die. This means that feeling unhappy is placed towards the bottom of the hierarchy of questions, whereas feeling that you want to die is nearer to the top. Respondents are initially asked a question that is somewhere in the middle of the hierarchy. Their answer to this determines which question they are asked next, until their position on the scale of severity has been determined. This system has the advantage of identifying the severity of someone's symptoms by using fewer questions than standard approaches, because the sequence of questions is tailored to each respondent. This reduces the burden on respondents by minimising the number of questions they need to answer.

Challenges and opportunities in implementing outcomes

The implementation of outcome measures is a highly complex and evolving area. Outcome measurement can be implemented at the level of the individual patient, as a way of enabling the patient, their friends and family and the clinical team responsible for their care to monitor their progress over time. This can be linked to particular treatments or approaches, as a way of providing an objective assessment of the impact of specific interventions for that patient, which supports clinical decision-making. Outcome measurement can also be used at the aggregate level as a way of determining whether a particular intervention is effective at a population level, to compare services and to improve quality.

Challenges

A number of challenges arise when implementing outcome measurement at an aggregate level. The user-led collective Recovery in the Bin warns that outcome measures can place too much emphasis on the individual patient to recover in a narrowly defined way, which does not take adequate account of the social and political realities affecting a person's wellbeing (Recovery in the Bin 2014). Their 'Unrecovery Star' includes items such as poverty and discrimination, as a way to redress this imbalance (Recovery in the Bin 2017).

The actual or perceived consequences of outcome measurement may affect the way that those producing the measurement behave and therefore the measurement itself. If certain outcomes are seen to be either rewarded or penalised, then this is likely to affect the way that those outcomes are assessed. This process can happen consciously or unconsciously and can take a number of forms. One way is simply to assign more or less favourable scores depending on the desired result. This process can be done inadvertently, particularly when there is genuine equivocation about which response option to choose. Another approach is to undertake measurement at a point in time when the desired response is more likely, for example asking someone to complete a measure immediately after receiving good news to obtain a more positive score.

Careful thought must be given to what incentives are attached to outcome data at the individual and systemic level. Communication with those stakeholders completing outcome measures is essential to build trust and to understand any perceived incentives, which may or may not have been anticipated.

Finally, there is the risk that services will 'cherry-pick' patients they believe are more likely to have good outcomes, neglecting those who will not. The latter may be more vulnerable to begin with, so it

Downloaded from https://www.cambridge.org/core. 14 Jun 2021 at 15:28:34, subject to the Cambridge Core terms of use.

MCQ answers
1 c 2 d 3 e 4 b 5 a

is essential that they are not denied the support they need. The use of validated case-mix data can be used to prevent cherry-picking by adjusting for characteristics of a population that may lead to less favourable outcomes.

Opportunities

Data should be collected in a consistent and reliable way, so that the quality of resultant information can be widely trusted. Digital approaches have provided significant improvements in the collection, analysis and presentation of information. Outcomes are now often routinely integrated into electronic patient records, meaning that responses can be recorded faster and more reliably. Digital technology offers greater opportunities for data to be collected over time and presented back to patients and teams to provide a longitudinal graphical representation of changes.

The forthcoming RCPsych paper will recommend using small numbers of well-validated outcomes, consistently implemented and coherently analysed. This echoes the technical guidance issued by NHS England and NHS Improvement, which recommends using only three to seven outcome measures (NHS England and NHS Improvement 2016c). One example where outcome measurement has been implemented at a national level is through the Improving Access to Psychological Therapies (IAPT) programme (National Collaborating Centre for Mental Health 2019). The IAPT programme aims to expand the availability of psychological therapies and monitor the effectiveness of these interventions through integral outcome monitoring. The manual prescribes the type and frequency of measurement for a range of conditions. It also outlines information about how to maintain data quality, stressing the importance of paired measurement, in order to track progress over time.

The Long Term Plan for Mental Health highlights that all providers must submit data to the Mental Health Services Data Set (MHSDS) and IAPT dataset (NHS England 2019b). Transparency regarding the achievement of outcomes and levels of quality supports patient choice, enables benchmarking of care across equivalent services, supports workforce planning and ensures the effective use of resources.

Conclusions

Outcome measurement is a hot topic in healthcare as a whole, including mental health services. Psychiatrists will increasingly be expected to have a basic understanding of how to measure outcomes effectively within their own scope of practice. It is essential to choose outcome measures carefully, ensuring that they are appropriate for the context of use and capture the whole range of relevant domains, not just clinical symptoms. Outcome measures must focus on measuring what is most important and this needs to prioritise the patient's perspective. Implementing outcome measurement offers great potential, but has a number of challenges. Psychiatrists must be aware of these pitfalls and take active steps to guarantee that outcomes are measured robustly, in a way that can lead to meaningful improvement.

Author contributions

All authors fulfil the ICMJE criteria for authorship.

Declaration of interest

J.C. is the Lead for Outcomes in Mental Health and the Specialist Advisor for Mental Health Payment Systems for the Royal College of Psychiatrists.

ICMJE forms are in the supplementary material, available online at https://doi.org/10.1192/bja.2020.58.

References

American Psychiatric Association (2000) *Diagnostic and Statistical Manual of Mental Disorders* (4th edn text revision) (DSM-IV-TR). APA.

Busner J, Targum SD (2007) The clinical global impressions scale: applying a research tool in clinical practice. *Psychiatry (Edgmont)*, **4**: 28–37.

COMET Initiative (2020) *Core Outcome Measures in Effectiveness Trials*. COMET Initiative (http://www.comet-initiative.org/). Accessed 17 Apr 2020.

Dawber TR, Meadors GF, Moore FE Jr (1951) Epidemiological approaches to heart disease: the Framingham Study. *American Journal of Public Health*, **41**: 279–81.

Donabedian A (2005) Evaluating the quality of medical care, 1966. *Milbank Quarterly*, **83**: 691–729.

Fazel S, Wolf A, Larsson H, Lichtenstein P, Mallett S, Fanshawe TR (2017) Identification of low risk of violent crime in severe mental illness with a clinical prediction tool (Oxford Mental Illness and Violence tool [OxMIV]): a derivation and validation study. *Lancet Psychiatry*, **4**: 461–468.

First MB, Williams JBW, Karg RS, et al (2015) *Structured Clinical Interview for DSM-5 – Research Version (SCID-5-RV)*. American Psychiatric Association.

Fossey M, Parsonage M (2014) *Outcomes and Performance in Liaison Psychiatry: Developing a Measurement Framework*. Centre for Mental Health.

Gargon E, Gorst SL, Williamson PR (2019) Choosing important health outcomes for comparative effectiveness research: 5th annual update to a systematic review of core outcome sets for research. *PLoS One*, **14**: e0225980.

Gibbons RD, Weiss DJ, Frank E, et al (2016) Computerized adaptive diagnosis and testing of mental health disorders. *Annual Review of Clinical Psychology*, **12**: 83–104.

Hampson M, Killaspy H, Mynors-Wallis L (2011) *Outcome Measures Recommended for Use in Adult Psychiatry (Occasional Paper OP78)*. Royal College of Psychiatrists.

Institute for Strategy & Competitiveness (2020) *Measure Outcomes & Cost for Every Patient*. Harvard Business School (https://www.isc.hbs.edu/health-care/value-based-health-care/key-concepts/Pages/measure-outcomes-and-cost.aspx). Accessed 22 Jul 2020.

International Consortium for Health Outcomes Measurement (2020a) *What Matters Most: Patient Outcomes and the Transformation of Healthcare*. ICHOM.

Downloaded from https://www.cambridge.org/core. 14 Jun 2021 at 15:28:34, subject to the Cambridge Core terms of use.

International Consortium for Health Outcomes Measurement (2020b) *Our Methodology*. ICHOM (https://www.ichom.org/standard-sets/#methodology). Accessed 26 Mar 2020.

Keetharuth AD, Brazier J, Connell J, et al (2018) Recovering Quality of Life (ReQoL): a new generic self-reported outcome measure for use with people experiencing mental health difficulties. *British Journal of Psychiatry*, **212**: 42–9.

Kilbourne AM, Beck K, Spaeth-Rublee B, et al (2018) Measuring and improving the quality of mental health care: a global perspective. *World Psychiatry*, **17**: 30–8.

Kingdon D (2019) Outcome-based providing and commissioning: pathways and standards. *BJPsych Bulletin*, **43**: 282–86.

Leekam S (2013) Diagnostic Interview for Social and Communication Disorders. In *Encyclopedia of Autism Spectrum Disorders* (ed FR Volkmar). Springer.

Lewis G, Killaspy H (2014) Getting the measure of outcomes in clinical practice. *Advances in Psychiatric Treatment*, **20**: 165–71.

National Collaborating Centre for Mental Health (2019) *The Improving Access to Psychological Therapies Manual*. NCCMH.

National Institute for Health and Care Excellence (2014) *Developing NICE Guidelines: The Manual. Processes and Methods (PMG20) (Last updated: 17 July 2020)*. NICE (http://www.nice.org.uk/process/pmg20).

NHS England (2019a) *What is NHS RightCare?* NHS England (https://www.england.nhs.uk/rightcare/what-is-nhs-rightcare/).

NHS England (2019b) *Further Progress on Care Quality and Outcomes: Adult Mental Health Services. NHS Long-Term Plan*. NHS.UK (https://www.longtermplan.nhs.uk/online-version/chapter-3-further-progress-on-care-quality-and-outcomes/better-care-for-major-health-conditions/adult-mental-health-services/).

NHS England (2020) *Specialised Services Quality Dashboards*. NHS England (https://www.england.nhs.uk/commissioning/spec-services/npc-crg/spec-dashboards/). Accessed 26 Mar 2020.

NHS England, National Collaborating Centre for Mental Health, National Institute for Health and Care Excellence (2016) *Implementing the Early Intervention in Psychosis Access and Waiting Time Standard: Guidance*. NICE.

NHS England, NHS Improvement (2016a) *Delivering the Five Year Forward View for Mental Health: Developing Quality and Outcomes Measures*. NHS England, NHS Improvement.

NHS England, NHS Improvement (2016b) *Oxford Health NHS Foundation Trust Outcome Based Commissioning Model for Mental Health: Updated Outcome Measures*. NHS Improvement, NHS England.

NHS England, NHS Improvement (2016c) *Linking Quality and Outcome Measures to Payment for Mental Health: Technical Guidance*. NHS England, NHS Improvement.

Northwestern University (2020) *Intro to PROMIS®*. Northwestern University (http://www.healthmeasures.net/explore-measurement-systems/promis/intro-to-promis). Accessed 26 Mar 2020.

Obbarius A, Van Maasakkers L, Baer L, et al (2017) Standardization of health outcomes assessment for depression and anxiety: recommendations from the ICHOM Depression and Anxiety Working Group. *Quality of Life Research*, **26**: 3211–25.

Porter ME, Teisberg EO (2006) *Redefining Health Care: Creating Value-Based Competition on Results*. Harvard Business School Press.

Prinsen CAC, Mokkink LB, Bouter LM, et al (2018) COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, **27**: 1147–57.

Recovery in the Bin (2014) *Recovery in the Bin (RITB) – Key Principles (updated 2017)*. RITB (https://recoveryinthebin.org/ritbkey-principles/). Accessed 22 Jul 2020.

Recovery in the Bin (2017) *Unrecovery Star*. RITB (https://recoveryinthebin.org/unrecovery-star-2/). Accessed 26 Mar 2020.

Terwee CB, Prinsen CAC, Chiarotto A, et al (2018) COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Quality of Life Research*, **27**: 1159–70.

Trigwell P, Kurstow J, Santhouse A, et al (2015) *Framework for Routine Outcome Measurement in Liaison Psychiatry (FROM-LP) (Faculty Report FR/LP/02)*. Royal College of Psychiatrists.

Trigwell P, Kustow J (2016) A multidimensional Framework for Routine Outcome Measurement in Liaison Psychiatry (FROM-LP). *BJPsych Bulletin*, **40**: 192–4.

Webbe J, Sinha I, Gale C (2018) Core outcome sets. *Archives of Disease in Childhood, Education & Practice Edition*, **103**: 163–6.

Williamson PR, Altman DG, Bagley H, et al (2017) The COMET Handbook: version 1.0. *Trials*, **18**: 280.

Wilson PWF, D'Agostino RB, Levy D, et al (1998) Prediction of coronary heart disease using risk factor categories. *Circulation*, **97**: 1837–47.

MCQs

Select the single best option for each question stem

1 The Structured Clinical Interview for DSM-5 – Research Version (SCID-5-RV) is an example of:

a a patient-reported outcome measure
b a performance measure
c a diagnostic instrument
d a risk assessment
e a clinician-reported outcome measure.

2 The COnsensus Standards for health Measurement Instruments (COSMIN) group approach prioritises:

a structural validity
b test–retest reliability
c the clinician's perspective
d content validity
e reducing costs.

3 Which of the following is not a guiding principle of outcome measurement for the Royal College of Psychiatrists?

a measures should be clear and unambiguous
b the focus should be on what is important to patients and carers
c measures should be relevant to patients and clinicians
d measures should be simple and easy to use
e measures should focus only on psychiatric symptoms.

4 Which of the following is not part of the value-based healthcare delivery (VBHCD) hierarchy of outcomes?

a time for recovery and time to return to other activities
b arranging services into integrated practice units
c survival
d degree of health or recovery
e disutility of care or treatment process.

5 Computer adaptive testing:

a ranks items related to a particular concept in a hierarchy of severity
b is not recommended by the Patient Reported Outcomes Measurement Information System (PROMIS)
c involves additional respondent burden over paper-based measurement
d is inappropriate for psychiatric illnesses
e is recommended by the Framework for Routine Outcome Measurement in Liaison Psychiatry (FROM-LP).

Downloaded from https://www.cambridge.org/core. 14 Jun 2021 at 15:28:34, subject to the Cambridge Core terms of use.