

## Routine mental health outcome measurement in the UK

ALASTAIR J. D. MACDONALD<sup>1</sup> & ANDREW J. B. FUGARD<sup>2</sup>

<sup>1</sup>Trust Outcomes Team, BRC Nucleus, South London & Maudsley NHS Foundation Trust, Maudsley Hospital, London, and <sup>2</sup>Research Department of Clinical, Educational and Health Psychology, University College London, UK

### Abstract

This paper offers a short history of routine clinical outcomes measurement (RCOM) in UK mental health services. RCOM developments in primary and secondary care are described, with reference to measures currently in widespread use or likely to be implemented. Assessment procedure and completion rates are discussed. Some of the forces operating in this field are enumerated. Comparison is made with UK attempts at routine outcomes measurement in public education. This field is thus reviewed for lessons for RCOM, and opportunities and challenges considered.

### Introduction: definitions and circumspections

Routine clinical outcomes measurement (RCOM) is taken here to mean the measurement of health status change (i.e. between at least two points in time) in a service-user population, usually with the intention of inferring how much – or little – clinical interventions have helped. No sense can be made of health change without also understanding interventions and the context in which these are applied. We distinguish outcomes measurement from Donabedian's other components of health service quality – process and structure (Donabedian, 1966) – and also from public health approaches to the mental health of whole populations, but we would include attempts to evaluate specific service delivery changes by RCOM.

Within the general embrace of a health service 'free at the point of access', the UK has no single national health service (NHS). Scotland, Northern Ireland and, since 2001, Wales have separate arrangements for health service policy, management and delivery. In Scotland there is no mandated or national system for RCOM, although large patient outcomes surveys have been carried out. In Wales and Scotland 'outcomes frameworks' have been developed to measure the impact of policies on the mental health of the whole population, for instance the average scores of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) (Table 1) (Tennant et al., 2007) from the Scottish Health Survey. In Northern Ireland the emphasis has been on measuring mental health recovery, but without yet clear agreement of how this can be done. What follows therefore predominately relates to England.

Of great importance in the use of rating scales in any context are their psychometric properties. In RCOM the primary concerns are validity, inter-rater reliability and test-retest reliability. There have been relatively few studies in UK clinical populations of psychometric properties of measures coming into widespread use, such as the Health of the Nation Outcome Scales. One reason for this could be an assumption that once the properties are established in one population, that this is likely to generalize to others. However, contexts can vary, and just as randomized controlled trials of treatments need to be replicated in different settings, so too should evaluations of psychometric properties. Given the breadth of measures and scarcity of relevant evidence, psychometric properties are not provided in this paper.

### Mental health services in the UK and their patients

Services are provided by the NHS in primary care settings (often but not always involving initial contact with general medical practitioners (GPs)), in secondary specialist mental health services (usually after referral from GPs), and in tertiary services such as secure forensic milieus (Deakin & Bhugra, 2012). Most mental health issues occur in and are dealt with in primary care (King et al., 2008), either through informal self-funded counselling, private psychotherapy services, charitable organizations, e.g. for relationship or bereavement problems, or funded counselling services attached to general practices, schools, colleges, universities and some workplaces.

Correspondence: Professor Alastair J.D. Macdonald, MD, FRCPsych, Clinical Advisor, Trust Outcomes Team, BRC Nucleus, South London & Maudsley NHS Foundation Trust, Maudsley Hospital, Denmark Hill, London SE5 8AZ, UK. E-mail: alastair.macdonald@kcl.ac.uk

(Received 3 December 2014; accepted 27 January 2015)

ISSN 0954-0261 print/ISSN 1369-1627 online © 2015 Institute of Psychiatry  
DOI: 10.3109/09540261.2015.1015505

Table 1. Outcomes measures in use, or likely to be used in the future, in the UK.

Name	Type	Measurement aim	Clinical setting	Disorders	Age range	Notes on UK use
Health of the Nation Outcome Scales (HoNOS) (Wing et al., 1998)	CROM	Clinical change in symptoms, physical health, functioning and social problems	Secondary mental healthcare, adults	All in setting	18 and over	Widely used
Health of the Nation Outcome Scales for older people (HoNOS65+) (Burns et al., 1999)	CROM	Clinical change in symptoms, physical health, functioning and social problems	Secondary mental healthcare, older adults	All in setting	65 and over	Decreasingly used, most older patients now rated with HoNOS
Health of the Nation Outcome Scales: Child and Adolescent Mental Health (HONOSCA) (Gowers et al., 1999)	CROM	Clinical change in symptoms, physical health, functioning, social problems and educational problems	Secondary mental healthcare, CAMHS	All in setting	4–18	Widely used but mostly in inpatient units
Health of the Nation Outcome Scales for people with Learning Disabilities (HoNOS-LD) (Roy et al., 2002)	CROM	Clinical change in symptoms, seizures, functioning and social problems	Secondary mental healthcare, learning disabilities	All in setting	18 and over	Not yet in widespread routine use
Health of the Nation Outcome Scales for Secure Settings (HoNOS-Secure) (Dickens et al., 2007)	CROM	Clinical change in symptoms, functioning, physical health, social problems and risk	Secondary and tertiary mental healthcare, forensic	All in setting	18 and over	Widely used
Children's Global Assessment Scale (CGAS) (Shaffer et al., 1983)	CROM	Global clinical change	Secondary mental healthcare, CAMHS	All in setting	4–18	Widely used
Treatment Outcomes Profile (TOP) (Marsden et al., 2008)	CROM	Clinical change in dependency, risk, crime and social functioning	Secondary mental healthcare, addictions	Substance misuse	18 and over	Widely used
DIALOG (Priebe et al., 2007)	PROM, structured clinical communication tool	Clinical change in symptoms, physical health, functioning, social problems and personal safety	Secondary mental healthcare, adults	Severe mental illness	18 and over	Session-by-session tool. Not yet in widespread routine use
Warwick-Edinburgh Wellbeing Scale (WEMWBS) (Tennant et al., 2007)	PROM	Clinical change in mood, energy, cognition and relationships	Secondary mental healthcare, adults	Severe mental illness	18 and over	Not yet in widespread routine use
Also Short version (SWEMWBS) (Stewart-Brown et al., 2009)						
Mental Health Recovery Star (MHRS) (MacKeith & Burns, 2008)	PROM, structured clinical communication tool	Recovery	Secondary mental healthcare, adults	Severe mental illness	18 and over	Not yet in widespread routine use

(Continued)

Table 1. (Continued)

Name	Type	Measurement aim	Clinical setting	Disorders	Age range	Notes on UK use
Questionnaire on the Process of Recovery (QPR) (Law et al., 2014)	PROM, process measure	Recovery	Secondary mental healthcare, adults	Severe mental illness	18 and over	Not yet in widespread routine use
DEMQOL (Smith et al., 2005)	PROM	Change in quality of life	Secondary mental healthcare, adults with dementia	Dementia disorders	65 and over	Not yet in widespread routine use
DEMQOL-PROXY (Rowen et al., 2012)	CROM	Change in quality of life	Secondary mental healthcare, adults with dementia	Dementia disorders	65 and over	Not yet in widespread routine use
Clinical Outcomes in Routine Evaluation Outcomes Measure (CORE-OM) (Evans et al., 2002) Also short version (CORE-10) (Barkham et al., 2013)	PROM	Clinical change in subjective well-being, symptoms, life functioning and risk of harm	Primary and secondary mental healthcare, adults having formal psychotherapy	Psychological disorders	18 and over	Widely used
Young Person's CORE (YP-CORE) (Twigg et al., 2014)	PROM	Clinical change in subjective well-being, symptoms, life functioning and risk of harm	Primary care and counselling services	Psychological disorders	11–16	Not yet in widespread routine use
Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997)	PROM, carer and teacher measures	Clinical change in positive and negative aspects of behaviour, social relationships and mood	Secondary mental healthcare, CAMHS	All disorders in setting	PROM (11–17) Parent/teacher 2–17 (two versions: 2–4, 4–17)	Increasingly widely used
The Systemic Clinical Outcome and Routine Evaluation Index of Family Function and Change (SCORE) (Fay et al., 2013; Jewell et al., 2013)	CROM	Family system change	Secondary mental healthcare, CAMHS	All disorders in setting	Children 12 and over and adults	Not yet in widespread routine use
Brief Parental Self-Efficacy Scale (Woolgar, 2014)	Carer-reported measure	Change in parenting skills	Primary and Secondary Mental Health	All disorders in setting	Parents of children 0–18	Not yet in widespread routine use
The Revised Child Anxiety and Depression Scale (RCADS) (Chorpita et al., 2000) Also parent version (Ebesutani et al., 2010)	PROM, CROM	Clinical change in specific symptoms	Care-CAMHS Primary mental healthcare, CYP IAPT	Anxiety and depression symptoms	8–18	Not yet in widespread routine use
Child Outcome Rating Scale ((C)ORS) (Miller et al., 2005)	PROM, structured clinical communication tool	Clinical change	Primary mental healthcare, CYP IAPT	All disorders in setting	6–12	Not yet in widespread routine use; Session-by-session rating

(Continued)

Table 1. (Continued)

Name	Type	Measurement aim	Clinical setting	Disorders	Age range	Notes on UK use
Generalized Anxiety Disorder 7 (GAD-7) (Spitzer, Kroenke, Williams, & Lowe, 2006)	PROM	Clinical change in anxiety symptoms	Primary mental healthcare, IAPT	Anxiety and depressive disorders	18 and over	In widespread use
Patient Health Questionnaire-9 (PHQ-9) (Kroenke, Spitzer, & Williams, 2001)	PROM	Clinical change in anxiety symptoms	Primary mental healthcare, IAPT	Anxiety and depressive disorders	18 and over	In widespread use
Recovering Quality of Life Measure (REQoL) (Brazier, 2015)	PROM	Mental health focussed recovery and quality of life items	Secondary Mental Health Care-Adults	All disorders excluding organic disorders and learning disabilities	18 and over	In development

CROM, Carer-reported outcome measure; CYP IAPT, Children and Young Persons Improving Access to Psychological Therapies; PROM, patient-reported outcome measure.

Depressive and anxiety disorders predominate. Severe mental illness is usually initially treated in secondary care by state-funded NHS services, but a few with short-term illnesses such as major depressive or bipolar disorder and only a small proportion of patients with chronic severe illness remain in secondary care – many are discharged back into the care of their GP once any acute phase has passed. Secondary care is community-orientated with patients assessed and treated in clinics, hospital accident and emergency departments or at home; inpatient admissions are short – a median of 17 days for working-age patients (Health and Social Care Information Centre UK, 2013). Child and adolescent mental health services (CAMHS) are even more community-orientated and mostly for those aged up to 18 years old together with their carers. For adults RCOM has developed separately in primary and secondary care in the UK. Generic services for people aged 18–65, old age services for most people aged 65 or over (except those ageing with chronic mental illness), learning disability and forensic services for all ages are all at different stages on different paths towards RCOM, with significant geographic variation.

### A brief history of RCOM in the UK

Some elements of RCOM (that of clinical change, intervention or context) have been implemented in mental health services in the UK and elsewhere for at least 150 years. Rates of ‘recoveries’, or even ‘cures’ were reported from many institutions across the UK and Europe (Tuke, 1854), e.g. sceptically reported by Hood (Hood, 1862), who, like Thurnham (1845) before him pondered on methodology – crucially the denominator. Do we measure outcomes for those whose treatment has ended, or, less gratifyingly, for all patients? In some cases changes in outcomes before and after management reform – such as the ‘institution of kindness for severity, and indulgence for restrictions’ – were used to justify them (Timbs, 1868).

Production of such outcomes data, almost certainly sustained (and corrupted) by commercial considerations, appears to have been extinguished in the UK with the advent of the NHS in 1948. Little systematic RCOM activity continued after the dawn of community care in the 1990s. Here and there in the UK the flame of interest was kept alight by solitary individuals and teams, but this was often despite local management or central government attitudes, at best indifferent and at worst hostile. It was very difficult to implement paper-based RCOM without the administrative machinery available in the asylums. Also ‘evidence-based medicine’ (in which the best outcomes were assumed and observational data relegated) took hold. When the notion of a hierarchy of evidence became more amenable to challenge

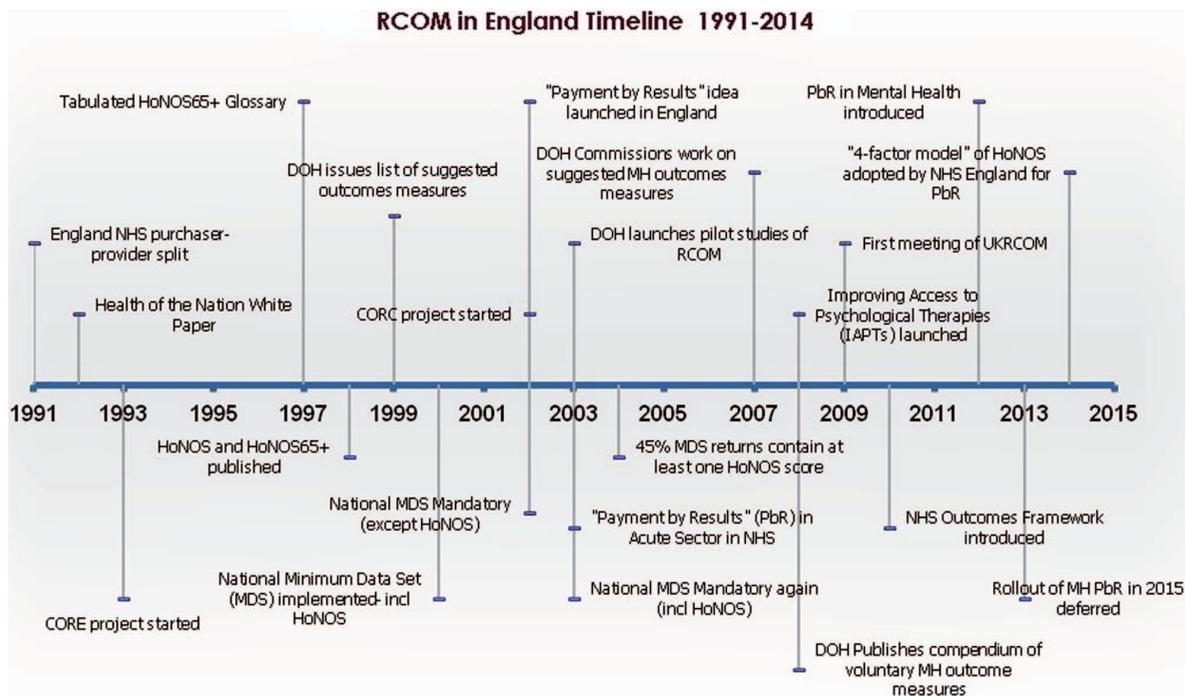


Fig. 1. Some landmarks in RCOM in England.

(Upshur & Tracy, 2004) and, at the same time, information technology began to achieve reasonable levels of reliability, the conditions became once again ripe for RCOM.

Fig. 1 shows some key national milestones in development of RCOM in England. What is not shown here is the emergence of purely local RCOM initiatives in the late 1990s and early 21st century, in parts of London and Hampshire for instance.

From this account we can arbitrarily identify six steps significant for RCOM in England, and one major theme. The first step, and one which separated England and Wales from Scotland, was the creation of a healthcare market in 1991 in which taxpayer's money was given to parts of the NHS completely separate from those providing care and treatment. The second significant step for secondary mental health care was the 1992 white paper 'The Health of the Nation' (Secretary of State for Health, 1992) containing the aim of improving the mental health of people with serious mental illness. The imperative was then to evidence this improvement. A 'clinician-reported outcome measure' (CROM) approach was favoured and the Health of the Nation Outcomes Scales (HoNOS) (Table 1) for working-age adults with serious mental illness in secondary care were commissioned (Wing et al., 1998), followed by variants for older adults, children and adolescents, people with learning disabilities, those in secure settings and with acute brain injury. Some of these were routinely implemented in some English services from 1997 (Macdonald, 2002), whilst the majority hung

back (Gilbody et al., 2002). The third step was the funding by the Mental Health Foundation charity of the development of a 'patient-reported outcome measure' (PROM) for routine use in psychological therapies, the Clinical Outcomes in Routine Evaluation Outcome Measure (CORE-OM) (Table 1) in 1993. The fourth step was the founding of the first RCOM group independent of NHS management – the Child Outcomes Research Consortium described below. The fifth step was massive investment in psychological therapies in primary care – the Improving Access to Psychological Therapies programme – with RCOM built in, again described below. The latest step has been the plan to implement 'payment by results' (actually payment by activity-related cost) in mental health services in a similar way to that developed in the acute hospital sector in 2003, only made possible with the purchaser–provider split 12 years earlier. All secondary mental health patients in England are now classified into 20 'clusters' of complexity and service need by the clinicians, for which it is planned different tariffs will be payable by commissioners. Relevant to RCOM is that the clusters are determined by mandatory use of the HoNOS.

The theme that emerges from the timeline is, for secondary care at least, central ambivalence towards outcomes measurement. Unlike Australia, there appeared to be reluctance to insist on any particular measures, and progress towards an outcomes-based culture in the NHS has been very slow. A list of voluntary measures was issued in 1999 and pilot studies using the HoNOS took place in 2003, but before

these studies reported, returns of HoNOS to government became ‘mandatory’. Four years later the Department of Health commissioned a committee to investigate and report on suitable measures, finally issuing a large menu of measures for voluntary selection three years after that. By now the use of HoNOS, despite being ‘mandatory’, was slightly more widespread for outcomes measurement, and it was decided to make HoNOS scores, with some additions, the main tool for cluster allocation and thus for ‘payment by results’ with its financial imperatives. This led to a very rapid deployment of HoNOS across England, so that by 2014 the vast majority of secondary mental healthcare service users in England had at least one HoNOS rating. Unfortunately for outcomes measurement, central emphasis on initial assessment ratings and only tardy support for follow-up ones meant that in services most advanced in RCOM the availability of at least pairs of ratings necessary to show change actually declined significantly. However, it is demonstrably the case that, because of this initiative, there are many more HoNOS data available for misunderstanding than there were before.

### **RCOM in different settings and client groups**

Below we describe the current position in the UK, and refer to the measures used in the different parts of the service described above. We report measures that are either in routine use or those we deem likely to become important in the future. For reasons stated above, we have not selected measures based on their psychometrics. Measures, their sources, and acronyms can be found in Table 1.

#### *RCOM in primary healthcare*

Although counselling and psychotherapy services have been steadily developing in primary healthcare for adults in the UK over the past two decades, there has been little formal RCOM until the recent central funding of the Improving Access to Psychological Therapies (IAPT) programme, a specific programme based mainly on cognitive behavioural interventions, described below. Several studies have reported the use of the CORE-OM (see Table 1) in a variety of counselling and psychotherapy services, not necessarily associated with GPs. However, CORE-OM for adults is much more associated with RCOM in psychological therapies in secondary services.

#### *Improving Access to Psychological Therapies*

Deployment of psychological therapies – mainly cognitive behavioural – in a stepped fashion to all primary care patients with psychological problems is

arguably the greatest expansion of mental health services anywhere in the world. A cost–benefit analysis argued that the cost to taxpayers ‘would be fully covered by the savings in incapacity benefits and extra taxes that result from more people being able to work’ (Layard et al., 2007). RCOM was built in to its design (Clark, 2011). Session-by-session PROMs, introduced in 2006 during pilot trials as outcomes, measured using session-by-session questionnaires, were worse for service users who failed to complete two measures (Clark et al., 2009). IAPT service users are asked to complete a series of measures including the Generalized Anxiety Disorder 7 (GAD-7) and the Patient Health Questionnaire-9 (see Table 1 for details). Additional measures are used with specific anxiety disorders such as obsessive–compulsive disorder since GAD-7 does not cover these (Clark, 2011). To renew contracts with commissioners, IAPT providers must complete measures to certain levels, submit them centrally, and increasingly must also demonstrate significant client improvement in the measures: payment by outcomes.

#### *Children and young people’s IAPT*

Children and Young People’s IAPT (CYP IAPT) (Wolpert et al., 2012) is now being rolled out in the UK and follows the lead of adult IAPT in requiring the use of session-by-session measures. As the Child Outcomes Research Consortium (CORC) (described below), closely involved in the development of CYP IAPT, operates by gathering and feeding back data from secondary CAMH services, they use a very similar system for session-by-session data regardless of whether a service is a member of IAPT: ‘CORC+’.

The Strengths and Difficulties Questionnaire (SDQ) (Table 1) is a five-dimensional measure covering symptoms of hyperactivity, emotional problems, conduct problems, peer problems, as well as the strength of being prosocial. Each subscale has five questions. The SDQ was extended to include questions asking about the impact of problems, for instance how severe problems are, for how long they have been present, and in what areas are they shown (home life, friendships, classroom, or leisure) (Goodman, 1999). There are self- (for 11–17 year olds), parent-, and teacher-rated (both in two versions, for 2–4 and 4–17 year olds) versions of the SDQ which all ask questions in the same general format.

An advantage of the SDQ for the UK is extensive normative data for the general population (<http://www.sdqinfo.com/norms/UKNorms.html>). Another advantage – currently available only for the parent version – is that an ‘added-value score’ (AVS) can be calculated using someone’s initial scores and scores 6 months later. (Ford et al., 2009; Rotheray et al., 2014) The AVS allows the impact of treatment

to be estimated above and beyond spontaneous improvement and factors such as regression to the mean. So far there have been two tests of the score's predictions using data from randomized controlled trials (RCTs) of parent training. As expected, the AVS for the control group was close to zero and the AVS for the treatment group was close to the difference between control group and treatment group found in the RCT.

CYP IAPT introduced several measures in addition to the SDQ. The Revised Child Anxiety and Depression Scale (RCADS) (see Table 1) is a 47-item measure and comes in both young person self-report and parent-report versions. The subscales are separation anxiety, social phobia, generalized anxiety disorder, panic disorder, obsessive-compulsive disorder, and major depressive disorder. The RCADS is recommended for use at assessment and case closure in full, and one (or two) subscales used, if applicable, session-by-session.

One of the main issues with the use of RCADS in the UK is that the current norms come from Hawaii. Also, to date there are no publications on the psychometric properties of the subscales when used individually. In the full 47-item measure, all the subscales are distributed randomly throughout; when used separately the items appear alongside each other; as yet it is unknown whether this affects how people respond.

## RCOM in secondary and tertiary mental healthcare

### CAMHS

Probably the most used CROM in CAMHS is the Children's Global Assessment Scale (CGAS) (see Table 1), an adaptation of the adult Global Assessment Scale (GAS) (Endicott et al., 1976). CGAS is a one-item rating between 1 and 100 of the overall level of functioning of a young person, where 1 is lowest and 100 highest. Scores in the range 1–70 indicate problems (e.g. 1–10 indicates the need for constant supervision, 61–70 indicates some difficulty in a single area) and 71–100 indicates normal functioning (91–100 is 'superior' functioning). Detailed descriptions and vignettes are used to guide the choice of rating. Interestingly, within a group of participants untrained in the CGAS, 84% of doctors and 73% of psychologists were within  $\pm 5$  points of an expert group consensus rating, suggesting that with the instructions in the measure alone high reliability can be reached by many professionals (Lundh et al., 2010).

Systemic interventions of various kinds are common in CAMHS and the desired outcome might be an improvement in family functioning rather than

(only) a reduction of an individual's symptoms. The Systemic Clinical Outcome and Routine Evaluation Index of Family Function and Change (SCORE) (see Table 1) measures these sorts of factors. Parent training typically targets 'conduct disorder' and 'oppositional defiant disorder' displayed by children, so measures of behavioural problems in the child are used as an outcome; for instance the conduct scale of the SDQ or the CORC (see below) and CYP IAPT parent-rated measure based on DSM-IV criteria (no published norms are yet available). Another measure being introduced is the Brief Parental Self-Efficacy Scale developed by Matt Woolgar (King's College London; no published norms are yet available).

Session-by-session measures have also been used in secondary care, notably the Partners for Change Outcome Monitoring System (Miller et al., 2005) in Lincolnshire NHS CAMHS (Timimi et al., 2013).

### Services for severe mental illness in adults

*Patient-reported measures.* PROMs have been piloted in patients with severe mental illness in some services in the UK but many of these have not been published or only as one-off feasibility trials of RCOM. Priebe et al. (2012) have reported the results of subjective quality of life items in DIALOG (see Table 1), a structured communication tool in mental health services. Some UK services have plans to pilot these for routine use.

Several outcomes measures were submitted to panels of service users by Crawford et al. (2011). Their highest ratings were given to scales measuring the side-effects of medication. The next highest rated was the Warwick-Edinburgh Well-Being Scale (WEMWBS) (see Table 1), a 14-item positively phrased Likert-style scale, with a 7-item shorter version developed by Rasch analysis. The latter has been piloted in several UK services, and has been found acceptable in many except those in which a high proportion of patients have dementia. Central government has recently funded a large-scale pilot in which it was used alongside HoNOS; of over 120,000 forms issued in five mental health services 28,204 (23.8%) were completed, a figure similar to that found in Australia with different measures.

The CORE-OM – either in its original 34-item version or short form, the CORE-10 – has been mainly used for the routine assessment of outcomes in psychological therapies in secondary mental healthcare, especially cognitive-behavioural therapy for anxiety and depression. The rigour of its development and psychometric assessment is unparalleled in routine measures used in the UK, and in particular normative data are published (Connell et al., 2007) and the test-retest reliability has been estimated

(Evans et al., 2002) which allows the calculation of reliable and clinically significant change using methods suggested by Jacobson & Truax (Jacobson & Truax, 1991). CORE developers request that anonymized data be submitted centrally, and analyses of these allow serious questions about effectiveness to be explored (Barkham et al., 2012).

One of the issues of PROMs is the degree with which, although completed by service users, they address actual concerns of service users – particularly in the domains now covered by the notion of ‘recovery’ – as opposed to those of clinicians, managers, service providers and governments; other stakeholders in the outcomes grid (Long & Jefferson, 1999). The Mental Health Recovery Star (Table 1) is designed with this in mind. It is completed by staff and service users together. Despite Killaspy et al.’s critique (2012), experience by United Kingdom Routine Clinical Outcomes Measurement in Mental Health Network (UKRCOM) members suggests that it is a useful tool for a collaborative approach to care planning and perhaps outcomes measurement.

Other recovery-orientated PROMS in development in the UK include the Recovering Quality of Life (ReQoL) (Table 1), a mental-health-focused alternative to the very generic EQ-5D (EuroQol group, 1990) in other health domains. The Questionnaire on the Process of Recovery (QPR) (Table 1) has been developed by service users and reflects change in the sense of recovery by them, also including process elements.

DEMQOL (Table 1) measures change in the quality of life for people with dementia, and there is a carer version.

#### *Clinician-reported measures*

*HoNOS*. The most used outcomes measures in secondary mental health services are the Health of the Nation Outcome Scales (HoNOS). These 12 scales for adult patients with severe mental illness score from 0–4 and cover symptoms, functioning, social relationships and environmental issues. They are rated using all available information – not as a questionnaire or interview – based on the worst state in the reference period. This is usually two weeks. HoNOS are technically plural; whether the 12 scales can sensibly be regarded as separate or as items in a single scale is contentious (Williams et al., 2014). Hereafter, we follow custom using the singular. Variants are available for different settings and client groups (Table 1). In the UK, HoNOS – either in the generic working-age or HoNOS-secure form – is the only truly mandated (i.e. with financial penalties) outcomes measure for adults. Data on HoNOS scale score change over every ‘spell’ of secondary mental care in England are reported to a central body, which makes aggregated

data available to commissioners and providers. There are plans to make these data available publicly. HoNOS65+ is for the moment tolerated in some older adult services, and HONOSCA is mandated only for CAMHS inpatient units. Because of its widespread use, and also its commandeering ‘payment by results’, issues in using HoNOS have a significant bearing on the development of RCOM in the UK, some of which affect many other measures in use or development. We discuss HoNOS issues of implementation, training, feedback and psychometrics below.

#### *Treatment Outcomes Profile (addictions)*

From 2001 all addiction services in England were managed through a national agency, which in 2006 commissioned the development of a 20-item CROM – the Treatment Outcomes Profile (TOP) (Table 1) which is now returned at key points in patient care as part of the contract for all substance misuse services to a national database (Marsden et al., 2009). These data have been used to explore different methods of assessing ‘reliable’ change including that of Jacobson and Truax (1991) in over 18,000 substance abuse patients (Marsden et al., 2011).

#### *Carer-reported measures*

Outside CAMHS there are no carer-reported outcomes measures in routine use in mental health services in the UK. This is undeniably a cause for concern.

### **RCOM assessment procedures and completion rates**

Unless it is retrospective, no outcomes measure is of any value unless it is completed at least twice. All PROMS in routine use are designed to be completed by the patient with no special help; some for psychological therapy outcomes are returned by post by patients waiting for treatment. Those used for severe mental illness, however, often require assistance from the clinical team; in the pilot study of SWEMWBS (Table 1) described above, of the 25% completed questionnaires more than half required some help, mostly from staff. This may be a source of bias when PROMs are repeated. Another issue is that clinicians uncertain or fearful about RCOM may select patients to complete PROMs and introduce bias. CORE-OM is completed at least twice, pre- and post-therapy, and data from the CORE National Research Database for Primary Care indicate that approximately 40% of 64,610 patients had useable paired ratings at the end of therapy (Bewick et al., 2006). CORE-OM questionnaires can be analysed manually, entered on

a PC or via the Internet and can be licensed for use in electronic patient record systems. Some systems allow outcomes charts to be printed for use in therapy. Like all outcomes measures, aggregated data cannot be easily achieved without use of electronic systems, and integrated data entry with clinical systems allows the proper involvement of other key context variables in any sensible analysis, but often at the cost of data useful in the individual clinical setting. HoNOS data are now compulsorily returned centrally as part of the Mental Health Clustering Tool (which is not itself an outcomes measure) from all secondary mental health providers, so all electronic patient record systems in the UK allow clinicians to enter HoNOS ratings directly, often without recourse to paper forms. Centrally mandated standards of reporting intervals are lower than those already previously adopted for RCOM by pioneer services. In the latter a typical regime would involve a rating at the start of an episode of treatment with any given team, at the end, and every 6 months in between if applicable. Some providers still have only administrative systems, in which clerical staff enter data from forms completed by clinicians, a practice which is dying out. Paired HoNOS completion rates for spells are not yet reported from the centre, and no benchmarking data are available. Rates of HoNOS65 + paired ratings for completed episodes have been reported as around 40% (Macdonald & Trauer, 2010), although within services there are teams achieving 80% from our experience (unpublished). Feedback of CROMs data to clinicians is in its infancy in the UK. Even in one of the pioneering services in which such activity is explicitly funded, regular active feedback presentations occur only for a few months before being interrupted by staff changes, information technology (IT) problems or reorganization of services. Data presented to teams shows change in individual HoNOS scale scores with context (age, diagnosis, gender, ethnicity) and compares these with data from functionally similar teams elsewhere in the service. This stimulates debate and discussion – often requests for extra information e.g. from sub-group analysis. This process is described in one of its earlier incarnations by one of us (Macdonald, 2002). At present the feedback of HoNOS and HoNOS65 + is complemented by similar feedback of CGAS and CORE data (see Table 1) in the same secondary mental health service, but it is one of only very few that currently fund such crucial activity.

### **Barriers to and facilitators of RCOM**

Here we review these in relation to the UK experience; many relevant forces can operate both as barriers and facilitators at the same time.

### *Cost of RCOM*

A sensible RCOM system involves the release and training of all clinical staff in the use of CROMs (and also in issuing and helping patients with PROMS), data entry, extraction – together with context data – and analysis, construction and delivery of feedback sessions and reports, and development, e.g. trapping intervention and other relevant data. Investment in RCOM may not have yields in terms of reflective practice until iterated – perhaps for years, and in terms of actual improved outcomes (as opposed to gamed ones) even longer. This is a major disincentive for most organizations, and in the UK only a few secondary services have taken any but a few steps along this long road.

### *Independence of individual NHS organizations*

Within changing financial constraints and, as we see above, somewhat disorganized health policy shifts, primary and secondary care NHS organizations are relatively free to invest in RCOM – as opposed to yet more process monitoring – as they see fit. Outside IAPT most primary mental healthcare services choose not to invest in RCOM. Until the recent advent of ‘payment by results’ and obligatory reporting of HoNOS this applied to secondary mental health services. There are still big differences in RCOM development between UK services responding to this ‘top-down’ pressure and those who had already embarked on RCOM because they wanted to. Even in these a major factor operating against RCOM is frequent reorganization (Oxman et al., 2005) of local teams; for reflective practice in a team to use aggregated outcomes data the team must remain stable in terms of membership and clientele for at least a couple of years, a criterion decreasingly met in the UK.

### *Influence of state-funded mental health services*

It might be thought that services almost entirely funded by taxes would welcome RCOM simply on the basis of ‘value’ (Porter & Teisberg, 2006) but this has only been the case in the UK in the case of the huge investment in IAPT described above. UK government ministries are relatively independent of each other in the approach they take to the services they fund and their demand for outcomes information. One ministry (e.g. Health) can proceed hesitantly and incoherently with RCOM whilst another (e.g. Education) can force a top-down outcomes system with scant consultation. We, under the former ministry, perhaps wishing for more coherent direction from above can usefully glance towards the latter. In the 1990s the UK Department of Education became

rabid about routine educational testing, eventually pushing through dramatic reforms in the teeth of professional opposition (Whetton, 2009). However, the summative and evaluative functions of routine testing became dominant, including published league tables, leading some to suggest that UK state education has, to a greater or lesser extent, been replaced by schooling for test-passing. Without comparison between scores at the beginning and end of a period of intervention, league tables are meaningless (William, 1992). Goodhart's Law (briefly, any measure that becomes a target ceases to be a measure (Goodhart, 1981) was also ignored so that the outcomes of education in the UK in the 21st century remain essentially unknown. Despite huge investment, the possibilities of routine outcomes measurement in terms of reflective professional practice and assessment of novel teaching methods were lost. Educational development in the UK continues to proceed by whim (ministerial rather than professional), when it could so much better proceed by the sort of systematic professional approach to which we aspire in RCOM.

*The semantic halo (Lewis, 1967) of the word 'outcomes'*

For many outside any health service it would seem axiomatic that they would be interested in the results of their ministrations and the fact that this is clearly not the case, even in orthopaedics where the 'end result idea' originated (Kaska & Weinstein, 1998), is a mystery beyond the reach of this paper. In the UK, however, the word 'outcomes' has recently been heard at every level of government (Macdonald, 2014) and, as ever imperfectly translated into action or actual resources, is helping drive the RCOM process forwards, as is the word 'value' (Porter & Teisberg, 2006). As described by these authors, value is defined as 'outcomes that are important to the patient' divided by cost. The simplicity and validity of the numerator is perhaps misleading (Long, 1997), especially in mental health. Despite this 'top-down' interest, a caricature of the position in the UK would have the word 'outcomes' heard either by mental health clinicians operating outside an evidence-based framework (e.g. counsellors or psychoanalysts, and many nurses) as reductionist bean-counting twaddle, or by many from within this ethos (e.g. cognitive-behavioural therapists and psychiatrists) as a tiresome and probably intellectually unsound necessity, riddled with bias and unreliability; a sop to those paying for services. Behind many objections lies the fear that ineffectiveness will be revealed or inappropriately deduced, services for the most vulnerable shut down and jobs lost. As in education, objections are manifest as practical ones – for instance about the time taken

to enter data or attend feedback sessions. In the UK this fear has not been generally assuaged by contrary evidence (e.g. Macdonald & Trauer, 2010), but is less obvious in services that have set up feedback systems. The present move to suck up outcomes data to central repositories without local feedback and under the moniker 'payment by results' is likely to exacerbate this issue.

*Carrots and sticks*

Experience in one of the pioneer secondary mental health services suggested that regular feedback and increasing sophistication of RCOM analysis reached a plateau of implementation that needed to be complemented by managerial force. The Chief Executive of the service mentioned in the previous sentence therefore reviewed RCOM progress in each part of the service on a regular basis with relevant senior managers, and expressed his displeasure at signs of backsliding. In terms of the trajectory of rating frequency this had no discernible impact. However, when commissioners, external to the service altogether, later set targets for completion of paired ratings with financial implications, a strong positive effect was seen, yet not sustained when these targets were later replaced by others unrelated to RCOM. This seemed to emphasize the view that, in the long term, feedback and use by teams of RCOM data will determine its long-term success as much as managerial or financial pressure. However, the recent infatuation with 'payment by results' – even if 'results' here have nothing to do with outcomes – has had a dramatic effect in enhancing the initial recording of RCOM measures but to the detriment of later ones, and thus outcomes. It has introduced a powerful incentive for services to 'game' outcomes in the same way as has already been suggested is happening with mortality data (Taylor, 2013) (despite threats to criminalize this activity in the NHS).

Despite evidence supporting feedback to patients of their own PROM outcomes data in psychotherapeutic settings (Slade et al., 2008), only in some psychotherapy services in the UK does this occur, and this has yet to be reported in secondary care. It would seem likely that combining data from PROMs and CROMs might helpfully inform the clinical conversation in many settings, and would clearly improve data quality of the latter.

*Status of information systems*

As we suggested, the emergence of IT in the late 20th century made RCOM – at least the analysis of aggregated data – a possibility, and the later development of electronic patient records (EPR) and paperless

services allowed it to become a sustainable reality. However, the history of IT in UK health systems has not been edifying. Most mental health EPR systems in the UK contain good mechanisms for collecting context data such as age, sex, diagnosis, etc., rudimentary mechanisms for collecting outcomes measures, and no reliable method of collecting and classifying interventions – medical or other. Extraction of data for the genesis of feedback has been given no priority in system design at all. The dataset used by CYP IAPT is not yet mandated so is supported by only a few systems.

#### *Independent UK outcomes interest groups*

Founded in 2002, the Child Outcomes Research Consortium (CORC) sought to help develop local RCOM, train staff in the use of approved measures, gather outcomes and context data, analyse them and feed information back to participating CAMHS (Fugard et al., 2014; Fleming et al., 2014). CORC is the main driver of RCOM in English CAMHS. It was originally a collaboration between five NHS services: Bedfordshire & Luton, Enfield, Barnet & Haringey, Hertfordshire, Leeds, and Tavistock & Portman. In 2004 membership was opened to all who wished to join – members pay subscription costs and in return receive support, e.g. training, analysis of outcomes, and access to IT systems for data collection. By 2014 there were 76 members, most of which are NHS service providers in the UK (mostly England). There are also non-NHS providers and services from outside the UK, e.g. Norway and Sweden. CORC initially used what is now referred to as a ‘snapshot’ approach to RCOM; data were collected at the beginning of interventions and 6 months later. This was later extended to 6-monthly snapshots to monitor outcomes from young people receiving longer periods of care. CORC’s focus is on PROMs, notably the SDQ but also includes CROMs such as the CGAS and HoNOSCA (see Table 1 for details of these measures). With CORC came the first appreciation that data feedback to their source was crucial to the success of RCOM, and their work has become increasingly influential in central government – it is now difficult to imagine how policy could be developed without them, although their advice is not by any means always heeded.

In 2009 a group of NHS trusts and individuals formed a complementary group for adult mental health – the United Kingdom Routine Clinical Outcomes Measurement in Mental Health Network. This informal group has since met two or three times a year to share experience, agree minimal standards for RCOM collection intervals, give and receive advice and support in developing RCOM for adults, discuss outcomes policy and implementation, and

latterly to influence the Government agenda, at least in England. Whilst issuing no definitive documents several of the key principles of successful RCOM described above have emerged from this informal group, and which are clearly communicated to, if not always welcomed by, central government.

#### **Future directions**

Given the fine balance in the UK between drivers of and obstacles to RCOM, we can anticipate only slow progress here. Central to this are the much greater development of routine feedback of outcomes data to clinicians and their use in clinician/service-user contacts. Of the three dimensions of outcome (health status change, context and intervention) necessary for meaningful feedback, usable data on the latter are almost entirely absent in most UK services. New techniques of natural language processing of electronic clinical records (Wu et al., 2013) are now being applied to the extraction of interventions, thus reducing the burden of extra data-gathering on clinicians or service users. With increasing accuracy it is possible to scan digital clinical notes and letters and extract data on the use of medications and formal psychotherapeutic interventions. There is also a vision that clinical change data themselves (improvement or deterioration) might one day be reliably garnered directly from records without the use of questionnaires. On the other hand, the misguided appetite for commissioning and purchasing healthcare outcomes themselves (£ per HoNOS point improvement) under the flag of ‘value’ (Porter & Teisberg, 2006) will, if thoughtlessly indulged, almost certainly lead to the same fate for RCOM in the UK as has befallen education, in which case we will re-enter the RCOM dark ages here again.

#### **Acknowledgements**

Thanks to Elaine Murphy for guidance on RCOM in the 19th century.

**Declaration of interest:** Alastair Macdonald is a member of NHS England’s Quality and Outcomes Group for Mental Health Currency Development. He is a co-founder of UKRCOM; views of some of its members are represented in this paper. He holds the copyright for HoNOS65+. He is supported by the South London & Maudsley NHS Foundation Trust implementing HoNOS and CORE-OM but has no other affiliations or relevant remuneration. Andy Fugard previously worked for the Child Outcomes Research Consortium (CORC) and helped develop the outcomes evaluation framework for the Children and Young People’s Improving Access to

Psychological Therapies (CYP IAPT) programme. He has no other affiliations or relevant remuneration. The authors alone are responsible for the content and writing of the paper.

## References

- Barkham, M., Bewick, B., Mullin, T., Gilbody, S., Connell, J., Cahill, J., ... Evans, C. (2013). The CORE-10: A short measure of psychological distress for routine use in the psychological therapies. *Counselling and Psychotherapy Research*, 13, 3–13.
- Barkham, M., Stiles, W.B., Connell, J., & Mellor-Clark, J. (2012). Psychological treatment outcomes in routine NHS services: What do we mean by treatment effectiveness? *Psychology and Psychotherapy*, 85, 1–16.
- Bewick, B., Trusler, K., Mullin, T., Grant, S., & Mothersole, G. (2006). Routine outcome measurement completion rates of the CORE-OM in primary care psychological therapies and counselling. *Counselling and Psychotherapy Research*, 6, 33–40.
- Burns, A., Beevor, A., Lelliott, P., Wing, J., Blakey, A., Orrell, M., ... Hadden, S. (1999). Health of the Nation Outcome Scales for elderly people (HoNOS65+). *British Journal of Psychiatry*, 174, 424–427.
- Chorpita, B.F., Yim, L., Moffitt, C., Umemoto, L.A., & Francis, S.E. (2000). Assessment of symptoms of DSM-IV anxiety and depression in children: A revised child anxiety and depression scale. *Behaviour Research and Therapy*, 38(8), 835–855.
- Clark, D.M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: The IAPT experience. *International Review of Psychiatry*, 23(4), 318–327.
- Clark, D.M., Layard, R., Smithies, R., Richards, D.A., Suckling, R., & Wright, B. (2009). Improving access to psychological therapy: Initial evaluation of two UK demonstration sites. *Behaviour Research and Therapy*, 47(11), 910–920.
- Connell, J., Barkham, M., Stiles, W.B., Twigg, E., Singleton, N., Evans, O., & Miles, J.N. (2007). Distribution of CORE-OM scores in a general population, clinical cut-off points and comparison with the CIS-R. *British Journal of Psychiatry*, 190, 69–74.
- Crawford, M.J., Robotham, D., Thana, L., Patterson, S., Weaver, T., Barber, R., ... Rose, D. (2011). Selecting outcome measures in mental health: The views of service users. *Journal of Mental Health*, 20(4), 336–346.
- Deakin, N., & Bhugra, D. (2012). State of psychiatric services in the UK. *International Review of Psychiatry*, 24(4), 379–384.
- Dickens, G., Sugarman, P., & Walker, L. (2007). HoNOS-secure: A reliable outcome measure for users of secure and forensic mental health services. *Journal of Forensic Psychiatry and Psychology*, 18(4), 507–514.
- Donabedian, A. (1966). Evaluating the quality of medical care. *Milbank Memorial Fund Quarterly*, 44, 166–206.
- Ebesutani, C., Bernstein, A., Nakamura, B.J., Chorpita, B.F., & Weisz, J.R. (2010). A psychometric analysis of the revised child anxiety and depression scale – Parent version in a clinical sample. *Journal of Abnormal Child Psychology*, 38(2), 249–260.
- Endicott, J., Spitzer, R.L., Fleiss, J.L., & Cohen, J. (1976). The global assessment scale. A procedure for measuring overall severity of psychiatric disturbance. *Archives of General Psychiatry*, 33(6), 766–771.
- EuroQol Group. (1990). EuroQol – A new facility for the measurement of health-related quality of life. *Health Policy*, 16(3), 199–208.
- Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., & Audin, K. (2002). Towards a standardised brief outcome measure: Psychometric properties and utility of the CORE-OM. *British Journal of Psychiatry*, 180, 51–60.
- Fay, D., Carr, A., O'Reilly, K., Cahill, P., Dooley, B., Guerin, S., Stratton, P. (2013). Irish norms for the SCORE-15 and 28 from a national telephone survey. *Journal of Family Therapy*, 35, 24–42.
- Fleming, I., Jones, M., Bradley, J., & Wolpert, M. (2014). Learning from a learning collaboration: The CORC approach to combining research, evaluation and practice in child mental health. *Administration and Policy in Mental Health and Mental Health Services Research*. doi:10.1007/s10488-014-0592-y
- Ford, T., Hutchings, J., Bywater, T., Goodman, A., & Goodman, R. (2009). Strengths and Difficulties Questionnaire Added Value Scores: Evaluating effectiveness in child mental health interventions. *British Journal of Psychiatry*, 194(6), 552–558.
- Fugard, A.J.B., Stapley, E.J., Ford, T., Law, D., Wolpert, M., & York, A. (2014). Analysing and reporting UK CAMHS outcomes?: an application of funnel plots. *Child and Adolescent Mental Health*, doi: 10.1111/camh.12086
- Gilbody, S.M., House, A.O., & Sheldon, T.A. (2002). Psychiatrists in the UK do not use outcomes measures. National survey. *British Journal of Psychiatry*, 180, 101–103.
- Goodhart, C. (1981). Problems of monetary management: The UK experience. In A.S. Courakis (Ed.), *Inflation, Depression, and Economic Policy in the West* (pp. 111–146). Lanham, MD: Rowman & Littlefield.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38(5), 581–586.
- Goodman, R. (1999). The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. *Journal of Child Psychology and Psychiatry*, 40(5), 791–799.
- Gowers, S.G., Harrington, R.C., Whitton, A., Lelliott, P., Beevor, A., Wing, J., & Jezzard, R. (1999). Brief scale for measuring the outcomes of emotional and behavioural disorders in children. Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA). *British Journal of Psychiatry*, 174, 413–416.
- Health and Social Care Information Centre UK. (2013). Hospital Episode Statistics, Admitted Patient Care, England – 2012–13. Retrieved from <http://www.hscic.gov.uk/catalogue/PUB12566>
- Hood, W.C. (1862). *Statistics of Insanity; Embracing a Report of Bethlehem Hospital from 1846 to 1860, Inclusive*. London: David Batten.
- Jacobson, N.S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12–19.
- Jewell, T., Carr, A., Stratton, P., Lask, J., & Eisler, I. (2013). Development of a children's version of the SCORE Index of Family Function and Change. *Family Process*, 52(4), 673–684.
- Kaska, S.C., & Weinstein, J.N. (1998). Historical perspective. Ernest Amory Codman, 1869–1940. A pioneer of evidence-based medicine: The end result idea. *Spine*, 23(5), 629–633.
- Killaspy, H., White, S., Taylor, T.L., & King, M. (2012). Psychometric properties of the Mental Health Recovery Star. *British Journal of Psychiatry*, 201, 65–70.
- King, M., Nazareth, I., Levy, G., Walker, C., Morris, R., Weich, S., ... Torres-Gonzalez, F. (2008). Prevalence of common mental disorders in general practice attendees across Europe. *British Journal of Psychiatry*, 192(5), 362–367.
- Kroenke, K., Spitzer, R.L., & Williams, J.B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9), 606–613.
- Law, H., Neil, S.T., Dunn, G., & Morrison, A.P. (2014). Psychometric properties of the questionnaire about the process of recovery (QPR). *Schizophrenia Research*, 156(2–3), 184–189.

- Layard, R., Clark, D., Knapp, M., & Mayraz, G. (2007). Cost-benefit analysis of psychological therapy. *Natl Inst Econ Rev*, 23(202), 90–98.
- Lewis, C.S. (1967). Life. In *Studies in Words* (pp. 269–305, 2nd ed.). London: Cambridge University Press.
- Long, A. (1997). Key issues in outcomes measurement. *International Journal of STD and AIDS*, 8, 663–667.
- Long, A., & Jefferson, J. (1999). The significance of outcomes within European health sector reforms: Towards the development of an outcomes culture. *International Journal of Public Administration*, 22(3), 385–424.
- Lundh, A., Kowalski, J., Sundberg, C.J., Gumpert, C., & Landen, M. (2010). Children's Global Assessment Scale (CGAS) in a naturalistic clinical setting: Inter-rater reliability and comparison with expert ratings. *Psychiatry Research*, 177, 206–210.
- Macdonald, A.J. (2002). The usefulness of aggregate routine clinical outcomes data: The example of HoNOS65+. *Journal of Mental Health*, 11(6), 645–646.
- Macdonald, A.J., & Trauer, T. (2010). Objections to routine clinical outcomes measurement in mental health services: Any evidence so far? *Journal of Mental Health*, 19(6), 517–522.
- Macdonald, A. (2014). Commissioning for better health outcomes. *Commissioning Monthly*, 1(6), 14–18.
- MacKeith, J., & Burns, S. (2008). *Mental Health Recovery Star*. London: Mental Health Providers Forum and Triangle Consulting.
- Marsden, J., Eastwood, B., Bradbury, C., Dale-Perera, A., Farrell, M., Hammond, P., ... Wright, C. (2009). Effectiveness of community treatments for heroin and crack cocaine addiction in England: A prospective, in-treatment cohort study. *Lancet*, 374(9697), 1262–1270.
- Marsden, J., Eastwood, B., Wright, C., Bradbury, C., Knight, J., & Hammond, P. (2011). How best to measure change in evaluations of treatment for substance use disorder. *Addiction*, 106(2), 294–302.
- Marsden, J., Farrell, M., Bradbury, C., Dale-Perera, A., Eastwood, B., Roxburgh, M., & Taylor, S. (2008). Development of the Treatment Outcomes Profile. *Addiction*, 103(9), 1450–1460.
- Miller, S.D., Duncan, B.L., Sorrell, R., & Brown, G.S. (2005). The partners for change outcome management system. *Journal of Clinical Psychology*, 61(2), 199–208.
- Oxman, A.D., Sackett, D.L., Chalmers, I., & Prescott, T.E. (2005). A surrealist mega-analysis of reorganization theories. *Journal of the Royal Society of Medicine*, 98(12), 563–568.
- Porter, M.E., & Teisberg, E.O. (2006). *Redefining health care: Creating value-based competition on results*. Boston: Harvard Business School Press.
- Priebe, S., Golden, E., McCabe, R., & Reininghaus, U. (2012). Patient-reported outcome data generated in a clinical intervention in community mental health care – Psychometric properties. *BMC Psychiatry*, 12, 113.
- Priebe, S., McCabe, R., Bullenkamp, J., Hansson, L., Lauber, C., Martinez-Leal, R., ... Wright, D.J. (2007). Structured patient–clinician communication and 1-year outcome in community mental healthcare: Cluster randomised controlled trial. *British Journal of Psychiatry*, 191, 420–426.
- Rotheray, S., Racey, D., Rodgers, L., McGilloway, S., Berry, V., & Ford, T. (2014). Innovations in practice: Further evidence on the effectiveness of the strengths and difficulties added value score as an outcome measure for child and adolescent services. *Child and Adolescent Mental Health*, 19(4), 270–273.
- Rowen, D., Mulhern, B., Banerjee, S., Hout, B., Young, T.A., Knapp, M., ... Brazier, J.E. (2012). Estimating preference-based single index measures for dementia using DEMQOL and DEMQOL-Proxy. *Value in Health*, 15(2), 346–356.
- Roy, A., Matthews, H., Clifford, P., Fowler, V., & Martin, D.M. (2002). Health of the Nation Outcome Scales for People with Learning Disabilities (HoNOS-LD). *British Journal of Psychiatry*, 180, 61–66.
- Secretary of State for Health. (1992). *Health of the Nation: A Strategy for Health in England*. London: HMSO.
- Shaffer, D., Gould, M.S., Brasic, J., Ambrosini, P., Fisher, P., Bird, H., & Aluwahlia, S. (1983). A children's global assessment scale (CGAS). *Archives of General Psychiatry*, 40(11), 1228–1231.
- Slade, K., Lambert, M.J., Harmon, S.C., Smart, D.W., & Bailey, R. (2008). Improving psychotherapy outcome: The use of immediate electronic feedback and revised clinical support tools. *Clinical Psychology and Psychotherapy*, 15(5), 287–303.
- Smith, S.C., Lamping, D.L., Banerjee, S., Harwood, R., Foley, B., Smith, P., ... Knapp, M. (2005). Measurement of health-related quality of life for people with dementia: Development of a new instrument (DEMQOL) and an evaluation of current methodology. *Health Technology Assessment*, 9(10), 1–93; iii–iv.
- Spitzer, R.L., Kroenke, K., Williams, J.B., & Lowe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine*, 166(10), 1092–1097.
- Stewart-Brown, S., Tennant, A., Tennant, R., Platt, S., Parkinson, J., & Weich, S. (2009). Internal construct validity of the Warwick–Edinburgh Mental Well-being Scale (WEMWBS): A Rasch analysis using data from the Scottish Health Education Population Survey. *Health and Quality of Life Outcomes*, 7, 15.
- Taylor, P. (2013). Rigging the death rate. *London Review of Books*, 35(7), 12–15.
- Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., ... Stewart-Brown, S. (2007). The Warwick–Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. *Health and Quality of Life Outcomes*, 5, 63.
- Thurnham, J. (1845). *Observations and Essays on the Statistics of Insanity including an Enquiry into the Causes Influencing the Results of Treatment in Establishments for the Insane to which are added the Statistics of the Retreat near York*. London: Simpkin & Marshall.
- Timbs, J. (1868). *Curiosities of London: Exhibiting the Most Rare and Remarkable Objects of Interest in the Metropolis, with Nearly Sixty Years' Personal Recollections*. London: Longmans, Green, Reader, and Dyer.
- Timimi, S., Tetley, D., Burgoine, W., & Walker, G. (2013). Outcome Orientated Child and Adolescent Mental Health Services (OO-CAMHS): A whole service model. *Clinical Child Psychology and Psychiatry*, 18(2), 169–184.
- Tuke, D. H. (1862). A Table which exhibits the Comparative Statistics of various Asylums in Great Britain, Holland, France, Germany, and Austria. In W.C. Hood (Ed.), *Statistics of insanity: embracing a report of Bethlem Hospital from 1846 to 1860 inclusive* (p. 6). London: David Batten.
- Twigg, E., Barkham, M., Bewick, B., Mulhern, B., Connell, J., & Cooper, M. (2014). The Young Person's CORE: Development of a brief outcome measure for young people. *Counselling and Psychotherapy Research*, 9(3), 160–168.
- Upshur, R., & Tracy, C. (2004). Legitimacy, authority, and hierarchy: Critical challenges for evidence-based medicine. *Brief Treatment and Crisis Intervention*, 43(3), 197–204.
- Whetton, C. (2009). A brief history of a testing time: National curriculum assessment in England 1989–2008. *Educational Research*, 51(2), 137–159.
- William, D. (1992). Value-added attacks: Technical issues in reporting national curriculum assessments. *British Educational Research Journal*, 18(4), 329–341.
- Williams, B., Speak, B., Hay, P., & Muncer, S.J. (2014). An evaluation of the independence of the Health of the Nation Outcome Scales. *Australasian Psychiatry*, 22(5), 473–475.
- Wing, J.K., Beevor, A.S., Curtis, R.H., Park, S.B., Hadden, S., & Burns, A. (1998). Health of the Nation Outcome Scales

- (HoNOS). Research and development. *British Journal of Psychiatry*, 172, 11–18.
- Wolpert, M., Fugard, A.J.B., Deighton, J., & Görzig, A. (2012). Routine outcomes monitoring as part of Children and Young People's Improving Access to Psychological Therapies (CYP IAPT) – Improving care or unhelpful burden? *Child and Adolescent Mental Health*, 17(3), 129–130.
- Woolgar, M (2014). The Brief Parental Self-efficacy Scale. Personal communication.
- Wu, C.Y., Chang, C.K., Robson, D., Jackson, R., Chen, S.J., Hayes, R.D., & Stewart, R. (2013). Evaluation of smoking status identification using electronic health records and open-text information in a large mental health case register. *PLoS One*, 8(9), e74262.