

# Quality-of-life measurement in depression trials: A consumerist relic

**Susan McPherson**

University of Essex, UK

**Jeppé Oute**

University of South-Eastern Norway, Norway

**Ewen Speed** 

University of Essex, UK

Health

1–17

© The Author(s) 2022



Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/13634593221074887

[journals.sagepub.com/home/hea](https://journals.sagepub.com/home/hea)



## Abstract

Quality-of-life measurement in depression is advocated as a patient-centred indicator of recovery, but may instead enhance the mimetic authority of randomised controlled trials (RCTs) which have been roundly critiqued in mental health. In this paper we draw on the social life of methods approach to extend the well-developed critique of RCTs into the field of quality-of-life measurement. We accomplish this through consideration and critique of the conceptual and epistemological development of quality-of-life measurement in depression, including the role of psychometrics in its development. Examining conceptual developments from the 1970s onwards, we consider how the scientific literature on quality-of-life in depression aligns with behavioural economics and consumerism but falls short of engaging with genuinely patient-centred approaches to recovery. We argue that quality-of-life measures in depression were developed within a consumerist model of healthcare in which the medical model was a central pillar and ‘choice’ a rhetorical device only. While quality-of-life instrument development was largely funded by industry, psychometrics provided no coherent solution to the ‘affective fallacy’ (high correlations between quality-of-life and depressive symptoms). Industry has largely abandoned the measures, while psychotherapy research has increasingly endorsed them. We argue that in their design and implementation, quality-of-life measures for depression remain based on a commercial model of healthcare, are conceptually flawed and do not support concepts of patient-centred healthcare.

## Keywords

issues in research methodology, mental health, quality of life

---

### Corresponding author:

Susan McPherson, School of Health and Social Care, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK.

Email: [smcpher@essex.ac.uk](mailto:smcpher@essex.ac.uk)

## Introduction

Randomised Controlled Trials (RCTs) as a ‘gold standard’ form of evidence has been comprehensively critiqued on numerous epistemological grounds in general and specifically in relation to mental health and depression – see for example Luyten et al. (2006). Despite this critique, recently, in challenging the national UK depression guideline, UK psychological therapy researchers and professionals called on the body responsible for the guidelines (the National Institute for Clinical Excellence; NICE) to foreground quality-of-life outcomes in the guideline (Thornton, 2018). The argument is that quality-of-life represents a more patient-centred view of outcome measurement and that this patient-centredness should be reflected in systematic reviews underpinning guideline recommendations.

In this paper we examine the conceptual and epistemological origins and development of quality-of-life as an outcome measure for depression. We consider whether foregrounding this as an outcome domain over symptom outcomes in depression trials would genuinely improve the patient-centredness of guidelines or merely result in increasing the mimetic authority of RCTs. Much of the legitimacy claimed by quality-of-life measurement in depression rests upon appeals to the scientific authority of RCTs. Yet, in examining the many failed attempts to dislodge the hegemony of RCTs in psychiatry, McGoey suggests that critics have engaged in a discursive process which unwittingly reifies RCTs:

*If a practitioner does want to challenge the authority of an individual RCT, or the reliability of individual treatment guidelines, he or she must have recourse to corroborative evidence to support his or her dissent, preferably in the form of more RCTs. It is the very methodological weaknesses of RCTs that imbues them with the authority they hold: for to deny the reliability of a particular study, one must reach for more data, more studies, larger RCTs, in order to justify the validity of one’s objections (McGoey, 2010: 71).*

This ‘mimetic authority’ is compared to Power’s argument that ‘failed audits tend to produce calls for more audits, rather than for reconsiderations of auditing systems in general’ (McGoey, 2010: 72). McGoey’s case study is the 2008 Maudsley debate in which Irving Kirsch uses data obtained through Freedom of Information requests to argue that the true effect size of published and unpublished trials of antidepressants is considerably smaller than the figure reported in published literature. Rather than providing a devastating challenge to the hegemony of RCTs and antidepressant efficacy, McGoey argues this served only to reify the paradigm by appearing to accept effect sizes derived from RCTs as a useful metric. As such, the hegemonic orthodoxy of RCTs, despite some notable challenges (see, e.g. Deaton and Cartwright, 2018) endures as a self-perpetuating tautology.

In examining the value of quality-of-life outcomes in depression trials, we will argue that psychometric measures of depression symptoms and quality-of-life have a shared heritage: a scientific approach concerned with measuring internal states with precision as though they had weight or mass. Psychometrics has provided a transformative technology for social science to mimic hard science; for industry to promote psychotropic medicines; and for healthcare rationing in the guise of Evidence Based Medicine (EBM) to

flourish. With this in mind, can RCTs really be reformed and made more patient-centred by measuring quality-of-life instead of symptoms?

In this paper, we combine arguments concerning the reification of RCTs, commercial interests in the measurement of quality-of-life in depression, critiques of psychometrics, the rise of consumerism and neoliberalism and the more recent ‘happiness’ agenda. We begin by examining the popularisation of quality-of-life as a medical concept in the 1970s (Post, 2014), coupled with happiness and wellbeing. We then examine the development of quality-of-life measurement in depression in published scientific literature during the 1990s and beyond. We consider how the scientific literature defines and applies quality-of-life to depression outcome research; and how the development of the construct has aligned with discourses of behavioural economics and emerged from within consumerist discourse of choice. We examine the application of psychometric technologies to manage the problematic overlap of concepts of quality-of-life and symptoms in depression. In attending to the role of psychometrics in the conceptual development of the field, we draw on the social life of methods approach (Savage, 2013). We consider the claims that quality-of-life embraces subjectivity and show how the literature fails to interact genuinely with patient-centredness.

Our analysis of the field sits within a wider social science critique of the enterprise of measurement and notions of reducing knowledge to that which is measurable. Whooley, for example, has examined the attempt to create metrics for those psychiatric entities presented in the most recent edition of the Diagnostic and Statistical Manual of Mental Disorders, revealing the ‘extra-scientific, non-empirical issues, which are then written into the metrics themselves’ (Whooley, 2016: 34). More broadly, Muller (2018), drawing on Hayekian critiques of logical positivism, has reviewed scholarly work on metrics and metrification in health, education and beyond, referring to an all-encompassing ‘metric fixation’ or ‘tyranny of metrics’ across numerous domains of public life. In this article, we focus specifically on the psychometric development of quality-of-life measures used in depression trials and how the available statistical techniques may have influenced the way in which the concept of quality-of-life in depression has taken shape. In this context, we aim to reveal that focussing on quality-of-life measures in depression trials is unlikely to provide a solution to (and may even exacerbate) the concerns psychological professions and researchers have about the over-reliance on RCTs within EBM and their influence in directing rationing and insurance restrictions.

## **Popularisation of ‘quality-of-life’**

The 1970s has been described in various disciplines as heralding a new turn to individualism. The American writer Tom Wolfe is said to have coined the term ‘Me Decade’ referring to the 1970s, attributing this new turn towards self-expression and search for self-fulfilment to a hangover from post-war prosperity and materialism, leaving many in the West wanting something more in order to feel satisfied with life (McNamara, 2005). Wolfe refers to this period as the ‘third great awakening’, likening the rise in self-interest to a secular version of a religious awakening (McNamara, 2005). The same decade delivered Thatcher, Reagan and the rise of neoliberal government in the West; some commentators have argued that the rise of neoliberal economics and

individualism were intrinsically linked (Boltanski and Chiapello, 2005). This is the socio-political context within which quality-of-life emerges as a construct of interest in scientific literature.

In this context, the ‘Easterlin paradox’ proposed in 1974 that happiness was linked to wealth within state boundaries but that richer countries were not happier overall than poorer countries. This observation and the debates and counter arguments that followed were the foundations for a science of ‘happiness’ (Tomlinson and Kelly, 2013), much of which turned a subjective emotional state into a bounded construct that could be measured. Around the same time, other related constructs became more prominent in the scientific literature including quality-of-life, life satisfaction and wellbeing, which all reflected a concern for the subjective inner states of individuals, that is, a concern with the potential of the self. In expressing concern for the lack of clarity and definition among these new related concepts, Moons et al. (2006) suggest that the flourishing of interest in quality-of-life (which they see as the umbrella term) in the 1970s derived from an increase in life expectancy for those living with chronic conditions as well as an increasing availability of medical technologies. With so many different treatments to choose from, there needed to be new ways to decide among and ration these and so simple mortality indicators were no longer sufficient; quality-of-life could provide a secondary fine grain way to prioritise treatments. Similarly, Armstrong and Caldwell (2004) consider the emergence of quality-of-life as a response to dilemmas of social progress (such as over-population and pollution); technological successes in fields such as cancer and renal medicine; and people living with chronic illness where mortality was an inadequate indicator of medical success.

This notion of somehow defining and then quantifying wellbeing in relation to economic policy and rationing was not novel and while it fits well with neoliberal ideals, it might be considered to have had earlier roots in Bentham’s utilitarianism which proposed a utility scale of 0–1. Utilitarianism came well before any science of psychometrics was established, while more contemporary neoliberal approaches to rationing benefitted (or sprang from) advanced psychometric technologies. The maturation of psychometrics has made possible the flourishing of a range of new inner-state related constructs and their precise measurement. The establishment of psychometrics as a legitimate science is marked by the founding of the Psychometric Society in 1935 in Chicago (Jones and Thissen, 2006). However, it also has earlier roots in late 19th century developments in anthropology which saw the emergence of laboratory methods for measuring and quantifying a multitude of different human physical features and comparison of individuals (Jones and Thissen, 2006).

It is conceivable that these early developments in psychometrics lay the foundation of possibility for a ‘self’ to move into the scientific gaze in the 1970s – a self that could be examined, assessed and interventions to improve the self, designed and evaluated. Google Ngrams indicate that the terms ‘quality-of-life’, ‘self-report’ and ‘construct validity’ all begin to flourish in published literature in the 1970s, suggesting some shared contingency between the developing tools of psychometrics and the idea that there are inner states which can be classified, labelled and measured by asking people to answer questions about their own feelings.

## Interwoven histories of depression and happiness metrics

Having considered how quality-of-life as a construct emerged alongside happiness and wellbeing, we now consider how the measurement of antonymic mood states (depression and happiness) have separate histories which collided at certain key points in time. These points of collision provide a context for the emergence of quality-of-life as something that could be subject to metrification specifically in people with depression.

Early depression questionnaires which emerged in the 20th century required researchers to ask questions and record answers, in ways that enabled them to maintain control over what were ostensibly unreliable or contradictory responses; however, statistical techniques designed to mitigate responder subjectivity and unreliability helped make self-report scales of depression possible (McPherson and Armstrong, 2021). Cronbach's test of internal validity, for example, was first published in 1951 and other techniques relating to 'construct validity' emerged in 1955, consolidated around 1980 (Jones and Thissen, 2006). These techniques enabled ways of ensuring that a questionnaire could be considered to be measuring an actual internal entity or construct rather than a diffuse set of individual views, opinions or unrelated feelings.

The 1960s saw the introduction of a range of self-report questionnaires which allowed the patient to directly answer questions about their mood such as the Beck Depression Inventory. With a proliferation of questionnaires to measure inner subjective states, these could be organised into separate entities, labelled, quantified and used in a range of fields including psychology, psychiatry, healthcare rationing as well as state economics, laying the foundation for happiness, quality-of-life, wellbeing and satisfaction to come to the fore in the 1970s. Important for the manufacturers of the leading treatment for depression at the time (drug companies), these measures were statistically crafted in order to be 'sensitive to change' which meant the items were selected to ensure the measure would show improvement over a short course of antidepressant treatment (McPherson and Armstrong, 2021).

As lexical antonyms, depression and happiness would perhaps inevitably find their paths interweaving in the timeline of metrification. Techniques developing in psychiatric research were gradually adopted into the fields of economics and policy formation where the science of happiness had been fermenting. Frawley (2015: 63) identifies the key moment for the science of happiness as Seligman's 1999 statement on positive psychology as a 'science of. . . seeking to improve quality-of-life'. This is thought to have led to an escalation of scientific approaches to happiness which informed political developments such as the introduction across various Western nations of happiness or wellbeing indexes as an alternative to indicators of economic growth. This process of marrying what became known as behavioural economics to national policy has come under heavy criticism from social scientists on the grounds of scientism and conceptual incoherence (Frawley, 2015).

A key point where the science of happiness and depression measurement overlapped was in the UK in 2006. Here behavioural economics and depression outcomes research collided in the form of The Depression Report (Layard et al., 2006), which led to the establishment of a new national psychological therapy service. The logic was that an evidence synthesis of trials in the recently published national guideline for depression

found Cognitive Behavioural Therapy (CBT) to be the most helpful treatment for depression; and that therefore a national CBT service would provide economic effectiveness by enabling depressed people to recover, go back to work and stop claiming benefits. Within the report were tenets of classic utilitarianism in that if CBT enabled people to recover, relieving them of ‘suffering’, then providing it at scale could deliver national wealth in the form of life satisfaction while also preserving monetary forms of wealth.

*The good news is that we now have evidence-based psychological therapies that can lift at least a half of those affected out of their depression or their chronic fear. . . . Only one in four of those who suffer from depression or chronic anxiety is receiving any kind of treatment. The rest continue to suffer, even though at least half of them could be cured at a cost of no more than £750. This is a waste of people’s lives. It is also costing a lot of money (Layard et al., 2006: 1).*

This collision between happiness and depression research was perhaps inevitable given the mutual contingency of a range of contributing factors in the UK, including (but not limited to) the principles of EBM, the institution of NICE in 1998 and advancing behavioural economics. All of these developments adopted the same set of tools provided by psychometrics. The case for this national CBT service in the UK required acceptance of the hegemonic paradigm of EBM and its unproblematic application to depression, which in turn relied on acceptance of a diagnostic and medical approach to classification of mental disorders. Yet, while happiness research has been labelled by critics as ‘scientism’, the idea of measuring quality-of-life (which shares a common heritage) is now being put forward as a means of addressing the epistemological problems with using RCTs in depression outcome research to inform guidelines and the rationing of treatments (e.g., see Rost and McPherson, 2018). This muddle reinforces the idea of RCTs as a ‘triumph of flawed experiments’ (McGoey, 2010).

## **Consumerist model: What patients want**

Another important element that frames much of these developments is the role of the consumerist model in healthcare which regards patients as consumers and healthcare as a commodity. Whilst there is often assumed to be conceptual overlap between patient-centred care and consumerism in their advocacy of patient choice and decision making, ‘patient-centred care never saw health as a commodity that could be bought and sold, dependent on the response to consumer choice for survival. . . [thus] “patient-centred care” may have been used cynically by political and commercial institutions to persuade patients and consumers to want what is good for the institution’ (Latimer et al., 2017: 2). With this in mind, we consider the potentially cynical role of pharmaceutical companies in the development of quality-of-life measures for depression.

In the 1980s, pharmaceutical companies had been pressing for a relaxation of FDA rules in the US that prevented them marketing prescription drugs directly to patients (Wilkes et al., 2000). This followed an increasing trend towards consumerism in healthcare sometimes couched in terms of giving patients more choice and control while also offering the industry opportunities to increase profits (Adeoye and Bozic, 2007). This is part of the picture of capitalist realism described by Fisher as emerging in the late 1980s

(Fisher, 2009). The 1990s saw an exponential rise in direct-to-consumer advertising of drugs with mounting pressure eventually seeing the FDA relax the rules in 1997 (Adeoye and Bozic, 2007).

It is within this context that the first questionnaire tool developed specifically for depression is published in a peer-reviewed journal: the Quality-of-life in Depression Scale (QLDS) (Hunt and McKenna, 1992). The QLDS is presented as a measure which treats the patient as a consumer whose views should be taken into account in determining the value of treatments. The work was funded by Lilly Industries Limited and the authors employed by Galen Research, a UK contract research organisation. Over the next twelve years the group published a series of peer-reviewed papers describing the ongoing validation of the measure in different countries and languages including the Netherlands (Tuyman-Qua et al., 1997), Belgium (Grégoire et al., 1994), Spain (Cervera-Enguix et al., 1999) and Norway (Berle and McKenna, 2004). By 2001 the authors claimed to have created nine different language versions (McKenna et al., 2001). The authors pitch the QLDS in consumerist terms as a measure to capture subjective patient views as opposed to objective symptoms.

*. . . the acceptability of the treatment to the patient and its impact on her life must be established. A growing emphasis on the views and opinions of the 'consumers' of health services has become apparent recently. . . . Evaluation of those elements of a patient's life which transcend clinical parameters has come to be known as 'quality-of-life' measurement (Hunt and McKenna, 1992: 308).*

The problem the authors seemed to be addressing was that depression RCTs using symptom outcome measures were beginning to reveal flaws as a business model. Patients were not adhering to medication, rejecting the instructions of their doctor and becoming anxious about taking toxic substances. The solution they proposed was to develop a tool which would demonstrate how depression treatments provided patients with what they really wanted. The rhetoric appears to put patients' views first, above those of doctors, offering choice and control.

*The issue of who is the best judge of the efficacy of treatment has also come under scrutiny. Several studies have shown that there is often disagreement between doctor and patient on outcome. The doctor may be satisfied that the patient has improved but the patient may feel worse than before. Conversely, the patient may feel herself to be cured, whilst the physician remains unconvinced that this is the case (Hunt and McKenna, 1992: 310).*

Throughout the series of validation studies, the authors emphasise that measure development was based on a theory about the needs of the patient and an understanding that 'life gains its quality from the ability and capacity of the individual to satisfy certain human needs' (Hunt and McKenna, 1992: 312). There is an emphasis on the development phase which involved interviewing a group of patients with experience of depression 'to explore the impact of the symptoms of depression on the patients' ability to fulfil their needs and to examine the effects of medication on the patients and their illness' (p. 313). The QLDS appears to have been designed with a view to being able to reflect the extent to which treatments met the needs of patients. It seems possible that this development

was part of a wider move within the pharmaceutical industry to make a more direct alliance with patients, by-passing professionals. Developing outcome measures which could reflect what patients want could enable a new means of communicating the benefits of antidepressants to patients in a future envisaged by industry in which they would eventually be allowed to directly advertise to patients. Other measures that were used in depression trials from the 1990s onwards were also developed with pharmaceutical funding including the Quality-of-life Enjoyment and Satisfaction Questionnaire (QLESQ; Endicott et al., 1993), this time funded by Pfizer; and the Social Adaptation Self-Evaluation Scale (SASS) funded by Upjohn (Bosc et al., 1997). It has also been common for depression trials to use measures of ‘functioning’, like the General Assessment of Functioning; or measures of subjective health such as the SF36 developed by the RAND Corporation; and to refer to these as quality-of-life measures.

However, neither the UK nor Europe followed the FDA in changing direct-to-consumer advertising laws, which may have contributed to the apparent gradual abandonment by drug companies of quality-of-life measures in depression trials. Another explanation for this abandonment and failure to report quality-of-life data in clinical study reports put forward by Sharma (2018) may be that the data collected by pharmaceutical companies reveals that antidepressants negatively impact quality-of-life, although this is speculation informed by the authors’ difficulties in obtaining information from the pharmaceutical companies they contacted.

## **Construct validity and the affective fallacy**

Having considered the role of pharmaceutical companies in developing quality-of-life measures, we show how measure developers turned to psychometric techniques to try to demonstrate ‘construct validity’ but ran into difficulties when confronted with the problem that depressive ‘symptoms’ (as constructed by psychiatry) tended to overlap with ‘quality-of-life’ (as constructed by measure developers). The practice of defining a new psychological construct and then designing a questionnaire tool to measure it had, by the 1990s, become a common activity among quantitative researchers in psychological science, particularly those concerned with evaluating treatments. In spite of statistical sophistication that had developed in psychometrics, the key problem facing any psychometric project was a philosophical one: to demonstrate the actuality of the construct being measured. This meant that the thing being measured had to be clearly different from another thing or construct that had already been established as being scalable through psychometrics (‘discriminant validity’). The tool that measured this new thing must also perform as well as another tool designed to measure the same thing (‘convergent validity’). A tool that met both these criteria could be deemed to have construct validity: it measures an actual thing in full without accidentally measuring part or all of something else at the same time. The idea of singling out quality-of-life in depression as a thing in its own right that was separate from both quality-of-life in general and from depression symptoms was logically and conceptually ambitious not least because neither depressive symptoms nor quality-of-life had been reliably established as discrete entities.

Defining quality-of-life as a discrete entity has never been fully realised in the general nor in the specific sense. It has been described as an umbrella term for ‘functioning,



health status, perceptions, life conditions, behaviour, happiness, lifestyle, symptoms' (Moons et al., 2006: 892). In other contexts these terms have been used interchangeably or quality-of-life has come under the umbrella of one of these other terms. The QLDS developers, for example, refer to quality-of-life variously as a form of wellbeing; as encompassing 'psycho-social elements which are not normally accessible to the doctor' and elsewhere as the 'perceptions and preferences of the patient' (Hunt and McKenna, 1992: 309–310).

Some authors have attempted to differentiate functioning from quality-of-life in depression whereas others explicitly use measures of 'functioning' to measure quality-of-life. The QLDS authors frequently differentiate quality-of-life from physical functioning, as do Mazumdar et al. (1996) in describing a study using the General Life Satisfaction Scale for depressed older people. In the latter study, authors depict functioning as having been the focus of 'health-related' quality-of-life measures as opposed to actual quality-of-life. This distinction variously differentiates activities patients might engage in such as employment or leisure from patient satisfaction with those activities. Mazumdar et al. (1996) later list social function, disability, physical health and quality-of-life as aspects of functioning; yet at another point operationalise quality-of-life as a combination of wellbeing and coping which together represent 'positive functioning'.

In looking at the potential of the EQ5D and the SF6D (considered in other contexts to be measures of 'subjective health') to evaluate quality-of-life in depression, Sobocki et al. (2007) note that 'impaired quality-of-life denotes functional limitations and perceived difficulties in everyday life caused by a disease or illness' (p. 153). In practice, the EQ5D has items covering mobility, self-care, physical/social function, role, pain, mental health and vitality. The SF36, SASS, QLESQ and other common measures used in depression trials all vary in terms of overlapping concepts and subscales. This confusion does not only reflect a lack of coherence or consensus that has been much remarked upon, but poses fundamental epistemological difficulties for establishing a psychometric approach to quality-of-life in depression.

Another common problem in the literature is the attempt to differentiate depressive symptoms from quality-of-life in depression, which often gets tangled up in a further narrative around quality-of-life being subjective and symptoms being objective. The QLDS, for example, is introduced in the context of a clear separation between 'objective' symptoms and subjective wellbeing. Yet symptoms in any area of psychiatry cannot be objectively measured, being as they are mediated through patient perception, followed by self-report, observation, clinician or researcher judgement or all of these. To justify why some items are classed as a symptom (e.g. lack of sleep) rather than a subjective experience, authors claim that objective symptoms 'influence' quality-of-life but are not quality-of-life (Tuynman-Qua et al., 1997).

Further illustrating this conceptual problem, depression symptom scales had by the 1990s become the go-to quality-of-life measure for studies in cancer, hypertension and several other physical diseases (Bech, 1996). This is not surprising given that depression 'symptoms' are, when looking at questionnaire items, things that overlap with the things that tend to be classed as quality-of-life when referring to physical illness (mood, sleep, worry, fatigue and so on). In order then to measure quality-of-life in depression, it would be necessary to establish some components of quality-of-life that could be established as

different to symptoms (as manifest in depression measures) and so be a thing with construct validity. This would remain consistently as part of the folklore of quality-of-life measurement in depression, with authors regularly attempting to disprove or prove the problem of measure redundancy or the ‘affective fallacy’:

*QOL does not appear to be an epiphenomenon of mood state. If the affective fallacy was operating, to the extent that it rendered the measure redundant, one would expect a substantially larger variance in QOL to be accounted for by mood state* (Swan et al., 2009: 494).

*The results indicate that QoL and depressive symptoms are two different constructs, and thus QoL could be assessed as an additional treatment outcome* (Kolovos et al., 2016: 466).

The decision as to whether a correlation between a quality-of-life measure and a depression severity measure represented good discriminant validity or good convergent validity appears arbitrary. In testing the validity of the QLDS, a correlation with depression severity of 0.81 was announced as a good indicator of validity as some correlation was expected ‘given that both measures are dependent on the same factor, “objective level of depression”’ (McKenna and Hunt, 1992: 328). Grégoire et al. (1994) found correlations ranging from 0.39 to 0.76 depending on timepoint and also concluded this was satisfactory:

*. . . we have to stress that the two measures we are discussing cannot be considered to be the two sides of the same coin. Indeed, the two evaluations are overlapping, partly but not totally* (Grégoire et al., 1994: 18).

In an examination of the usefulness of the WHOQOL in depression, the authors conclude that a correlation with depression severity of 0.59 is highly significant, raising a concern about ‘measure redundancy’, although also suggesting the correlation occurs because the sample is clinically depressed:

*. . . highly significant correlation between depression severity and QOL may simply reflect the characteristics of this clinical sample. In a non-clinical sample or a sample of medically ill non-depressed subjects in which depression scores are lower and normally distributed, the correlation between depression severity and QOL may be different* (Naumann and Byrne, 2004: 168).

This lack of consensus around appropriate statistical cut-offs is not uncommon in developing fields of metrics and statistics, but of particular interest are the explanations authors give for the  $r$  value they find as though the statistic speaks for itself and the conclusion of there being two separate entities (or not), self-evident from the statistic. This may be accounted for by the financial sponsorship of most of these measure validation studies, since, in order for quality-of-life measures to be a useful marketing tool for industry, they needed to be manufactured in such a way that they could produce positive results while also appearing to be different from symptom measures, more patient-focussed, more concerned with individual needs. The measures had to be both the same as and different to symptom measures: ‘partly but not totally’ overlapping. Any correlation

it seemed, could be interpreted as reflecting a good enough mid-point between convergent and divergent validity.

## Limiting subjectivity through metrics

As noted earlier, quality-of-life measures have been developed within a consumerist framework, with an appearance of being concerned with patient-centred outcomes, but without sufficient engagement with identifying and verifying those outcomes. In this regard, tools have been developed which claim to represent something about patients' subjective experience as opposed to their symptomatology. We pose a question here as to whether the apparent primacy of patient subjectivity, coupled to claims of patient-centredness, may account for claims that quality-of-life measures are more indicative of the clinical aims of psychotherapies than other therapeutic approaches. To some extent, the items on these questionnaires provide some face-validity for this idea. Questionnaires like the QLESQ address satisfaction with home life, work, leisure, education and so on, rather than focus on reductionist, behaviourist and context-free items such as hours of sleep, frequency of crying, appetite and so on. Yet subjectivity in quality-of-life measures is confined to a tightly defined and delimited set of experiences that patients are allowed report on, belying the concern with subjective experience (i.e. subjective experience cannot be tightly delimited). This is inherent in the closed-response questionnaire format of quality-of-life measures. It is also reflected in the way in which measures have been developed and the way in which patient experience is separated from the clinical gaze, suggestive of an underlying assumption that a truly subjective understanding of patient experience may well be in opposition to the interests of the clinician.

For example, in the QLDS literature, quality-of-life is often set up as relating to 'perceptions and preferences of the patient' as opposed to those of the clinician (Hunt and McKenna, 1992: 309). The patient is deemed to be the best judge of treatment efficacy because the doctor is uninterested in or unable to know about psycho-social elements of the patient's life. At the same time authors reveal a lack of trust in the patient's view noting that quality-of-life 'is highly influenced by depressed mood' (Tuyman-Qua et al., 1997: 200) (i.e. something that can be clinically measured), while another group comments that 'subjective reports of QOL may be influenced by psychiatric symptoms' (Naumann and Byrne, 2004: 160). This trail of logic suggests that neither clinicians nor patients seem to be able to truly know whether their life has been improved or not by a treatment because clinicians are not patients and psychiatric patients cannot know their own mind.

Similarly, patients cannot be trusted to complete the questionnaires correctly in reporting on their subjective experience. The QLDS group conducted patient interviews in developing the measure. Patients were invited to comment on the draft questionnaire and although they indicated that the yes/no response format would force them to make stark choices which may not reflect their more nuanced experience, the developers admitted that they kept this format because patients would otherwise tend towards the middle choice (Hunt and McKenna, 1992). This limited faith in patient subjectivity was also evident in the process of translation. Native speakers of the language with experience of depression were invited to be part of the development process but their role was 'to

consider the translations provided by the previous panel and decide whether they were acceptable' (Cervera-Enguix et al., 1999: 393). Lay members could comment on the wording to ensure it made sense but they could not alter the underlying meaning of the English version or adapt any concepts. Subjectivity appears to have been screened out all the while appearing to be screened in. In the US and Canadian translations the item 'I am reluctant to answer the door or telephone' was considered inappropriate to people living in areas of high crime so was replaced with 'I just want people to leave me alone' (McKenna et al., 2001). Thus, socio-cultural issues which might have a direct impact on one's psychological wellbeing (crime) are seen as confounding variables which have to be reinterpreted in order to measure the true feeling, as though the feeling itself is not socially contingent. This highlights the problem of claiming that a measure of quality-of-life in depression measures a universal and unidimensional construct or thing when that thing is understood to be influenced by cultural and value systems. A thing that is acknowledged as being subjective cannot logically be the subject of a questionnaire which demonstrates construct validity. It cannot be universal or unidimensional by virtue of it being subjective.

## **QALYs, death and the perfect life**

Finally, we consider the use of quality-of-life measures for the purposes of rationing healthcare and the epistemological problems with applying this approach to depression in particular. The utilitarian heritage of quality-of-life in depression is seen at its most stark in the notion of Quality Adjusted Life Years (QALYs). As in Bentham's metric, QALYs work on a scale of 0–1 in which 0 is equivalent to death and 1 represents a perfectly lived life. Economist Alan Maynard has been described as the pioneer of health economics, having established the field in the 1970s at York University (Cooxson et al., 2016). In 1997, as EBM was being adopted into mainstream government policy making, Maynard declared:

*... the clinician who supports an evidence-based approach would argue that scarce resources be allocated on the basis of the interests of the individual patient and efficacy. By contrast, the economist or public-health physician would contend that scarce health-care resources be allocated according to the interests of society as a whole (the population-health ethic) and on the basis of efficiency (Maynard, 1997: 126).*

NICE was established in 1998 to address UK population health and to remove variations in quality of care and so in essence was married to the stance of the health economist. NICE explicitly includes health economic modelling using QALYs in its procedures as do several other neoliberal welfare states. To determine QALYs, a particular type of quality-of-life measure is required, designed to produce a 'utility' score between 0 and 1. The EQ5D is the most commonly used scale for this. 'However, utility estimates in purely depressed, UK patient populations derived from the EQ5D or SF6D have not hitherto been available. This is surprising, given the important role of bodies such as NICE and the consequences of the rulings made upon their commissioned technology assessments' (Mann et al., 2009: 570). NICE produced a draft update depression guideline in 2018 which identifies five studies on which to base utility estimates for depression

(National Institute for Health and Care Excellence, 2018). Of these five, NICE selected the EQ5D scores from Sobocki et al. (2007) for several of their economic models. The Sobocki study was funded by Lundbeck and used the EQ5D to find out if the quality-of-life of depressed people prescribed antidepressants in Sweden improved over 6 months.

*To calculate patient preference utility weights (index scores) from the answers to the EQ-5D instrument, population-based social tariffs are usually employed, that is, tariffs based on health state valuations in general population samples . . . In the absence of specific social tariffs for Sweden at the time of conducting this study, the EQ-5D index tariffs derived by Dolan et al. were employed (Sobocki et al., 2007: 154).*

In other words, at the point at which some form of meaning or value (social value judgement) is introduced into the method which might be culturally sensitive in some way (however crude), there has been limited basis for doing so within the national context (UK guidelines for depression) and so values are imputed from different cultural contexts. This implies that life holds the same meaning or quality for all; that quality-of-life is a universal and uniform construct, where subjectivity is not relevant, counter to the rhetoric behind quality-of-life metrification. Moreover, the concept of a universal 0–1 scale for this concept collapses when applied to depression. As an extreme example, Sobocki et al. (2007) grappled with the problem that some of their research participants were producing negative values whereas the cut-off is automatically set to 0. This means that in spite of depressed people attempting to express a feeling that living can be worse than death, the statistical method was based on a logic in which nothing could be worse than death. In other words, the utilitarian scale of 0–1 places suicidal people in a blanket category of dead. NICE's use of this study data (as well as its use of data from 4 other studies in other models) was to extract the mean EQ5D scores reported for each symptom severity category, thereby benchmarking utility back to the 'objective' indicator of symptom remission (the CGI), 'the utility value of remission based on the improved or very much improved CGI-I 28 score is likely to express the utility of people in future remission states' (National Institute for Health and Care Excellence, 2018: 746).

Recovery, here used interchangeably with symptom 'remission' was now a state against which quality-of-life was being gauged. This is echoed in other literature where researchers deal with the absence of weights specific to depression. For example Byford (2013) in assessing the EQ5D in adolescents notes that although there are no weights available, 'health per se is a reasonable proxy for the value placed on health'(p. 103). Their study divides the sample into two groups down the middle of each objective indicator (severity, number of suicide attempts, comorbidity and so on) and conducts a Student *T*-test on the groups to establish construct validity (on the premise that the groups should be statistically different if the quality-of-life measure is valid). Here in QALYs, we therefore see the field of quality-of-life moving full circle back to the supposed objectivity of symptoms while clothed in a range of new technological jargon and statistical technique.

## Conclusion

Measurement of quality-of-life in depression trials has been advocated by psychotherapy professions and critics of the medical model of mental health. It has been

argued that quality-of-life measures provide a more patient-centred approach to evaluation and more closely align with what patients value in terms of their personal recovery. In this paper we have argued that quality-of-life measures have a shared heritage with depression symptom measures in their emanating from the technologies of psychometrics. They also have a shared heritage with happiness research and its predecessor utilitarianism which, while seeking to measure the value in life and develop policies to provide the maximum amount of happiness across the population, it applies a 'detectivist' approach to emotional wellbeing. Paradoxically, this 'detectivist' approach to seeking happiness does not fit with philosophical ideas about happiness and self-fulfilment which argue that genuine states of eudamonia are destroyed by attempts to monitor or measure them, since these activities strip away context, meaning and agency (Morgan, 2014).

In its starkest form, these technologies rank people who feel suicidal as already dead, not unlike Soerensen's (2002) portrayal of depressed people as the 'living dead' by virtue of their passivity and dysfunction, the very negation of the 'useful life'. We argue that the quest for valid measures of quality-of-life leads to epistemologically unsound and circular journeys in order to identify a unitary construct that is implausibly both universal and subjective. Rather than being patient-centred, we argue that these attempts to quantify quality-of-life in people with depression derive from consumerist models of depression treatments which employ patient choice as a marketing device. Having failed in this, industry has rolled back from using quality-of-life outcome measures, while psychological therapy professions are advocating them to support a methodological enterprise (RCTs) that they had deemed inappropriate to apply to depression.

In looking to better understand how different treatments impact on patient recovery in depression, it may be necessary to abandon the metric fixation that has dominated psychiatry as well as many other domains of social life in neoliberal states. Genuinely patient-centred and patient-led approaches might, for example, involve researchers and guideline developers working closely with user-led organisations to design and deliver research, enabling more user-led research and enhancing coproduction strategies in research, practice, guideline and policy development. In this vein, Rose et al. (2006) proposed the need for a 'multiple perspectives paradigm' in mental health which would enable service users and carers as well as professionals to have equal roles in all aspects of research design and conduct. The authors (from the Service User Research Enterprise unit at King's College London) give an example of an evidence synthesis in 2003 on ECT led by service user researchers commissioned by the UK Department of Health. Service user researchers identified hidden biases across the research field that were not evident to others in the research team. Although this review was incorporated into NICE good practice guidelines for ECT, this approach is still not common practice in guideline development, perhaps because health services researchers continue to believe naively that EBM can be reformed in some way (see Wieringa et al., 2017). As we have shown here concerning quality-of-life measures in depression trials, reforming an epistemologically unsound paradigm does not offer a way forward to patient-centred care; a way forward is only possible with approaches in which patients are genuinely centred in all aspects of care delivery, research and policy formation.

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## ORCID iDs

Susan McPherson  <https://orcid.org/0000-0002-9478-9932>

Ewen Speed  <https://orcid.org/0000-0002-3850-922X>

## References

- Adeoye S and Bozic KJ (2007) Direct to consumer advertising in healthcare. *Clinical Orthopaedics and Related Research* 457: 96–104.
- Armstrong D and Caldwell D (2004) Origins of the concept of quality of life in health care: A rhetorical solution to a political problem. *Social Theory & Health* 2(4): 361–371.
- Bech P (1996) Quality of life measurements in major depression. *European Psychiatry* 11(3): 123–126.
- Berle JØ and McKenna SP (2004) Quality of Life in Depression Scale (QLDS): Adaptation and evaluation of the psychometric properties of the Norwegian version. *Nordic Journal of Psychiatry* 58(6): 439–446.
- Boltanski L and Chiapello E (2005) The new spirit of capitalism. *International Journal of Politics Culture and Society* 18(3-4): 161–188.
- Bosc M, Dubini A and Polin V (1997) Development and validation of a social functioning scale, the Social Adaptation Self-evaluation Scale. *European Neuropsychopharmacology* 7 Suppl 1(1): S57–70; discussion S71.
- Byford S (2013) The validity and responsiveness of the EQ-5D measure of health-related quality of life in an adolescent population with persistent major depression. *Journal of Mental Health* 22(2): 101–110.
- Cervera-Enguix S, Ramírez N, Giralá N, et al. (1999) The development and validation of a Spanish version of the quality of life in depression scale (QLDS). *European Psychiatry* 14(7): 392–398.
- Cookson R, Goddard M and Sheldon T (eds)(2016) *Maynard Matters: Critical Thinking on Health Policy*. England: University of York, CHE.
- Deaton A and Cartwright N (2018) Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine* 210: 2–21.
- Endicott J, Nee J, Harrison W, et al. (1993) Quality of life enjoyment and satisfaction questionnaire: A new measure. *Psychopharmacology Bulletin* 29(2): 321–326.
- Fisher M (2009) *Capitalist Realism: Is There No Alternative?* Winchester: John Hunt Publishing.
- Frawley A (2015) Happiness research: A review of critiques. *Sociology Compass* 9(1): 62–77.
- Grégoire J, de Leval N, Mesters P, et al. (1994) Validation of the quality of life in depression scale in a population of adult depressive patients aged 60 and above. *Quality of Life Research* 3(1): 13–19.
- Hunt SM and McKenna SP (1992) The QLDS: A scale for the measurement of quality of life in depression. *Health Policy* 22(3): 307–319.

- Jones LV and Thissen D (2006) A history and overview of psychometrics. In: Rao CR and Sinharay S (eds) *Handbook of statistics, 26: Psychometrics*. Amsterdam, The Netherlands: North Holland, pp.1–27.
- Kolovos S, Kleiboer A and Cuijpers P (2016) Effect of psychotherapy for depression on quality of life: Meta-analysis. *The British Journal of Psychiatry* 209(6): 460–468.
- Latimer T, Roscamp J and Papanikitas A (2017) Patient-centredness and consumerism in health-care: An ideological mess. *Journal of the Royal Society of Medicine* 110(11): 425–427.
- Layard R, Clark D, Bell S, et al. (2006) *The depression report: A new deal for depression and anxiety disorders*. London: London School of Economics. Available at: <http://cep.lse.ac.uk/pubs/download/dp0829.pdf>
- Luyten P, Blatt SJ, Van Houdenhove B, et al. (2006) Depression research and treatment: Are we skating to where the puck is going to be? *Clinical Psychology Review* 26(8): 985–999.
- Mann R, Gilbody S and Richards D (2009) Putting the ‘Q’ in depression QALYs: A comparison of utility measurement using EQ-5D and SF-6D health related quality of life measures. *Social Psychiatry and Psychiatric Epidemiology* 44(7): 569–578.
- Maynard A (1997) Evidence-based medicine: An incomplete method for informing treatment choices. *Lancet* 349(9045): 126–128.
- Mazumdar S, Reynolds CF 3rd, Houck PR, et al. (1996) Quality of life in elderly patients with recurrent major depression: A factor analysis of the general life functioning scale. *Psychiatry Research* 63(2–3): 183–190.
- McGoey L (2010) Profitable failure: Antidepressant drugs and the triumph of flawed experiments. *History of the Human Sciences* 23(1): 58–78.
- McKenna SP, Doward LC, Kohlmann T, et al. (2001) International development of the Quality of Life in Depression Scale (QLDS). *Journal of Affective Disorders* 63(1–3): 189–199.
- McKenna SP and Hunt SM (1992) A new measure of quality of life in depression: testing the reliability and construct validity of the QLDS. *Health policy* 22(3): 321–330. [https://doi.org/10.1016/0168-8510\(92\)90005-v](https://doi.org/10.1016/0168-8510(92)90005-v)
- McNamara C (2005) The pursuit of happiness, American style: Tom Wolfe’s study of status and freedom. *Perspectives on Political Science* 34(1): 16–26.
- McPherson S and Armstrong D (2021) Psychometric origins of depression. *History of the Human Sciences*. <https://doi.org/10.1177/09526951211009085>
- Moons P, Budts W and De Geest S (2006) Critique on the conceptualisation of quality of life: A review and evaluation of different conceptual approaches. *International Journal of Nursing Studies* 43(7): 891–901.
- Morgan A (2014) The happiness turn: Axel Honneth, self-reification and ‘sickness unto health’. *Subjectivity* 7(3): 219–233.
- Muller J (2018) *The Tyranny of Metrics*. Princeton, NJ: Princeton University Press.
- National Institute for Health and Care Excellence (2018) *Depression in adults: Treatment and management - Full guideline (consultation draft)*. National Institute for Health and Care Excellence.
- Naumann VJ and Byrne GJ (2004) WHOQOL-BREF as a measure of quality of life in older patients with depression. *International Psychogeriatrics* 16(2): 159–173.
- Post MW (2014) Definitions of quality of life: What has happened and how to move On. *Topics in Spinal Cord Injury Rehabilitation* 20(3): 167–180.
- Rose D, Thornicroft G and Slade M (2006) Who decides what evidence is? Developing a multiple perspectives paradigm in mental health. *Acta Psychiatrica Scandinavica* 113(429): 109–114.
- Rost F and McPherson S (2018) Stakeholder position statement on the NICE guideline for depression in adults. Available at: <https://cdn.ymaws.com/www.psychotherapyresearch>.



- org/resource/group/bc770a1b-ec30-4101-a101-30e12b059ac5/downloads/spruk\_nice\_position\_statemen.pdf
- Savage M (2013) The ‘Social Life of Methods’: A critical introduction. *Theory Culture & Society* 30(4): 3–21.
- Sharma T (2018) Effects of selective serotonin reuptake inhibitors (SSRIs) and serotonin - norepinephrine reuptake inhibitors (SNRIs) on suicidality, violence, and quality of life. *University of Copenhagen*. Available at: <https://ctu.dk/wp-content/uploads/2020/12/2018-Ph.D.-thesis-Tarang-Sharma.pdf>
- Sobocki P, Ekman M, Agren H, et al. (2007) Health-Related quality of life measured with EQ-5D in patients treated for depression in primary care. *Value in Health* 10(2): 153–160.
- Soerensen A (2002) Den biologiske psykiatris renæssance og det antidepressive menneske. *Slagmark – Tidsskrift for idehistorie* 35: 159–183. <https://doi.org/10.7146/sl.v0i35.228>
- Swan A, Watson HJ and Nathan PR (2009) Quality of life in depression: An important outcome measure in an outpatient cognitive-behavioural therapy group programme? *Clinical Psychology & Psychotherapy* 16(6): 485–496.
- Thornton J (2018) Depression in adults: Campaigners and doctors demand full revision of NICE guidance. *BMJ* 361: k2681.
- Tomlinson MW and Kelly GP (2013) Is everybody happy? The politics and measurement of national wellbeing. *Policy and Politics* 41(2): 139–157.
- Tuynman-Qua H, de Jonghe F and McKenna SP (1997) Quality of Life in Depression Scale (QLDS). Development, reliability, validity, responsiveness and application. *European Psychiatry* 12(4): 199–202.
- Whooley O (2016) Measuring mental disorders: The failed commensuration project of DSM-5. *Social Science & Medicine* 166: 33–40.
- Wieringa S, Engebretsen E, Heggen K, et al. (2017) Has evidence-based medicine ever been modern? A labour-inspired understanding of a changing EBM. *Journal of Evaluation in Clinical Practice* 23(5): 964–970.
- Wilkes MS, Bell RA and Kravitz RL (2000) Direct-To-Consumer prescription drug advertising: Trends, impact, and implications. *Health Affairs* 19(2): 110–128.

### Author biographies

Susan McPherson is Professor of Psychology and Sociology in the School of Health and Social Care at the University of Essex with research interests in the intersections of evidence based mental health, social sciences and medical humanities. Jeppe Oute is Associate Professor in mental health care at the Department of Health, Social and Welfare Studies at the University of South-Eastern Norway and works in the field of recovery and critical mental health. Ewen Speed is Professor of Medical Sociology in the School of Health and Social Care at the University of Essex with research interests in state organisation of healthcare and policy implementation processes.