Article:
Aragón, María José, Gravelle, Hugh orcid.org/0000-0002-7753-4233, Castelli, Adriana orcid.org/0000-0002-2546-419X et al. (6 more authors) (2024) Measuring the overall performance of mental healthcare providers. Social Science & Medicine. 116582. ISSN 1873-5347

https://doi.org/10.1016/j.socscimed.2024.116582

Social Science & Medicine 344 (2024) 116582

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

Measuring the overall performance of mental healthcare providers

María José Aragón [a], Hugh Gravelle [b], Adriana Castelli [b], Maria Goddard [b], Nils Gutacker [b], Anne Mason [b], Donna Rowen [c], Russell Mannion [d], Rowena Jacobs [b,*]

[a] HCD Economics, Daresbury Innovation Centre, Keckwick Lane, Daresbury, Warrington, WA4 4FS, UK
[b] Centre for Health Economics, University of York, Heslington York, YO10 5DD, UK
[c] Sheffield Centre for Health and Related Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK
[d] Health Services Management Centre, School of Social Policy, Park House, University of Birmingham, Edgbaston, Birmingham, B15 2RT, UK

ARTICLE INFO

Handling Editor: Joanna Coast

Keywords:
Health benefits
Non-health benefits
Costs
Performance measurement
Mental health providers

ABSTRACT

To date there have been no attempts to construct composite measures of healthcare provider performance which reflect preferences for health and non-health benefits, as well as costs. Health and non-health benefits matter to patients, healthcare providers and the general public. We develop a novel provider performance measurement framework that combines health gain, non-health benefit, and cost and illustrate it with an application to 54 English mental health providers. We apply estimates from a discrete choice experiment eliciting the UK general population's valuation of non-health benefits relative to health gains, to administrative and patient survey data for years 2013–2015 to calculate equivalent health benefit (eHB) for providers. We measure costs as forgone health and quantify the relative performance of providers in terms of equivalent net health benefit (eNHB): the value of the health and non-health benefits minus the forgone benefit equivalent of cost. We compare rankings of providers by eHB, eNHB, and by the rankings produced by the hospital sector regulator. We find that taking account of the non-health benefits in the eNHB measure makes a substantial difference to the evaluation of provider performance. Our study demonstrates that the provider performance evaluation space can be extended beyond measures of health gain and cost, and that this matters for comparison of providers.

1. Introduction

Both regulators of healthcare providers and policymakers need to be able to distinguish good from poor performance. Performance assessment and ranking of providers helps to understand the reasons for variation, identify best practice, tackle poor performance and evaluate the impact of regulatory decisions. A challenge for provider performance assessment is to capture a broad array of performance dimensions that matter, alongside costs.

In the extra-welfarist perspective (Brouwer et al., 2008) the key outcome of healthcare provision is its effect on health. This focus is justified by assuming that the objective of public and private funders of healthcare is to maximise health within a given budget. However, patients may also value other aspects of healthcare. Specifically, they also care about non-health process attributes of healthcare, such as how long they wait for treatment, how easy it is to access care, and how they are treated whilst receiving care (Rowen et al., 2022; Soekhai et al., 2019; Entwistle et al., 2012). Policymakers and funders also care about non-health benefits. For example, the World Health Organisation's health system performance framework takes responsiveness (defined as including respect for persons, prompt attention, quality of basic amenities, and choice of provider) as a fundamental objective in addition to health (Murray and Frenk, 2000). The English National Health Service (NHS) recognises the importance of non-health benefits in its legislation and policy (Department of Health, 2010, Department of Health and Social Care, 2023; NHS, 2017), and even collects data to monitor its performance against these policy aims, such as the Friends and Family Test (https://www.england.nhs.uk/fft/) based on patient assessments of aspects of the care delivery process such as dignity and privacy.

The question is not whether we should measure non-health benefits, but rather how to combine these with data on health and costs to derive an overall assessment of provider performance so that providers can be ranked. In cost-effectiveness analysis, there is support for extending the attributes taken into account in the evaluation of specific healthcare *interventions* beyond the usual health effects and cost which are considered (Coast, 2004; Coast et al., 2008; Devlin and Sussex, 2011;

* Corresponding author.
E-mail address: rowena.jacobs@york.ac.uk (R. Jacobs).

https://doi.org/10.1016/j.socscimed.2024.116582
Received 3 January 2023; Received in revised form 27 December 2023; Accepted 8 January 2024
Available online 12 January 2024

Wildman et al., 2016; Wildman and Wildman, 2019). However, there is no consensus on the best approach to achieve this (Briggs, 2016; Frazão et al., 2018).

A small number of studies have used estimates of patient health benefits and cost to compare healthcare *providers* rather than healthcare interventions. Appleby et al. (2013) used gains in quality-adjusted life years (QALYs) and costs for English hip replacement providers and Karnon et al. (2013) compared survival and costs for patients presenting with acute chest pain in four public hospitals in Australia. Timbie and Normand (2008) and Timbie et al. (2008) compared risk-adjusted costs and survival for heart attack patients in Massachusetts hospitals.

With only two dimensions of performance (health and cost) scatter plots can identify efficient and inefficient (undominated and dominated) providers. Timbie and Normand (2008) and Timbie et al. (2008) extended their estimation of cost and survival by using an assumed monetary value per life saved to calculate incremental *net* health benefits to measure provider performance. Data envelopment analysis and stochastic frontier methods (Jacobs et al., 2006) can be used as the number of performance dimensions increase. But judgements about the relative marginal value of the different performance dimensions, are essential to create a single measure.

Nonetheless, regulators and policymakers find it useful to rank providers by overall performance, which requires a single measure. There appear to have been no previous attempts to construct quantitative measures of overall healthcare provider performance which take account of preferences for non-health benefits, along with health benefits and cost (Ryan and Tompkins, 2014).

In this paper we develop a novel method of incorporating non-health aspects of the healthcare process into quantitative assessments and rankings of healthcare provider performance. We build on the principles within economic appraisals of weighing the costs and benefits of *interventions*, by applying these ideas to assessing the performance of healthcare *providers*. We extend the evaluation space to include aspects of provider responsiveness (such as continuity of care, clear communication, being treated as a person) and access, in addition to measures of health benefit and production costs. We show how stated preference estimates obtained from a discrete choice experiment (DCE) with a sample of the UK general public can be used to measure health and non-health benefits on a common scale – *equivalent health benefit* (eHB) – which can then be combined with information on cost to calculate an *equivalent net health benefit* (eNHB) performance indicator. This can then be used to rank providers.

We illustrate the eNHB approach by applying it to English NHS mental healthcare providers to demonstrate the first quantitative comparison of their overall performance in terms of costs and (health and non-health) benefits. Aspects of care other than direct health gain are particularly important in mental health services given that many mental health conditions are enduring or recurring and healthcare contacts can be for prolonged periods of time. An evidence review highlighted the value placed by mental health service users on relationship building, continuity of care and involvement in care (Newman et al., 2015). Relationships form the core of service users' experiences (Gilburt et al., 2008) and higher patient satisfaction with mental health services is associated with care continuity and involvement (Stamboglis and Jacobs, 2020). Policies also stress the importance of patient centred care and the need to involve service users in care decisions (Mental Health Taskforce, 2016; NHS England, 2016). However, given the complexity of the nature of mental health care, there has been mixed success in developing performance assessment frameworks for these providers (Jacobs and McDaid, 2010; Urbanoski and Inglis, 2019).

In the next section we outline the institutional background in NHS mental healthcare. Section 3 describes the methods we use to measure the overall performance of NHS mental health providers and Section 4 presents results. In section 5 we discuss the key steps and requirements for incorporating non-health benefits into a provider performance assessment framework.

2. Institutional background

In the English NHS, healthcare is funded almost entirely from general taxation. Mental health services differ from physical health services in both organisation and delivery. Although some mental health services are provided in primary care, patients with more serious mental health problems are typically managed by publicly owned mental health providers (known as Mental Health Trusts) which provide care, without charge, in both inpatient and community settings. Patients access the services of one of the 54 Trusts mainly by referral from their primary care doctor, although some are referred by the justice system, social services, and some self-refer. Most Trusts serve the local population in their catchment areas, although some receive national referrals for more specialist services. They often cover large geographic areas and usually have several community and inpatient sites. Trusts are largely funded by block contracts negotiated with local commissioners which hold budgets, for all types of care, derived from a national needs-related capitation formula. Trusts are expected to break even each year.

Mental health disorders account for 4% of hospital admissions in England (QualityWatch, 2015) and mental health treatment costs account for 12% of the total NHS budget (McCrone et al., 2008). The average length of stay for those who are treated in hospital is around 35 days; much longer than the average of around 8 days for patients admitted for physical health conditions (Cavallaro et al., 2023; Bell, 2022). For over 90% of the 1.8 million people in contact with secondary mental health services each year, care is delivered outside hospital by community-based mental health teams run by the local Trust.

3. Methods

In this section we outline how we construct our proposed performance measure (equivalent net health benefit – eNHB). (A fuller description is in the Online Appendix.) The study was approved by NHS North East Newcastle and North Tyneside Research Ethics Committee. The reference number for the study is: 240,427.

3.1. Equivalent net health benefit

Creating a single measure from multiple dimensions of performance requires that all dimensions of performance are valued in the same units. The net health benefit (NHB) framework is routinely used in the assessment of health technologies and interventions (Stinnett and Mullahy, 1998; Drummond et al., 2015) to measure cost-effectiveness. We extend this framework to incorporate non-health aspects of healthcare provision and apply it to performance assessment. Our overall performance measure is the *equivalent net health benefit* (eNHB):

$$eNHB_j = eHB_j - w_c c_j = h_j + \sum_{k=1}^{K} w_k q_{jk} - w_c c_j \quad (1)$$

where h_j is the health benefit of treatment in provider j, q_{jk} is the k-th non-health benefit, w_k is the marginal valuation of this type of benefit in terms of health (i.e. the marginal rate of substitution between q *and* h), c_j is the cost of providing care, and w_c is the opportunity cost of expenditure in terms of units of benefit forgone. eHB is the equivalent health benefit, i.e. the sum of health and non-health benefits valued in units of heath. The way w_k and w_c are obtained is described in sections 3.3 and 3.4 respectively.

3.2. Data to measure benefits and costs

We use two key datasets to measure health and non-health benefits. First, the Mental Health Minimum Data Set (MHMDS) has data on all adults using secondary mental health services funded by the English NHS. MHMDS has information on patients' mental health, socio-demographic characteristics, home address, waiting time from referral to start of treatment, and treatment. Patients are assessed by a

healthcare professional at the start of an episode of care and allocated to one of 20 needs-based categories, known as clusters. Much of our data collection and analyses are at the level of the *cluster-episode*: the time a patient remains within the same cluster in the same provider. If patients change their cluster allocation or change provider, this triggers a new cluster-episode. We use MHMDS to measure health benefit, as well as three patient-level measures of non-health benefit using the methods described below. Second, we use provider-level information on patients' experience of Mental Health Trusts from the annual Community Mental Health Survey (CMHS) for six other measures of non-health benefit such as provider responsiveness.

To calculate the costs of cluster-episodes, we cost the most common types of activity recorded in the MHMDS (see Ride et al., 2020 for details). For patients admitted to hospital, we use the cluster-specific unit costs per admitted patient day. For services provided in the community we use the Electronic Staff Record and estimates of unit costs for 2016 by the Personal Social Services Research Unit (Curtis and Burns, 2016).

After allowing for changes to CMHS and missing data in MHMDS, we measure the performance of all 54 mental health providers using data for years 2013–2015.

3.3. Definition and measurement of health and non-health benefits

In order to identify important dimensions of health and non-health benefits for assessing mental health provider performance, we apply the results of a previous study that used six focus groups (three of mental health service users and three of mental healthcare professionals) (Powell and Rowen, 2022). The dimensions they identified consist of: effectiveness of treatment (h_j) and nine other non-health quality attributes (q_{jk}). The attributes they identified were described in such a way that we could measure them using data in MHMDS and CMHS. We measure four of the ten attributes using data from MHMDS patient level data aggregated to provider level: waiting times, ease of access, appropriate discharge, and effectiveness of treatment (health benefit). The other six dimensions are measured using provider level means of patient responses to questions in the CMHS (see Table 1). We provide a more detailed description of the measurement of the attributes in the Online Appendix.

The health benefit (attribute 10) is calculated using data on the change in the patient's health between the start and end of the cluster-episode. Generic health-related quality of life measures, such as the EQ-5D instrument, do not perform well in mental health patient populations (Brazier et al., 2014). Instead, we use the ReQoL-UI (Recovering Quality of Life Utility Index), a validated measure of the quality of life for people with mental health problems (Keetharuth et al., 2020). MHMDS does not include the ReQoL instrument but does have information on the Health of the Nation Outcome Scale (HoNOS) score at the beginning and end of cluster-episodes where we assume the difference in score represents the health gain from the episode of care. HoNOS is a widely used measure of symptoms and severity for mental health patients with 12 items rated on a 0–4 scale (Wing et al., 1998). Using results from regression analysis of matched ReQoL-UI and HoNOS scores collected for 649 patients in 14 Mental Health Trusts (Keetharuth and Rowen, 2020), we estimate ReQoL-UI as a linear function of the HoNOS score. Data limitations prevented us from estimating the duration of the health benefit, and in the absence of any available evidence, we assume that the benefit of treatment (the health gain) lasts one year after the end of the cluster episode, yielding a measure commensurate with the QALY to which a cost-effectiveness threshold can be applied (see section 3.4).

Comparison of provider quality measures and costs requires risk adjustment for patient-level factors outside the control of providers (Iezzoni, 2012; Moran and Jacobs, 2018). The six trust-level quality measures (attributes 3 to 8) derived from the CMHS are already adjusted for the age and sex mix of respondents for each question in each trust and no further adjustment is applied. We also interpret a failure of appropriate discharge (attribute 9) as a 'never-event' that should not

Table 1
Benefit attributes for mental healthcare.

Attributes	Dataset	Definition	DCE description
1. Waiting times	MHMDS	Days from referral date to cluster-episode start date (only calculated for the first cluster-episode)	The time you wait to receive healthcare is appropriate for your needs
2. Ease of access	MHMDS	Treated in nearest site	The healthcare you receive is provided in your local area
3. Person-centred care	CMHS	% most positive response to 'Were you involved as much as you wanted to be in agreeing what care you will receive?' (Q12)	You are involved as much as you want to be in agreeing what care you receive
4. Co-ordinated approach	CMHS	% most positive response to 'How well does this person organise the care and services you need?' (Q10)	The person or people you see organise the care and services you need well
5. Continuity	CMHS	% most positive response to 'In the last 12 months, have the people you see for your care or services changed?' (Q17)	You are able to see the same person or people throughout your healthcare
6. Communication, capacity and resources	CMHS	Average of % most positive response to 'Did the person or people you saw listen carefully to you?' (Q.4) and 'Were you given enough time to discuss your needs and treatment?' (Q5)	The person or people you see listen carefully to you and give you enough time to discuss your needs and treatment
7. Treated as a person	CMHS	% most positive response to 'Overall in the last 12 months, did you feel that you were treated with respect and dignity by NHS mental health services?' (Q42)	You are treated with dignity and respect
8. Recovery focus	CMHS	% most positive response to 'Do the people you see through NHS mental health services help you with what is important to you?' (Q39)	You are supported to do the things in your life that you want to do
9. Appropriate discharge	MHMDS	Indicator equal to one if there is no new spell starting between seven and 30 days of the end of the previous spell.	You are not discharged before you are ready
10. Effectiveness of treatment (health benefit)	MHMDS	Change in HoNOS score between start and end of cluster-episode converted to change in ReQoL-UI	For your next year of life you will have 20%, 50%, 80% quality of life

Note. DCE: discrete choice experiment. CMHS: Community Mental Health Survey – outcomes reflect rates, i.e. provider-level mean values of binary variables. MHMDS: mental health minimum data set (patient-level). HoNOS: Health of the Nation Outcome Scale. ReQoL UI: Recovering quality of life utility index. CMHS question numbers are from the 2015 and 2016 questionnaires (https://nhssurveys.org/surveys/).

occur, irrespective of the characteristics of the patient. We therefore case-mix adjust three MHMDS based benefit attributes (waiting times, ease of access and effectiveness) and patient cost, using regression models with patient characteristics, including age, gender, ethnicity, marital status, Mental Health Act (detention) status, income deprivation, and whether a care co-ordinator is assigned (typically for those with enduring illness).

Estimates of the marginal rate of substitution, or the valuation of the health (h_j) and non-health attributes (q_{jk}) are taken from a published online discrete choice experiment (DCE) that was undertaken with 1018 members of the UK general population (Rowen et al., 2022). The sample was drawn from the UK general public to elicit preferences as they are voters, tax payers (and, therefore, fund the NHS) and potential users of the NHS. The marginal values (w_k) of the nine non-health attributes in terms of health are the marginal rates of substitution between health (h_j) and non-health attributes (q_{jk}) and are taken from the model estimates using the DCE data (for further details see Rowen et al., 2022).

3.4. Combining costs and benefits as equivalent net health benefit (eNHB)

We use *eNHBj* to compare the relative performance of providers against each other, and not to make judgements about the absolute value of providers (e.g. whether their output has a greater value than their cost). All health expenditure has opportunity costs in the form of other benefits that could have been generated with these resources. Estimates of these opportunity costs can be used to make the cost of mental health care provision commensurate with the measures of health (and non-health) benefits. The National Institute for Health and Care Excellence (NICE) generally recommends new technologies if additional health benefits are produced at less than £20,000-£30,000 per QALY (Devlin and Parkin, 2004; Appleby et al., 2007), but has used a weighting system equivalent to a threshold of £50,000 per QALY for drugs providing relatively short extensions of life in patients defined as end of life (Paulden, 2017). Claxton et al. (2015) estimate the marginal cost to the NHS of producing an additional QALY at a much lower value of around £13,000. None of these valuations explicitly incorporates the likely non-health benefits gained as a result of expending NHS resources. In the absence of estimates of the opportunity costs of mental healthcare expenditure in terms of eHB forgone, we use $w_c = 1/£30{,}000$ in our base case.

3.5. Robustness checks and extensions

We examine the robustness of provider rankings in a number of ways. First, we conduct sensitivity analysis using a range of alternative cost weight values ($w_c = 1/£10{,}000$, $1/£20{,}000$, $1/£50{,}000$).

Second, we conduct sensitivity analysis using a range of alternative attribute weights. Where our main analyses used preferences elicited from the general public to determine the relative weightings of the different attributes, we examine the use of published weights derived from mental health service users and mental healthcare professionals (Rowen et al., 2022). Next, we set the non-health benefit attribute weights to zero so that only health gains and costs matter.

We also compare provider rankings based on equivalent net health benefit with the published rankings of the sector regulator, the Care Quality Commission (CQC). The CQC produces a four-category (*Outstanding*, *Good*, *Requires Improvement* and *Inadequate*) rating of English mental health providers. The rating is based on five criteria (safety, effectiveness, caring, responsiveness to people's needs, leadership) assessed by a combination of routine quantitative data, qualitative investigations, and site visits (Care Quality Commission, 2017; Allen et al., 2020).

The CQC criteria, though not formally defined, differ from those embodied in the eNHB measure and while the CQC measure appears to take account of both health and non-health benefits, it takes no account of cost in comparing providers. Whilst we would therefore not expect to find an equivalence in ratings, it is of interest to see how the results used by regulators which are available in the public domain differ from those using our measure.

4. Results

Provider level summary statistics are shown in Table 2 which presents the descriptive statistics for the case-mix adjusted provider level effects *plus* the overall mean for the quality attributes and costs. The health gain, nine non-health attributes and cost shown in Table 2 are the main inputs to our provider performance analysis. On average, Trusts appear to perform reasonably well on ease of access with over two-thirds of them able to treat all locally resident patients. There are some outliers, with the worst performing Trust treating less than 10% of locally resident patients and the best performing Trust treating almost all of its locally resident patients. Performance on appropriate discharge appears to be good with Trusts discharging only 5% of patients inappropriately, on average.

Overall mean Trust performance on patient reported attributes from CMHS varies considerably across the attributes, with 75% of patients reporting that they are satisfied with being treated as a person but only 44% reporting they were involved as much as they wanted in the care they received. However, for any given CMHS attribute the variation in mean scores across Trusts is relatively small: the coefficients of variation range from 0.03 (treated as a person) to 0.09 (person centred care).

The average Trust level health effect (ReQoL gain) varies across Trusts, ranging from 0.101 to 0.114. (Mean values for provider level quality attributes when taking account of the marginal valuations of each benefit are shown in Appendix Table A.2).

We adjust only for common year trends for the provider effects for the CMHS attributes (3–8) because we use provider level survey data, and for appropriate discharge (attribute 9) because we regard it as a "never event". Consequently these provider effects are highly correlated with the raw provider means (see Appendix Table A.3). Correlations of

Table 2
Descriptive statistics for provider level effects.

Benefit attribute	Weight relative to ReQoL	Provider level summary statistics			
		Mean	SD	Min	Max
1. Waiting time: wait ≤75th centile (62 days)	$w_1 = 0.062$	1.055	0.110	0.750	1.306
2. Ease of access	$w_2 = 0.179$	0.701	0.173	0.096	0.927
3. Person-centred care	$w_3 = 0.090$	0.435	0.039	0.354	0.535
4. Co-ordinated approach	$w_4 = 0.205$	0.606	0.043	0.503	0.689
5. Continuity	$w_5 = 0.121$	0.571	0.046	0.457	0.675
6. Communication, capacity and resources	$w_6 = 0.134$	0.698	0.028	0.641	0.748
7. Treated as a person	$w_7 = 0.176$	0.751	0.023	0.694	0.802
8. Recovery focus	$w_8 = 0.104$	0.423	0.040	0.348	0.500
9. Appropriate discharge	$w_9 = 0.139$	0.948	0.022	0.888	1.002
10. Effectiveness: ReQoL gain	1	0.107	0.002	0.101	0.114
Cost					
Cost (£) per cluster-episode	$w_c = 1/£30{,}000$	13,383	910	11,703	16,043
Equivalent health benefit (eHB)		0.933	0.043	0.841	1.009
Equivalent net health benefit (eNHB)		0.487	0.059	0.320	0.603

Notes: Summary statistics are for provider effects calculated as risk adjusted provider fixed effects *plus* the overall mean for the attribute or cost. Equivalent health benefit: $eHB = h - w_c\ c$. Equivalent net health benefit: $eNHB = h + \sum_{k=1}^{9} w_k q_k - w_c c$. Number of providers: 54. Benefit attributes (except for appropriate discharge) and cost are case-mix risk adjusted as described in the Online Appendix.

risk adjusted and raw measures are much lower for ease of access (0.35), effectiveness (0.55), and cost (0.63) which are derived from patient-level data in MHMDS suggesting that risk adjustment materially affects these patient-level measures. Risk adjustment makes relatively little difference for the waiting time attribute where the correlation of risk adjusted and raw provider scores is 0.96.

Table 3 reports the correlations amongst risk adjusted provider level health, non-health attributes, and cost. The benefit attributes 3 to 8, measured using CMHS data, show strong positive correlations: all but 2 of the 15 correlation coefficients are statistically significant at the 5% level. Some of these attributes may therefore be picking up related aspects of quality. None of the four benefit measures derived from the patient cluster-episode level MHMDS data are statistically significantly correlated with any of the other benefit measures. Cost per cluster-episode is not significantly associated with any of the benefit attributes.

Fig. 1 compares the rank of providers when performance is measured by eNHB which takes account of health, cost and non-health attributes, and the rank of providers when performance is measured by net health benefit (NHB) in which only cost and health benefits matter (i.e. taking no account of non-health benefits). The Spearman correlation between the two rankings is 0.57 and the average absolute change in rank is 10.6 suggesting that including non-health benefits makes a difference to provider rankings. There is less change in ranking at the extremes than in the middle of the rankings.

Table 4 further examines the sensitivity of performance rankings to the weights on health attributes and cost. Varying the weight on cost has a smaller effect on the numerical *magnitude* of the eNHB performance measure (column (1)) than on the rankings by eNHB (column (2)). Even after increasing the cost weight three-fold (£1/30,000 to £1/10,000), the eNHB scores are still strongly correlated (0.93). The effect on the correlation of the *ranking* of providers is somewhat greater with an average absolute change in ranking of 6.5. However, there is little movement in the top and bottom quartiles of the rankings.

Where the main analyses used preferences elicited from the general public to determine the relative weightings of the different attributes, we found that using DCE weights derived from mental health service users or mental healthcare professionals have very little effect on the magnitudes of eNHB or on the rankings.

When we set the weights of the non-health attributes (1–9) to zero so that providers are evaluated by NHB ($h - w_c c$), the average absolute change in ranking is 10.6 and five of the 13 providers in the top quartile are moved into lower quartiles. This marked impact suggests the potential importance of incorporating non-health benefit attributes in the measurement of the performance of mental health service providers.

Fig. 2 compares the CQC four-fold categorisation of providers during the period 2014–2015 with the quartiles of the eNHB performance metric. The CQC rates only two providers as *Outstanding* and these are in the second and third quartiles of the eNHB distribution. The single provider rated as *Inadequate* is in the second eNHB quartile. There is slightly better correspondence between the CQC and eNHB rankings for the providers rated *Good* and *Requires Improvement* by the CQC. Thirteen of the 15 providers rated as *Good* are in the top two eNHB quartiles and 22 of the 36 rated as *Requires Improvement* are in the bottom two quartiles. Comparison of the CQC categorisation with our NHB measure, rather than with eNHB, produces a very similar pattern (see Appendix Table A.4)).

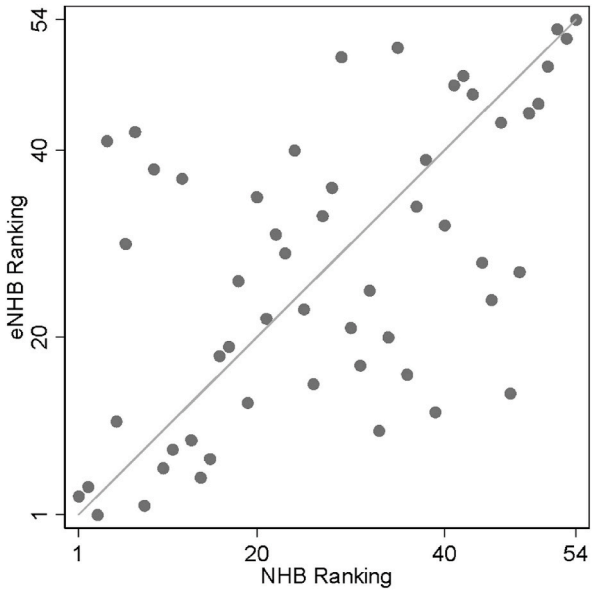

Fig. 1. eNHB vs NHB provider performance ranking
Note. eNHB: equivalent net health benefit: $h_j + \sum_{k=1}^{9} w_k q_{jk} - w_c c_j$. NHB: net health benefit: $h_j - w_c c_j$. Rank 1 is best performance.

In Appendix Table A.5 we report results from a comparison of the CQC categories and eNHB using three ordered logistic regressions of the CQC categories on NHB, on eNHB, and on the 11 components (health gain, non-health attributes, cost) of eNHB. The CQC ranking is weakly ($p < 0.1$) positively associated with eNHB but not with NHB. In the model with all 11 separate components of eNHB there are no statistically significant coefficients and five (including the health gain) are negative.

The differences between the sector regulator rankings and our proposed eNHB performance measure suggests that they are capturing different aspects of performance. Allen et al. (2020) found similarly poor agreement between CQC ratings of acute hospitals and those predicted from a large set of indicators.

Overall, the robustness checks to the eNHB measure suggest that the performance rankings are relatively stable to changes in varying the preferences used for the relative weightings and in applying different weights to the costs. The results also show that including non-health attributes matters for provider rankings. There was however lower agreement between the eNHB and a comparison with CQC ratings but

Table 3
Correlations between health, non-health attributes, and cost.

	1. Waiting time	2. Ease of access	3. Person-centred care	4. Co-ordinated approach	5. Continuity	6. Communication	7. Treated as a person	8. Recovery focus	9. Appropriate discharge	10. Effectiveness (ReQoL gain)
2	0.108	1								
3	0.142	0.079	1							
4	0.193	0.093	0.636***	1						
5	0.183	0.112	0.500***	0.543***	1					
6	0.211	0.122	0.829***	0.724***	0.634***	1				
7	0.167	0.080	0.518***	0.441**	0.367	0.448**	1			
8	0.204	0.039	0.832***	0.765***	0.545***	0.851***	0.410	1		
9	−0.151	−0.129	−0.215	−0.055	−0.253	−0.057	−0.291	−0.023	1	
10	−0.076	−0.117	−0.142	0.024	0.186	−0.022	−0.088	−0.070	0.044	1
Cost	−0.134	−0.332	−0.123	−0.051	0.005	−0.058	−0.120	0.024	0.046	0.178

Note. **, ***: Pearson correlation significant at the 1% and 5% level after applying Bonferroni adjustment.

Table 4
Sensitivity to weights on cost and non-health attributes.

	Correlation[a] with baseline eNHB (1)	Correlation[b] with baseline eNHB ranking (2)	Average absolute change in ranking (3)	Change in top quartile (4)	Change in bottom quartile (5)
Health gain equivalent of cost (w_c)					
w_c = 1/£10 k	0.93	0.82	6.5	2	1
w_c = 1/£20 k	0.99	0.97	2.6	1	1
baseline w_c = 1/£30 k	1	1	0	0	0
w_c = 1/£40 k	1.00	0.99	1.5	1	1
w_c = 1/£50 k	0.99	0.98	2.2	1	2
Weights on non-health attributes					
Service Users	1.00	0.99	1.7	1	1
Healthcare Professionals	0.98	0.96	2.9	1	1
Only health gain and cost matter (NHB): $w_1 = 0, \ldots, w_9 = 0$)	0.72	0.57	10.6	5	4

Note.
[a] Pearson correlation coefficient.
[b] Spearman rank order correlation coefficient. Change in top and bottom quartiles: number of different providers in quartile compared with baseline. The baseline case uses the general public marginal rates of substitution between health gain and other attributes. eNHB: equivalent net health benefit $h_j + \sum_{k=1}^{9} w_k q_{jk} - w_c c_j$. NHB: net health benefit $h_j - w_c c_j$.

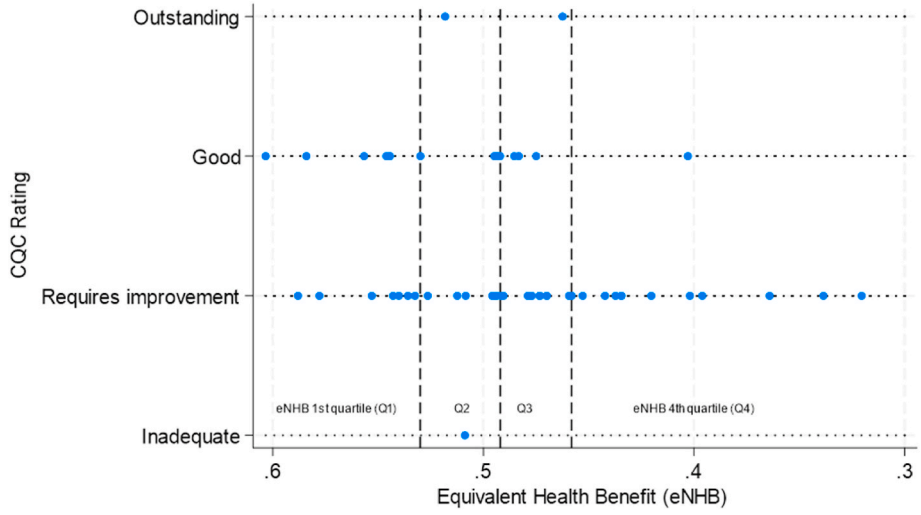

Fig. 2. Comparison of ranking of Care Quality Commission rating with ranking by eNHB quartile.

these performance measures likely focus on different aspects of performance.

5. Discussion

In this paper we demonstrate that it is possible to take account of preferences for health and non-health attributes alongside cost, in measuring the overall performance of healthcare providers. This requires:

- Choice of health and non-health attributes. We assume that the relevant dimensions of healthcare are those that matter to mental healthcare service users and mental healthcare professionals, and use ten existing attributes of mental healthcare quality (health gain and nine non-health attributes) identified in previous studies using focus groups (Powell and Rowen, 2022; Rowen et al., 2022).
- Measurement. Administrative data covers some of these performance dimensions (waiting time, being treated close to home). The English NHS also routinely collects information on the mental health of patients at the start and end of episodes of care. Other dimensions of non-health benefit that matter to service users and mental healthcare professionals, such as being treated as a person and continuity of care, are not routinely captured in the administrative data we use and were drawn instead from patient surveys.
- Valuation of the relevant attributes on a common scale. We apply results from an existing DCE estimating the general public's willingness to trade off non-health benefit dimensions against a measure of health-related quality of life for mental health (Rowen et al., 2022). To express cost in terms of health and non-health benefits forgone (i.e. the opportunity cost of NHS expenditure) we use existing estimates of the cost-effectiveness threshold in the English NHS (Devlin and Parkin, 2004; Appleby et al., 2007) and conduct sensitivity analyses around this value.

We find considerable variation across providers in health and non-health attributes and in cost. Most of the non-health attributes are positively correlated, though the correlations are statistically significant only for those collected via the patient survey. Six of the non-health attributes are negatively correlated with the change in health-related quality of life. This implies that better performance on non-health attributes is not necessarily linked to improvements in health. Providers that produced larger health benefits also had higher costs, but for six of the nine non-health dimensions, costs are lower for providers with better health benefit scores.

Taking account of nine non-health benefits made a substantial

difference to the evaluation of provider performance. The Spearman correlation coefficient of the ranking of providers by eNHB, which allows for health, non-health benefits and cost, with the ranking based only on NHB, which allows only for health and cost, was relatively low at 0.57. Provider performance rankings are much less sensitive to assumptions about the health equivalent weight on cost. For example, after a three-fold increase in the cost weight compared with that based on the NICE threshold, the Spearman correlation coefficient of the provider rankings was 0.82.

The sector regulator – the Care Quality Commission – produces an ordered four category rating of provider performance based on its assessment of safety, effectiveness, caring, responsiveness to people's needs, and leadership. Some of these attributes are based on the same patient survey that we used, though the way in which different performance attributes are combined to produce the CQC rating is not published. We find only weak positive correlation between our eNHB measure and the CQC rating, suggesting that the CQC criteria, or weightings placed on them by the CQC, differ considerably from those we derived from discrete choice experiments undertaken with members of the general population (Rowen et al., 2022). The low correspondence between eNHB and CQC ratings may not be surprising, given the differences in their focus. The choice of performance assessment measures may depend on both the purpose of the assessments and the needs of the audience for whom the approach is designed. For example, the CQC ratings are designed for regulators where system leadership may explicitly be highly valued because the CQC can hold to account those they deem responsible for poor performance in terms of leadership of the organisations they assess and they therefore gather specific qualitative evidence from leaders as part of their assessment. Our approach has a patient-centred focus, taking account of attributes that matter to patients where non-health benefits may be more highly valued.

Our exploratory analyses have some limitations. First, we assume that the duration of the health gain is the same for all clusters and providers (one year). However, if there is variation in the duration of health effects, then provider rankings would be affected. Second, some of the routine administrative and costing data we used are of poor quality, with many missing items and inconsistencies, particularly for the measurement of cost and health gain. Survey data may also suffer from response bias. Third, though the wording used in the DCE was selected to represent the patient survey data, there may have been a weak correspondence between some of the attributes identified as important in focus groups, those valued in the DCE, and the patient survey questions used to measure these attributes. Fourth, we did not explore the estimation of uncertainty around the performance measures derived from the eNHB approach because we want to be as transparent as possible in the development of the method and not introduce additional complexity since both the weighting on different elements and the performance level of different elements are estimated and are subject to uncertainty. Future applications of the eNHB approach may wish to use more sophisticated methods. Fifth, performance is a function of a range of aspects of provider behaviour which need to be combined in some functional form. Whilst a more complicated aggregation is possible, we chose a simple linear model for our DCE weights which is explicit. An alternative to an additive model would make it more difficult to estimate and then apply the weights to. Sixth, there are limitations to the HoNOS mental health outcome measure (Delaffon et al., 2012) and the sample size used to convert HoNOS to ReQoL-UI may be too small to capture potential heterogeneity in mental health needs. Finally, estimates of the opportunity costs of NHS expenditure take account only of displaced health. They may be higher or lower than the opportunity cost in terms of health and non-health benefit (eHB) depending on whether health and non-health benefits are substitutes or complements in the production of health services. We therefore had to make strong assumptions to value cost in terms of eHB, though sensitivity checks suggested that the effect on provider rankings was relatively minor.

Our exploratory analysis suggests some pointers for future research on measuring the performance of healthcare providers. The first requirement for future analyses is the routine collection of the health effect of treatment using measures of patient health pre- and post-treatment, rather than crude post-treatment measures (such as mortality, emergency readmissions and surgical revisions). Some non-health attributes of healthcare, such as waiting time or distance travelled to access care, can be captured in routine administrative data. But some other important dimensions of care, such as continuity of care or the quality of interactions with health service staff, are likely to require surveys of service users and carers. This is feasible: for example, the NHS now has patient surveys for primary care, maternity services, children services, inpatients, outpatients, accident and emergency, cancer services as well as the CMHS for mental health services.

Second, where incorporating non-health attributes, it is possible to obtain valuations by patients in units of the health measure, using DCEs. Alternatively, when patients have a choice amongst different providers of a given type of healthcare, it may be possible to estimate models of patient choice to obtain revealed, rather than stated preference valuations of non-health attributes in terms of health. The final and most difficult step in constructing a summary measure of overall provider performance, is to value health on the same scale as provider costs. We convert provider costs into health units using estimates of the NHS cost-effectiveness threshold. An alternative approach would be to use DCEs to estimate willingness to pay for health and non-health attributes (de Bekker-Grob et al., 2012) although there may be challenges to do this in a health system where care is traditionally free at the point of access, and attempts to explore this have yielded a wide range of estimated values (Donaldson et al., 2011).

Future work may also wish to include estimates of uncertainty to assess whether differences in relative performance rankings cannot be explained by chance variation alone. It may also be important to explore different ways to aggregate the underlying health and non-health benefits, including nonlinear approaches. Future research may also wish to replicate our approach using high quality data from a different context to mental health.

Our study demonstrates a novel method of extending the provider performance evaluation space beyond a measure of health gain and cost and we show that this matters for the comparison of providers. Whilst health gain and cost are important, they are not the only relevant dimensions when judging the overall performance of healthcare providers. This new approach can help regulators and policymakers to identify good and poor performance on dimensions that matter to patients, healthcare providers and the general public.

CRediT authorship contribution statement

María José Aragón: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation. **Hugh Gravelle:** Writing – review & editing, Methodology, Investigation, Conceptualization. **Adriana Castelli:** Writing – review & editing, Methodology, Investigation, Data curation. **Maria Goddard:** Writing – review & editing, Investigation, Funding acquisition. **Nils Gutacker:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Funding acquisition, Conceptualization. **Anne Mason:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Donna Rowen:** Writing – review & editing, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Russell Mannion:** Writing – review & editing, Investigation, Funding acquisition. **Rowena Jacobs:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization.

Data availability

The authors do not have permission to share data.

Acknowledgements

This project is part of the Health Foundation's Efficiency Research Programme. The Health Foundation is an independent charity committed to bringing about better health and health care for people in the UK. The project was undertaken on the Data Safe Haven, which is an ISO 27001 certified environment for handling sensitive data, and is provided by the University of York. We are grateful for support from the University of York Data Safe Haven team and the Research Computing team. Mental Health Minimum Dataset (MHMDS) are copyright © 2012/13–2015/16, the Health and Social Care Information Centre and re-used with the permission of the Health and Social Care Information Centre. All rights reserved. We are grateful to Dr Rita Santos for her help with the calculation of the distance between patient and provider.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.socscimed.2024.116582.

References

Allen, T., Walshe, K., Proudlove, N., Sutton, M., 2020. Using quality indicators to predict inspection ratings: cross-sectional study of general practices in England. Br. J. Gen. Pract. 70 (690), e55–e63. https://doi.org/10.3399/bjgp19X707141.

Appleby, J., Davlin, N., Parkin, D., 2007. NICE's cost effectiveness threshold. BMJ 335 (7616), 358–359.

Appleby, J., Poteliakhoff, E., Shah, K., Devlin, N., 2013. Using patient-reported outcome measures to estimate cost-effectiveness of hip replacements in English hospitals. J. R. Soc. Med. 106, 323–331.

Bell, A., 2022. Mental Health Services in the UK in 2022: what the Latest Adult Mental Health Care Benchmarking Network Data Can Tell Us, Centre for Mental Health. https://www.centreformentalhealth.org.uk/blogs/mental-health-services-uk-2022-what-latest-adult-mental-health-care-benchmarking-network-data-can-tell-us.

Brazier, J., Connell, J., Papaioannou, D., Mukuria, C., Mulhern, B., Peasgood, T., Lloyd Jones, M., Paisley, S., O'Cathain, A., Barkham, M., Knapp, M., Byford, S., Gilbody, S., Parry, G., 2014. A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures. Health Technol. Assess. 18.

Briggs, A., 2016. A view from the bridge: health economic evaluation—a value-based framework? Health Econ. 25 (12), 1499–1502.

Brouwer, W.B.F., Culyer, A.J., van Exel, N.J.A., Rutten, F.F.H., 2008. Welfarism vs. extra-welfarism. J. Health Econ. 27, 325–338.

Care Quality Commission, 2017. About Us: what We Do and How We Do it. https://www.cqc.org.uk/sites/default/files/20170504_CQC_About-us.pdf.

Cavallaro, F., Ewbank, L., Marszalek, K., Grimm, F., Tallack, C., 2023. Longer Hospital Stays and Fewer Admissions: How NHS Hospital Care Changed in England between 2019 and 2022. https://www.health.org.uk/publications/long-reads/longer-hospital-stays-and-fewer-admissions.

Claxton, K., Martin, S., Soares, M., Rice, N., Spackman, E., Hinde, S., Devlin, N., Smith, P., Sculpher, M., 2015. Methods for the estimation of the NICE cost effectiveness threshold. Health Technol. Assess. 19.

Coast, J., 2004. Is economic evaluation in touch with society's health values? Br. Med. J. 329, 1233–1236.

Coast, J., Smith, R., Lorgelly, P., 2008. Welfarism, extra-welfarism and capability: the spread of ideas in health economics. Soc. Sci. Med. 67, 1190–1198.

Curtis, L., Burns, A., 2016. Unit costs of health and social care 2016. In: Personal Social Services Research Unit. University of Kent, Canterbury.

de Bekker-Grob, E., Ryan, M., Gerard, K., 2012. Discrete choice experiments in health economics: a review of the literature. Health Econ. 21 (2), 145–172.

Delaffon, V., Anwar, Z., Noushad, F., Ahmed, A.S., Brugha, T.S., 2012. Use of health of the nation outcome scales in psychiatry. Adv. Psychiatr. Treat. 18, 173–179.

Department of Health, 2010. Healthy Lives, Healthy People: Our Strategy for Public Health in England. CM 7985. https://www.gov.uk/government/publications/healthy-lives-healthy-people-our-strategy-for-public-health-in-england. (Accessed 28 June 2023).

Department of Health and Social Care, 2023. The NHS Constitution for England. https://www.gov.uk/government/publications/the-nhs-constitution-for-england/the-nhs-constitution-for-england. (Accessed 28 June 2023).

Devlin, N., Parkin, M., 2004. Does NICE have cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. Health Econ. 13, 437–452. https://doi.org/10.1002/hec.864.

Devlin, N., Sussex, J., 2011. Incorporating Multiple Criteria in HTA. Methods and Processes. Office of Health Economics, London. https://www.ohe.org/wp-content/uploads/2014/07/350-IncorporatingMutlipleCriteria.pdf.

Donaldson, C., Baker, R., Mason, A., Jones-Lee, M., Lascar, E., et al., 2011. The social value of a QALY: raising the bar or barring the raise? BMC Health Serv. Res. 11, 8.

Drummond, M.F., Sculpher, M.J., Claxton, K., Stoddart, G.L., Torrance, G.W., 2015. Methods for the Economic Evaluation of Health Care Programmes. Oxford University Press, Oxford.

Entwistle, V., Firnigl, D., Ryan, M., Francis, J., Kinghorn, P., 2012. Which experiences of health care delivery matter to service users and why? A critical interpretive synthesis and conceptual map. J. Health Serv. Res. Policy 17, 70–78.

Frazão, T., Camilo, D., Cabral, E., Souza, R., 2018. Multicriteria decision analysis (MCDA) in health care: a systematic review of the main characteristics and methodological steps. BMC Med. Inf. Decis. Making 18, 90. https://doi.org/10.1186/s12911-018-0663-1.

Gilburt, H., Rose, D., Slade, M., 2008. The importance of relationships in mental health care: a qualitative study of service users' experiences of psychiatric hospital admission in the UK. BMC Health Serv. Res. 8, 92.

Iezzoni, L.I., 2012. Risk Adjustment for Measuring Health Care Outcomes.

Jacobs, R., McDaid, D., 2010. Performance measurement in mental health services. In: Smith, P.C., Mossialos, E., Papanicolas, I., Leatherman, S. (Eds.), Performance Measurement for Health System Improvement: Experiences, Challenges and Prospects, pp. 426–471.

Jacobs, R., Smith, P.C., Street, A., 2006. Measuring Efficiency in Health Care: Analytic Techniques and Health Policy. Cambridge University Press, Cambridge.

Karnon, J., Caffrey, O., Pham, C., Grieve, R., Ben-Tovim, D., Hakendorf, P., Crotty, M., 2013. Applying risk adjusted cost-effectiveness (RAC-E) analysis to hospitals: estimating the costs and consequences of variation in clinical practice. Health Econ. 22, 631–642.

Keetharuth, A., Rowen, D., 2020. Mapping the health of nation outcomes scale (HoNOS) onto the recovering quality of life utility index (ReQoL-UI). In: ScHARR HEDS Discussion Papers. University of Sheffield.

Keetharuth, A., Rowen, D., Björner, J.B., Brazier, J., 2020. Estimating a preference-based index for mental health from the Recovering Quality of Life (ReQoL) measure: valuation of ReQoL-UI. In: ScHARR HEDS Discussion Papers. University of Sheffield.

McCrone, P., Dhanasiri, S., Patel, A., Knapp, M., Lawton-Smith, S., 2008. Paying the Price: the Cost of Mental Health Care in England to 2026. The King's Fund.

Mental Health Taskforce, 2016. The Five Year Forward View for Mental Health.

Moran, V., Jacobs, R., 2018. Investigating the relationship between costs and outcomes for English mental health providers: a bi-variate multi-level regression analysis. Eur. J. Health Econ. 19, 709–718.

Murray, C.J.L., Frenk, J., 2000. A framework for assessing the performance of health systems. Bull. World Health Organ. 78 (6), 717–731.

Newman, D., O'Reilly, P., Lee, S.H., Kennedy, C., 2015. Mental health service users' experiences of mental health care: an integrative literature review. J. Psychiatr. Ment. Health Nurs. 22, 171–182.

NHS, 2017. Next Steps on the NHS Five Year Forward View. https://www.england.nhs.uk/wp-content/uploads/2017/03/NEXT-STEPS-ON-THE-NHS-FIVE-YEAR-FORWARD-VIEW.pdf. (Accessed 28 June 2023).

NHS England, 2016. Implementing the Five Year Forward View for Mental Health. https://www.england.nhs.uk/wp-content/uploads/2016/07/fyfv-mh.pdf. (Accessed 28 June 2023).

Paulden, M., 2017. Recent amendments to NICE's value-based assessment of health technologies: implicitly inequitable? Expert Rev. Pharmacoecon. Outcomes Res. 17, 239–242.

Powell, P.A., Rowen, D., 2022. What matters for evaluating the quality of mental healthcare? Identifying important aspects in qualitative focus groups with service users and frontline mental health professionals. The Patient: Patient-Centered Outcomes Research 15, 669–678.

QualityWatch, 2015. Focus on: People with Mental Ill Health and Hospital Use. Nuffield Trust and The Health Foundation.

Ride, J., Kasteridis, P., Gutacker, N., Aragon Aragon, M.J., Jacobs, R., 2020. Healthcare costs for people with serious mental illness in England: an analysis of costs across primary care, hospital care, and specialist mental healthcare. Appl. Health Econ. Health 18, 177–188.

Rowen, D., Powell, P.A., Hole, A.R., Aragón, M.J., Castelli, A., Jacobs, R., 2022. Valuing quality in mental healthcare: a discrete choice experiment eliciting preferences from mental healthcare service users, mental healthcare professionals and the general population. Soc. Sci. Med. https://doi.org/10.1016/j.socscimed.2022.114885.

Ryan, A.M., Tompkins, C.P., 2014. Efficiency and Value in Healthcare: Linking Cost and Quality Measures: A Paper Commissioned by the National Quality Forum. https://www.qualityforum.org/Publications/2014/11/Efficiency_and_Value_in_Healthcare_Linking_Cost_and_Quality_Measures_Paper.aspx.

Soekhai, V., de Bekker-Grob, E.W., Ellis, A.R., Vass, C.M., 2019. Discrete choice experiments in health economics: past, present and future. Pharmacoeconomics 37 (2), 201–226.

Stamboglis, N., Jacobs, R., 2020. Factors associated with patient satisfaction of community mental health services: a multilevel approach. Community Ment. Health J. 56, 50–64.

Stinnett, A.A., Mullahy, J., 1998. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Med. Decis. Making 18, S68–S80.

Timbie, J.W., Normand, S.-L.T., 2008. A comparison of methods for combining quality and efficiency performance measures: profiling the value of hospital care following acute myocardial infarction. Stat. Med. 27, 1351–1370.

Timbie, J.W., Newhouse, J.P., Rosenthal, M.B., Normand, S.-L.T., 2008. A cost-effectiveness framework for profiling the value of hospital care. Med. Decis. Making 28, 419–434.

Urbanoski, K., Inglis, D., 2019. Performance measurement in mental health and addictions systems: a scoping review. J. Stud. Alcohol Drugs Supplement, 114–130.

Wildman, J., Wildman, J.M., 2019. Combining health and outcomes beyond health in complex evaluations of complex interventions: suggestions for economic evaluation. Value Health 22 (5), 511–517. https://doi.org/10.1016/j.jval.2019.01.002.
Wildman, J., McMeekin, P., Grieve, E., Briggs, A., 2016. Economic evaluation of integrated new technologies for health and social care: suggestions for policy makers, users and evaluators. Soc. Sci. Med. 169, 141–148. https://doi.org/10.1016/j.socscimed.2016.09.033.
Wing, J.K., Beevor, A.S., Curtis, R.H., Park, S.G.B., Hadden, J., Burns, A., 1998. Health of the nation outcome scales (HoNOS): research and development. Br. J. Psychiatry 172, 11–18.