

A systematic review of clinician-rated instruments to assess adults' levels of functioning in specialised public sector mental health services

Philip M. Burgess^a, p.burgess@uq.edu.au

Meredith G. Harris^{a,b}, meredith.harris@uq.edu.au

Tim Coombs^c, timcoombs@live.com.au

Jane E. Pirkis^d, j.pirkis@unimelb.edu.au

^a The University of Queensland, School of Public Health, Australia

^b Policy and Epidemiology Group, Queensland Centre for Mental Health Research, Australia

^c Illawarra Institute for Mental Health, University of Wollongong, Australia

^d Centre for Mental Health, Melbourne School of Population and Global Health, The University of Melbourne, Australia

* Corresponding author: Philip Burgess, School of Public Health, The University of Queensland, Herston QLD 4006 Australia. Email: p.burgess@uq.edu.au

Abstract

Background

Functioning is one of the key domains emphasised in the routine assessment of outcomes that has been occurring in specialised public sector mental health services across Australia since 2002, via the National Outcomes and Casemix Collection (NOCC). For adult consumers (aged 18-64), the Life Skills Profile (LSP-16) has been the instrument of choice to measure functioning. However, review of the NOCC protocol has highlighted some limitations to the current approach to measuring functioning. A systematic review was conducted to identify, against a set of pre-determined criteria, the most suitable existing clinician-rated instruments for the routine measurement of functioning for adult consumers.

Method

We used two existing reviews of functioning measures as our starting point, and conducted a search of Medline and PsycInfo to identify articles relating to additional clinician-rated instruments. We evaluated identified instruments using a hierarchical, criterion-based approach. The criteria were: 1. Is brief (<50 items) and simple to score; 2. Is not made redundant by more recent instruments; 3. Relevant version has been scientifically scrutinised; 4. Considers functioning in a contemporary way; and 5. Demonstrates sound psychometric properties.

Results

We identified 20 relevant instruments, five of which met our criteria: the LSP-16, the Health of the Nation Outcome Scales (HoNOS), the Illness Management and Recovery Scale – Clinician Version (IMRS-C), the Multnomah Community Ability Scale (MCAS), and the Personal and Social Performance Scale (PSP).

Conclusions

Further work is required to determine which, if any, of these instruments satisfy further criteria relating to their appropriateness for assessing functioning within relevant service contexts, acceptability to clinicians and consumers, and feasibility in routine practice. This should involve seeking stakeholders' opinions (for example, about the specific domains of functioning covered by each instrument, and the language used in individual items) and testing completion rates in busy service settings.

Key words

measures; mental health services; functioning

Background

The International Classification of Functioning, Disability and Health (ICF) recognises functioning as an essential component of health and wellbeing (World Health Organization, 2001). The ICF emphasises functioning over disability, focusing on what people have the potential to do and actually do, irrespective of their mental (or physical) health conditions. The ICF stresses two key elements of functioning: ‘activity’ (the execution of tasks) and ‘participation’ (involvement in life situations) (World Health Organization, 2001).

Functioning is one of the key domains that has been emphasised in the routine assessment of outcomes that has been occurring in specialised public sector mental health services across Australia since 2002, via the National Outcomes and Casemix Collection (NOCC) (Burgess et al., 2015). Under the NOCC protocol, various outcome measurement instruments are administered for all consumers at set points in their episode of care. For adults (aged 18-64) receiving care in non-admitted settings, the main instrument used to assess functioning to date has been the Life Skills Profile (LSP-16) (Buckingham et al., 1998a; 1998b). The Health of the Nation Outcome Scales (HoNOS) (Craig et al., 2004; Wing, 1999a; 1999b; Wing et al., 1998; Wing et al., 1999) which is primarily used to assess severity of symptoms, also contains a small number of items that relate to functioning. The LSP-16 and the HoNOS are both clinician-rated. More information about the full NOCC suite of instruments and the framework that guides their administration can be found elsewhere (Department of Health, 2015).

Measures of functioning are also important for casemix classification and funding purposes. With respect to the latter, in 2016, the Independent Hospital Pricing Authority (IHPA) released the Australian Mental Health Care Classification (AMHCC) Version 1.0, a national classification for mental health care (Independent Hospital Pricing Authority (IHPA), 2016). It is based on available consumer-level clinical and treatment information, including information gathered from the instruments administered under the

NOCC protocol. Of relevance, the AMHCC Version 1.0 classification uses LSP-16 scores as one indicator of case complexity for adult consumers in community settings.

In 2013, NOCC was reviewed by the National Mental Health Information Development Expert Advisory Panel (NMHIDEAP) which gathered information from a variety of sources, including multi-modality stakeholder consultations and analysis of NOCC data. Some specific issues were identified through those consultations regarding the use of the LSP-16 with adult consumers. These included that it is not strengths-based, it uses outdated language, the wording of some items is unclear, and completion rates are lower than desired (National Mental Health Information Development Expert Advisory Panel (NMHIDEAP), 2013). Notwithstanding these issues, experience with the LSP-16 over an almost 20 year period is invaluable in informing broader considerations in the measurement of functioning. For adults, the review recommended that the NOCC suite of instruments be rationalised, and that a simple clinician-rated instrument be developed that assesses functioning and symptomatology and, potentially, other relevant domains. Such an instrument might take the form of a single existing instrument, or alternatively it might be a composite of several instruments, but either way it should be brief.

Since the review, NMHIDEAP has proposed a 'domain framework' that should guide developments in the measurement of functioning (National Mental Health Information Development Expert Advisory Panel (NMHIDEAP), 2015). This emphasises personal recovery, social recovery and clinical recovery.

NMHIDEAP has also suggested several options for how the new instrument should be developed: augmenting the HoNOS with a measure of functioning that replaces the LSP-16 and, if necessary, some additional clinically-relevant items; or constructing a new instrument that is purpose-designed to cover off all of the areas in the domain framework (again, this might have the HoNOS at its core).

We conducted the current systematic review in order to inform considerations about how functioning should be captured within the NOCC protocol. We did this as part of our role with the Australian Mental Health Outcomes and Classification Network (AMHOCN), which has been responsible for data

management, training and service development, and analysis and reporting related to NOCC since 2003 (Burgess et al., 2012). Our starting point was two reviews of functioning measures that had been conducted for different purposes. One of these looked at instruments that might be used in community managed organisations in Australia (Australian Mental Health Outcomes and Classification Network (AMHOCN) and Community Mental Health Australia (CMHA), 2013), and the other considered instruments that might be used in clinical services in New Zealand (Lutchman et al., 2007; Waikato Evaluation Team, 2005). Once we had considered the instruments that were shortlisted in these reviews, we conducted our own systematic review of the academic literature. We sought to identify articles that had been published since the original reviews, as well as any articles that might have been missed by these reviews. Our review aimed to answer the following question: *What are the most suitable existing clinician-rated instruments that might be used to routinely measure consumer functioning for adult consumers in Australian specialised public sector mental health services?*

Method

The current systematic review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher et al., 2009). We conducted an iterative search of Medline and PsycINFO from their respective years of inception to April 2016 for journal articles that described relevant functioning instruments. In the first iteration, we searched titles and abstracts using the following search string: ('mental' OR 'psychiatr*') AND ('social function*' OR 'personal function*' OR 'community function*' OR 'social abilit*' OR 'personal abilit*' OR 'community abilit' OR 'social perform*' OR 'personal perform*' OR 'community perform*' OR 'occupation* function' OR 'occupation* perform*' OR 'community participat*' OR 'community involve*' OR 'work' OR 'leisure' OR 'educat*' OR 'personal relationship*' OR 'interpersonal relationship*' OR 'social inclusion' OR 'living skill*' OR 'life skill*' OR 'self-care'). In the second iteration, we searched titles only for the names of identified instruments, in order to ensure that we picked up as many relevant articles on each as possible. We also searched the reference lists of key review papers and articles on individual instruments. Our search was restricted to English-language articles.

At the title and abstract screening and full-text screening stages, we excluded articles that made reference to instruments that could not be readily rated by a clinician without recourse to other information (e.g., consumer-rated instruments, instruments that required structured or semi-structured interviews with consumers or other informants, instruments that involved a systematic extraction of information from case notes). We also excluded articles on instruments that were designed for use with non-adult populations or clinically defined sub-populations (e.g., instruments designed for use with children and adolescents or older persons, instruments designed for use with people with intellectual disabilities, instruments designed for use in forensic mental health settings). In addition, we excluded articles on instruments that assessed only a limited aspect of functioning (e.g., instruments that were exclusively about activities of daily living, instruments that focused only on work performance).

Once we had identified our pool of relevant articles, we assessed whether each of the given instruments they described might be candidates for routinely assessing changes in functioning of consumers in Australian public sector mental health services. We did this using a hierarchical, criterion-based approach based on one that we used for a previous review of recovery instruments (Burgess et al., 2011). Under this approach, we progressively excluded instruments from further consideration if they did not meet a specific criterion. The criteria were:

1. Is brief (<50 items) and simple to score;
2. Is not made redundant by more recent instruments;
3. Relevant version has been scientifically scrutinised;
4. Considers functioning in a contemporary way; and
5. Demonstrates sound psychometric properties.

For each instrument meeting the above criteria, we extracted and summarised information describing its purpose and structure, and its psychometric properties. The psychometric properties considered were:

- Validity, or the extent to which the instrument measures what it intends to measure. Three types of validity were examined: construct validity; concurrent validity; and predictive validity.
- Reliability, or the extent to which the instrument gives stable, consistent results. Three aspects of reliability were considered: internal consistency; inter-rater reliability; and test-retest reliability;
- Sensitivity to change, or the extent to which, assuming the instrument is valid and reliable, it demonstrates the capacity to detect change over time.

Results

Overview of identified articles and instruments

An overview of the identified articles is provided in Figure 1. In total, our search identified 5,907 journal articles. Removal of duplicates and screening titles and abstracts left 335 full text articles, of which 81 were excluded when the full text was reviewed. The remaining 254 articles provided information about 20 clinician-rated instruments designed to measure functioning. Table 1 profiles the 20 instruments, describing them in terms of when and where they were developed, the domains they assess, and their item structure.

Hierarchical, criterion-based assessment of the instruments

Criterion 1: Is brief (<50 items) and simple to score

Figure 2 shows that 13 of the 20 instruments meet the first criterion. The exceptions are the FACE Core Assessment, the Level of Functioning Scale (LFS), the Life Functioning Assessment Inventory (LFAI), the Multi-Function Needs Assessment (MFNA), the Residential Competency Scale (RCS), the Social Adjustment Behavior Rating Scale (SABRS) and the Social Functioning Index (SFI). The LFAI is complex to score because it the domains it assesses are given status scores (reflecting general performance) and grade scores (reflecting more specific performance levels within the grade). The remaining exceptions range in length from 50 items (the FACE Core Assessment) to 134 items (the MFNA), making them unsuitable for use in routine outcome measurement. These instruments are excluded from further analysis.

Criterion 2: Is not made redundant by more recent instruments

Figure 2 shows that the majority of the remaining 13 instruments remain in contention when this criterion is examined. The two exceptions are the Global Assessment of Functioning (GAF) and the Social and Occupational Functioning Assessment Scale (SOFAS). The GAF was introduced in the revised version of the third edition of the Diagnostic and Statistical Manual (DSM-III-R) as a means of assessing 'adaptive functioning' (American Psychiatric Association, 1987), but was eliminated from subsequent versions of the DSM because it was regarded as being inadequate for assessing a construct like functioning that may be volatile and may not operate independently of symptomatology, and because of the training required for it to be used appropriately (Suzuki et al., 2015). The GAF was replaced by the SOFAS, on the grounds that the SOFAS assessed social and occupational functioning independently of symptom severity (Hendryx et al., 2001). In turn, the SOFAS has been superseded by the Personal and Social Performance Scale (PSP), which demonstrates stronger psychometric performance (Morosini et al., 2000). This sequence of instrument development and replacement led us to eliminate the GAF and the SOFAS from further consideration.

Criterion 3: Relevant version has been scientifically scrutinised

We considered whether the relevant version of each of the remaining 11 instruments had been subjected to scientific scrutiny. To satisfy this criterion, the given instrument had to have been assessed by investigators who were independent of the original instrument developers, and the results of that assessment had to have been published in the peer reviewed literature. It should be noted that the LSP-16 is included among these instruments. The LSP-16 is a short version of its parent instrument, the LSP-39 (Rosen et al., 1989). We focused on the LSP-16 as the 'relevant' instrument because this is the version of the instrument that is in current use in specialised public sector mental health services in Australia. Reference is made to studies scrutinising the LSP-39, as appropriate, however. Figure 2 indicates that six instruments satisfied this criterion. Those which have not been subjected to scientific scrutiny are the

Disability Rating Form (DRF), the Mini-ICF-APP, the Need of Support and Service Questionnaire (NSSQ), the Profile of Community Psychiatry Clients (PCPC) and the Uniform Client Data Instrument (UCDI). These were excluded from further examination.

Criterion 4: Considers functioning in a contemporary way

We evaluated whether the remaining six instruments consider functioning in a contemporary way. Figure 2 shows that we removed the Rehabilitation Evaluation Hall and Baker (REHAB) at this point. This instrument was developed in 1984, in the era of deinstitutionalisation, and was designed for use with residents of long term psychiatric facilities who were being relocated to community residential support settings. It takes a limited view of functioning, and not one that recognises the capacity of people with mental illness to lead contributing lives. It primarily deals with activities of daily living, and only includes relatively few items on other aspects of functioning. Most of these are framed negatively, falling into the 'deviant behaviours' subscale of the instrument.

Criterion 5: Demonstrates sound psychometric properties

Table 2 summarises the psychometric properties of the five remaining instruments. All five have been subject to independent psychometric testing by investigators other than the original developers. Figure 2 shows that all five have relatively sound psychometric properties, although some caveats are worth noting here. For example, the Health of the Nation Outcome Scales (HoNOS) has been extensively examined in its entirety, but less attention has been paid to the *social* subscale which contains the four items (Items 9-12) that relate to functioning that are relevant here. When this subscale and its component items have been assessed, they have sometimes performed less well than other elements of the instrument (particularly Items 11 and 12, which relate to *living conditions* and *occupation and activities*, where functioning is not independent of opportunities). The Life Skills Profile (LSP-16) has undergone more limited psychometric testing, however, and most of the information on its psychometric

properties comes from assessments of its parent instrument, the LSP-39. The Illness Management and Recovery Scale – Clinician Version (IMRS-C) has also undergone limited testing; further information on its inter-rater reliability and sensitivity to change would be desirable. Across all instruments, some consistent gaps were evident. Notably, we found only one or two studies examining the predictive validity for each of the IMRS-C, MCAS, PSP, HoNOS *social* subscale and the LSP-16 (as opposed to the LSP-39). Moreover, the measures used to establish predictive validity for each instrument were diverse, making it difficult to compare their relative performance. Information about sensitivity to change was also limited – being absent for the LSP-16 (as opposed to the LSP-39) and the MCAS, and available from only two studies for the IMRS-C, both conducted within a single program context.

Discussion

We used a hierarchical, criterion-based approach to identify candidate instruments for measuring functioning among adult consumers of specialised public sector mental health services. By the end of the elimination process, we had reduced 20 potential instruments to five: the Health of the Nation Outcome Scales (HoNOS); the Illness Management and Recovery Scale – Clinician Version (IMRS-C); the Life Skills Profile (LSP-16); the Multnomah Community Ability Scale (MCAS); and the Personal and Social Performance Scale (PSP). The HoNOS, the MCAS and the PSP were all shortlisted in the two previous reviews that we drew upon, and the LSP-16 was shortlisted in the Australian review (Australian Mental Health Outcomes and Classification Network (AMHOCN) and Community Mental Health Australia (CMHA), 2013) but not the New Zealand one (Lutchman et al., 2007; Waikato Evaluation Team, 2005). The IMRS-C was not identified in either of these previous reviews, so it did not feature in their shortlists.

The current review is a first step in further developing the measurement of functioning. All five of the above instruments are recommended for consideration as clinician-rated instruments that might be used to routinely measure adult consumers' functioning in Australian mental health services. However, further work is required to consider the appropriateness of the candidate instruments for assessing functioning in relevant service contexts, their acceptability to clinicians and consumers, and the feasibility of using them in routine practice. The consideration process should be systematic and structured. It should involve seeking stakeholders' opinions about, for example, the specific domains of functioning covered by each instrument, and the language used in individual items. Ideally, the process should also involve some real-world testing of clinicians' completion of the instruments in specialised community mental health settings. Completion rates for the most recent available year (2014-2015), for the two instruments that are already part of the NOCC suite, showed that the HoNOS was completed at 83% of review/discharge collection occasions, and the LSP-16 was completed at 71% (National Mental Health

Information Development Expert Advisory Panel (NHMIDEAP), 2013). Similar field testing of the other three instruments is desirable.

A key consideration in terms of appropriateness relates to the capacity of each of the instruments to measure outcomes meaningfully within relevant service contexts. Our review examined a range of psychometric properties, including sensitivity to change, and all of the instruments performed reasonably well on at least several of these. Further testing is needed, however, to address gaps regarding predictive validity and sensitivity to change, which presently limit the extent to which conclusions can be drawn about the way the instruments work across the domains of functioning they are intended to measure and the contexts in which they would be implemented, and to compare their relative performance. There has also been increasing discussion in the literature regarding the distinction between reflective and formative indicators, and the selection of appropriate measurement models for each (Bollen and Bauldry, 2011). Future investigations could consider these issues in relation to the measurement of functioning. Other factors need to be taken into account too, however. For instance, in specifying the time period covered by the instruments, it is necessary to ensure that no two rating periods overlap. The PSP asks about the consumer's general functioning, without specifying time period, so this is not an issue for this instrument. The HoNOS covers the previous two weeks, which is sufficiently short that the issue of potentially overlapping assessment periods is minimised in most cases. The LSP-16 covers the last three months, as does the IMRS-C. The MCAS has rating periods of three months and one year, depending on the specific item. Consideration might be given to exploring whether these instruments can be modified to cover shorter time periods. Precedents exist for these sorts of modifications; an alternative version of the MCAS exists which has a rating period of one month (Dickerson et al., 2003). Any such modifications would need to be tested.

The five shortlisted instruments each have items that address 'activities' and 'participation', identified as core elements of functioning in the ICF. In part this is because our initial exclusion criteria meant that instruments that only measured activities (or, more specifically, activities of daily living) were discarded

before they reached the point of review. The various instruments placed differing emphasis on these two elements, however, and included divergent domains within each of them. The HoNOS assesses *relationships, activities of daily living, living conditions and occupation and activities*. The IMRS-C covers *recovery, management and biology*. The LSP-16 focuses on *withdrawal, self-care, compliance and anti-social behaviour*. The MCAS considers *interference with functioning, adjustment to living, social competence and behavioural problems*. The PSP provides a rating that is based on *socially useful activities, personal and social relationships, self-care and disturbing and aggressive behaviours*. When the appropriateness, acceptability and feasibility of the five instruments is explored, consideration should be given to stakeholders beliefs about the precise domains that should be assessed and the relative emphasis that should be placed on ‘activities’ and ‘participation’. Future work could also build upon the current review by evaluating the psychometric properties of the identified instruments separately in relation to the measurement of ‘activities’ and ‘participation’.

The scope of the current review was restricted to clinician-rated measures of functioning for adult consumers that could be used in as part of the new instrument that was recommended by NMHIDEAP (National Mental Health Information Development Expert Advisory Panel (NHMIDEAP), 2013). This meant that we excluded consumer-rated instruments and clinician-rated instruments that sought information via consumer interviews or case note reviews, doing so at the stage of screening the abstracts and full-text of identified journal articles. More than 80 additional, but out-of-scope, instruments designed to measure functioning were eliminated at this pre-review stage. Some of these instruments undoubtedly have merit. For example, the Camberwell Assessment of Need Short Appraisal Schedule (CANSAS) (Phelan et al., 1995) is popular and has sound psychometric properties, but was excluded because it involves a structured interview in which clinician, consumer and carer views of need can be recorded separately. If the examination of appropriateness, acceptability and feasibility of the five instruments does not yield positive findings, then consideration might be given to broadening the search criteria and identifying additional instruments (albeit ones that might need to be modified to be fit for purpose).

Decisions about whether or not to use one of the five identified instruments – or to seek alternatives – should not be made in isolation. The clinician-rated instruments in the current NOCC suite are complemented by various consumer-rated instruments. At present, these primarily relate to levels of distress and other psychological symptoms, but there is an appetite for broadening these to include constructs like social inclusion and recovery. AMHOCN has reviewed existing recovery and social inclusion instruments (Burgess et al., 2011; Coombs et al., 2013), and has developed and trialled a new social inclusion instrument (the Living in the Community Questionnaire) (Australian Mental Health Outcomes and Classification Network (AMHOCN), 2015). There is an argument that these constructs are closely related to functioning, particularly the ‘participation’ element of functioning. There is also an argument that whereas a consumer’s level of functioning can be assessed by either a clinician or by the consumer him/herself, social inclusion and recovery are more appropriately measured by the consumer, because of their experiential nature. Consideration should be given to how the selected clinician-rated measure of functioning complements proposed consumer-rated social inclusion and recovery instruments.

Identifying an appropriate clinician-rated functioning instrument should not stop with adult consumers. The current review excluded instruments that were designed for specific populations, including children and adolescents and older people. Norms around functioning are clearly age-related to some extent, so it makes sense that functioning instruments that have utility for adult consumers may not do so for younger and older consumers. For younger consumers, levels of maturity will impact upon functioning. For older consumers, physical and cognitive abilities may play a role. Age-specific functioning instruments are required for these groups, and we would recommend a similar process for identifying them.

We acknowledge that our review had some limitations. Despite our best efforts, we may have missed some relevant and potentially useful instruments designed to assess functioning (e.g., if our search terms

did not pick up articles related to them, or if these articles were not indexed in the two academic databases we used). Also, we may have missed some articles relating to the instruments we did identify, so our examination of the psychometric properties of the final five may not have been exhaustive. In addition, the articles we did retrieve did not always provide optimal detail on the instruments they described (particularly with respect to the specific items on these instruments), so it is possible that we misinterpreted information about some of them. Finally, we cannot rule out possible publication bias. Studies showing that an instrument has good psychometric properties are more likely to be published than studies that do not. Having said that, our Criterion 3 required that included instruments had to have been scrutinised by investigators who were independent of the original instrument developers; this should have increased the extent to which the assembled evidence base included studies by investigators who did not have a vested interest in showing that a given instrument has sound psychometric properties.

These limitations aside, we believe that the current review can help to inform decisions about which clinician-rated instruments hold promise for assessing whether functioning improves, deteriorates or does not change for adult consumers of Australian mental health services. Further work is required to determine which, if any, of these instruments satisfy further criteria relating to appropriateness, acceptability and feasibility.

Acknowledgements

The authors would like to acknowledge feedback from members of the National Mental Health Information Development Expert Advisory Panel (NMHIDEAP) on a previous version of this systematic review.

Declaration of interest

The Australian Mental Health Outcomes and Classification Network (AMHOCN) is funded by the Australian Government Department of Health. The authors alone are responsible for the content and writing of the paper.

References

- Adair CE, McDougall GM, Mitton CR, et al. (2005) Continuity of care and health outcomes among persons with severe mental illness. *Psychiatric Services* 56: 1061-1069.
- Allan S and McGonagle I. (1997) A comparison of HoNOS with the Social Behaviour Schedule in three settings. *Journal of Mental Health* 6: 117-124.
- American Psychiatric Association. (1987) Diagnostic and Statistical Manual of Mental Disorders (3rd edition, revised). Washington, DC: American Psychiatric Association.
- American Psychiatric Association. (1994) Diagnostic and Statistical Manual of Mental Disorders (4th edition). Washington, DC: American Psychiatric Association.
- Amin S, Singh SP, Croudace T, et al. (1999) Evaluating the Health of the Nation Outcome Scales. Reliability and validity in a three-year follow-up of first-onset psychosis. *British Journal of Psychiatry* 174: 399-403.
- Andreas S, Harfst T, Dirmaier J, et al. (2007) A psychometric evaluation of the German version of the 'Health of the Nation Outcome Scales, HoNOS-D': On the feasibility and reliability of clinician-performed measurements of severity in patients with mental disorders. *Psychopathology* 40: 116-125.
- Andreas S, Harfst T, Rabung S, et al. (2010) The validity of the German version of the Health of the Nation Outcome Scales (HoNOS-D): A clinician-rating for the differential assessment of the severity of mental disorders. *International Journal of Methods in Psychiatric Research* 19: 50-62.
- Andrews G, Peters L and Teesson M. (1994) Measurement of Consumer Outcome in Mental Health: A Report to the National Mental Health Information Strategy Committee. Sydney: Clinical Research Unit for Anxiety Disorders.
- Andrews G, Teesson M, Stewart G, et al. (1990) Follow-up of community placement in the chronic mentally ill in New South Wales. *Hospital and Community Psychiatry* 41: 184-188.
- Angelini DE. (1982) Functional needs of the chronically mentally ill: Implications for service delivery. *Psychosocial Rehabilitation Journal* 5: 29-33.
- Apiquian R, Elena Ulloa R, Herrera-Estrella M, et al. (2009) Validity of the Spanish version of the Personal and Social Performance scale in schizophrenia. *Schizophrenia Research* 112: 181-186.
- Audin K, Margison FR, Clark JM, et al. (2001) Value of HoNOS in assessing patient change in NHS psychotherapy and psychological treatment services. *British Journal of Psychiatry* 178: 561-566.
- Aumack L. (1962) A social adjustment behavior rating scale. *Journal of Clinical Psychology* 18: 436-441.
- Australian Mental Health Outcomes and Classification Network (AMHOCN). (2015) Development of the Living in the Community (LCQ) Measure of Social Inclusion for Use in Mental Health: Final Report. Sydney: AMHOCN.
- Australian Mental Health Outcomes and Classification Network (AMHOCN) and Community Mental Health Australia (CMHA). (2013) National Community Managed Organisation (CMO) Outcome Measurement Project: Final Report to the Mental Health Information Strategy Standing Committee. Sydney: AMHOCN.
- Baker R and Hall J. (1984) User's Manual for Rehabilitation Evaluation. Aberdeen: Vine Publishing.
- Baker R and Hall J. (1988) REHAB: A new assessment instrument for chronic psychiatric patients. *Schizophrenia Bulletin* 14: 97-110.

- Barker S, Barron N, McFarland B, et al. (1994a) Multnomah Community Ability Scale: User's Manual. Portland, OR: Western Mental Health Research Center, Oregon Health Sciences University.
- Barker S, Barron N, McFarland BH, et al. (1994b) A community ability scale for chronically mentally ill consumers: I. Reliability and validity. *Community Mental Health Journal* 30: 363-383.
- Barker S, Barron N, McFarland BH, et al. (1994c) A community ability scale for chronically mentally ill consumers: Part II. Applications. *Community Mental Health Journal* 30: 459-472.
- Bassani DG, Dewa CS, Krupa T, et al. (2009) Factor structure of the Multnomah Community Ability Scale: Longitudinal analysis. *Psychiatry Research* 167: 178-189.
- Bebbington P, Brugha T, Hill T, et al. (1999) Validation of the Health of the Nation Outcome Scales. *British Journal of Psychiatry* 174: 389-394.
- Bech P, Bille J, Schutze T, et al. (2003) Health of the Nation Outcome Scales (HoNOS): Implementability, subscale structure and responsiveness in the daily psychiatric hospital routine over the first 18 months. *Nordic Journal of Psychiatry* 57: 285-290.
- Bollen KA and Bauldry S. (2011) Three Cs in measurement models: Causal indicators, composite indicators, and covariates. *Psychological Methods* 16: 265-284.
- Brissos S, Palhava F, Marques JG, et al. (2012) The Portuguese version of the Personal and Social Performance Scale (PSP): reliability, validity, and relationship with cognitive measures in hospitalized and community schizophrenia patients. *Social Psychiatry and Psychiatric Epidemiology* 47: 1077-1086.
- Broadbent M. (2001) Reconciling the information needs of clinicians, managers and commissioners: A pilot project. *Psychiatric Bulletin* 25: 423-425.
- Brooks R. (2000) The reliability and validity of the Health of the Nation Outcome Scales: validation in relation to patient derived measures. *Australian and New Zealand Journal of Psychiatry* 34: 504-511; discussion 520-501.
- Brown E, Rush A, Biggs M, et al. (2003) Clinician ratings vs. global ratings of symptom severity: A comparison of symptom measures in the bipolar disorder module, phase II, Texas Medication Algorithm Project. *Psychiatry Research* 117: 167-175.
- Browne G and Courtney M. (2004) Measuring the impact of housing on people with schizophrenia. *Nursing and Health Sciences* 6: 37-44.
- Browne S, Doran M and McGauran S. (2000) Health of the Nation Outcome Scales (HoNOS): Use in an Irish psychiatric outpatient population. *Irish Journal of Psychological Medicine* 17: 17-19.
- Buckingham W, Burgess P, Solomon S, et al. (1998a) Developing a Casemix Classification for Mental Health Services. Volume 1: Main Report. Canberra: Commonwealth Department of Health and Family Services.
- Buckingham W, Burgess P, Solomon S, et al. (1998b) Developing a Casemix Classification for Mental Health Services. Volume 2: Resource Materials. Canberra: Commonwealth Department of Health and Family Services.
- Burgess P, Coombs T, Clarke A, et al. (2012) Achievements in mental health outcome measurement in Australia: Reflections on progress made by the Australian Mental Health Outcomes and Classification Network (AMHOCN). *International Journal of Mental Health Systems* 6: 4.
- Burgess P, Pirkis J, Buckingham W, et al. (1999) Developing a casemix classification for specialist mental health services. *Casemix* 1: 4-20.
- Burgess P, Pirkis J and Coombs T. (2015) Routine outcome measurement in Australia. *International Review of Psychiatry* 27 264-275.
- Burgess P, Pirkis J, Coombs T, et al. (2011) Assessing the value of existing recovery measures for routine use in Australian mental health services. *Australian and New Zealand Journal of Psychiatry* 45: 267-280.
- Burgess P, Trauer T, Coombs T, et al. (2009) What does 'clinical significance' mean in the context of the Health of the Nation Outcome Scales? *Australasian Psychiatry* 17: 141-148.

- Cheah Y, Parker G, Hadzi-Pavlovic D, et al. (1998) Development of a measure profiling problems and needs of psychiatric patients in the community. *Social Psychiatry and Psychiatric Epidemiology* 33: 337-344.
- Clifford P. The FACE recording and measurement system: A scientific approach to person-based information. *Bulletin of the Menninger Clinic* 63: 305-331.
- Clifford P. (1994) The FACE Project: Final Report to the Department of Health. London: University College London.
- Clifford P. (1997a) Structuring the Clinical Record: The FACE Assessment and Outcome System (Appendices). London: University College London.
- Clifford P. (1997b) Structuring the Clinical Record: The FACE Assessment and Outcome System (Main Report to the Department of Health). London: University College London.
- Clifford P, Orbach G, Hobbins J, et al. (1999) Measuring disability and outcomes in routine practice with the FACE Core Assessment. *Bulletin of the Menninger Clinic* 63: 332-345.
- Coombs T, Nicholas A and Pirkis J. (2013) A review of social inclusion measures. *Australian and New Zealand Journal of Psychiatry* 47: 906-919.
- Corbiere M, Crocker AG, Lesage AD, et al. (2002) Factor structure of the Multnomah Community Ability Scale. *Journal of Nervous and Mental Disease* 190: 399-406.
- Craig T, Doherty I, Jamieson-Craig R, et al. (2004) The consumer-employee as a member of a Mental Health Assertive Outreach Team: (1) Clinical and social outcomes. *Journal of Mental Health* 13: 59-69.
- Department of Health. (2015) Mental Health National Outcomes and Casemix Collection: Technical Specification of State and Territory Reporting Requirements, Version 1.90. Canberra: Department of Health.
- Dickerson FB. (1997) Assessing clinical outcomes: the community functioning of persons with serious mental illness. *Psychiatric Services* 48: 897-902.
- Dickerson FB, Origoni AE, Pater A, et al. (2003) An expanded version of the Multnomah Community Ability Scale: anchors and interview probes for the assessment of adults with serious mental illness. *Community Mental Health Journal* 39: 131-137.
- Dickinson D and Coursey RD. (2002) Independence and overlap among neurocognitive correlates of community functioning in schizophrenia. *Schizophrenia Research* 56: 161-170.
- Eagar K, Trauer T and Mellis G. (2005) Performance of routine outcome measures in adult mental health care. *Australian and New Zealand Journal of Psychiatry* 39: 713-718.
- Egger ST, Weniger G, Prinz S, et al. (2015) Health of the Nation Outcome Scales in a psychiatric inpatient setting: assessing clinical change. *Journal of Evaluation in Clinical Practice* 21: 236-241.
- Färdig R, Lewander T, Fredriksson A, et al. (2011) Evaluation of the illness management and recovery scale in schizophrenia and schizoaffective disorder. *Schizophrenia Research* 132: 157-164.
- Ford R, Barnes A, Davies R, et al. (2001) Maintaining contact with people with severe mental illness: 5-year follow-up of assertive outreach. *Social Psychiatry and Psychiatric Epidemiology* 36: 444-447.
- Ford R, Ryan P, Beadsmoore A, et al. (1997) Intensive case management for people with serious mental illness - Site 2: Clinical and social outcome. *Journal of Mental Health* 6: 181-190.
- Gallagher J and Teesson M. (2000) Measuring disability, need and outcome in Australian community mental health services. *Australian and New Zealand Journal of Psychiatry* 34: 850-855.
- Garcia-Portilla MP, Saiz PA, Bousoño M, et al. (2011) Validation of the Spanish Personal and Social Performance scale (PSP) in outpatients with stable and unstable schizophrenia. *Revista de Psiquiatria y Salud Mental* 4: 9-18.
- Gingerich S and Mueser K. (2002) Illness Management and Recovery Implementation Resource Kit. Rockville, MD: Center for Mental Health Services, Substance Abuse and Mental Health Services Administration.

- Goldney RD, Fisher LJ and Walmsley SH. (1998) The Health of the Nation Outcome Scales in psychiatric hospitalisation: a multicentre study examining outcome and prediction of length of stay. *Australian and New Zealand Journal of Psychiatry* 32: 199-205.
- Hambridge J and Rosen A. (1994) Assertive community treatment for the seriously mentally ill in suburban Sydney: A programme description and evaluation. *Australian and New Zealand Journal of Psychiatry* 28: 438-445.
- Hamernik E and Pakenham KI. (1999) Assertive community treatment for persons with severe mental disorders: A controlled treatment outcome study. *Behaviour Change* 16: 259-268.
- Hasson-Ohayon I, Roe D and Kravetz S. (2007) A randomized controlled trial of the effectiveness of the illness management and recovery program. *Psychiatric Services* 58: 1461-1466.
- Hasson-Ohayon I, Roe D and Kravetz S. (2008) The psychometric properties of the Illness Management and Recovery scale: client and clinician versions. *Psychiatry Research* 160: 228-235.
- Hendryx M, Dyck DG, McBride D, et al. (2001) A test of the reliability and validity of the Multnomah Community Ability Scale. *Community Mental Health Journal* 37: 157-168.
- Hope JD, Trauer T and Keks NA. (1998) Reliability, validity and utility of the Health of the Nation Outcomes Scale (HoNOS) in Australian adult psychiatric services. *Schizophrenia Research* 29: 9-10.
- Hopkins M and Ramsundar N. (2006) Which factors predict case management services and how do these services relate to client outcomes? *Psychiatric Rehabilitation Journal* 29: 219-222.
- Hoyle RH, Nietzel MT, Guthrie PR, et al. (1992) The Disability Rating Form: A brief schedule for rating disability associated with severe mental illness. *Psychosocial Rehabilitation Journal* 16: 77-94.
- Hoyle RH, Nietzel MT, Guthrie PR, et al. (1993) The Disability Rating Form. *Psychosocial Rehabilitation Journal* 16: 153-160.
- Hsieh PC, Huang HY, Wang HC, et al. (2011) Intercorrelations between the Personal and Social Performance Scale, cognitive function, and activities of daily living. *Journal of Nervous and Mental Disease* 199: 513-515.
- Hui CL-M, Li Y-K, Leung K-F, et al. (2013) Reliability and validity of the Life Functioning Assessment Inventory (L-FAI) for patients with psychosis. *Social Psychiatry and Psychiatric Epidemiology* 48: 1687-1695.
- Hunter R, McLean J, Peck D, et al. (2004) The Scottish 700 Outcomes Study: A comparative evaluation of the Health of the Nation Outcome Scale (HoNOS), the Avon Mental Health Measure (AVON), and an Idiographic Scale (OPUS) in adult mental health. *Journal of Mental Health* 13: 93-105.
- Independent Hospital Pricing Authority (IHPA). (2016) Australian Mental Health Care Classification Version 1.0 User Manual. Sydney: IHPA.
- Issakidis C and Teesson M. (1999) Measurement of need for care: A trial of the Camberwell Assessment of Need and the Health of the Nation Outcome Scales. *Australian and New Zealand Journal of Psychiatry* 33: 754-759.
- Jansson L, Sonnander K and Wiesel F-A. (2005) Needs assessed by psychiatric health care and social services in a defined cohort of clients with mental disabilities. *European Archives of Psychiatry and Clinical Neuroscience* 255: 57-64.
- Jelastopulu E, Giourou E, Merikoulias G, et al. (2014) Correlation between the Personal and Social Performance scale (PSP) and the Positive and Negative Syndrome Scale (PANSS) in a Greek sample of patients with schizophrenia. *BMC Psychiatry* 14: 197.
- Johnston S, Salkeld G, Sanderson K, et al. (1998) Intensive case management: A cost-effectiveness analysis. *Australian and New Zealand Journal of Psychiatry* 32: 551-559.
- Juckel G, Schaub D, Fuchs N, et al. (2008) Validation of the Personal and Social Performance (PSP) Scale in a German sample of acutely ill patients with schizophrenia. *Schizophrenia Research* 104: 287-293.

- Kawata AK and Revicki DA. (2008) Psychometric properties of the Personal and Social Performance scale (PSP) among individuals with schizophrenia living in the community. *Quality of Life Research* 17: 1247-1256.
- Kazarian SS, Cole JD, Barnes SM, et al. (1989) The Residential Competency Scale: A new measure of residential functioning skills. *Journal of Community Psychology* 17: 297-303.
- Keller S and Hayes R. (1998) The relationship between the Allen Cognitive Level Test and the Life Skills Profile. *American Journal of Occupational Therapy* 52: 851-856.
- Kirkby K, Daniels B, Jones I, et al. (1997) A survey of social outcome in schizophrenia in Tasmania. *Australian and New Zealand Journal of Psychiatry* 31: 405-410.
- Kisely S, Campbell LA, Cartwright J, et al. (2010) Do the Health of the Nation Outcome Scales measure outcome? *Canadian Journal of Psychiatry* 55: 431-439.
- Kisely S, Campbell LA, Crossman D, et al. (2007) Are the Health of the Nation Outcome Scales a valid and practical instrument to measure outcomes in North America? A three-site evaluation across Nova Scotia. *Community Mental Health Journal* 43: 91-107.
- Kisely S, Preston N and Rooney M. (2000) Pathways and outcomes of psychiatric care: Does it depend on who you are, or what you've got? *Australian and New Zealand Journal of Psychiatry* 34: 1009-1014.
- Lavin D and Ryan P. (2012) Using quantitative research to measure recovery outcomes and correlates. *Irish Journal of Psychological Medicine* 29: 157-162.
- Linden M, Baron S and Muschalla B. (2009) Mini-ICF-Rating für Aktivitäts- und Partizipationsstörungen bei Psychischen Erkrankungen (Mini-ICF-APP). Hogrefe AG, Bern: Verlag Hans Huber.
- Little J. (2013) Multilevel confirmatory ordinal factor analysis of the Life Skills Profile-16. *Psychological Assessment* 25: 810-825.
- Lovaglio PG and Monzani E. (2011) Validation aspects of the Health of the Nation Outcome Scales. *International Journal of Mental Health Systems* 5: 20.
- Lovaglio PG and Monzani E. (2012) Health of the nation outcome scales evaluation in a community setting population. *Quality of Life Research* 21: 1643-1653.
- Lutchman R, Thompson A, Tait H, et al. (2007) In search of a standardised, comprehensive assessment of functioning. *New Zealand Journal of Occupational Therapy* 54: 33-38.
- Massey OT, Pokorny LJ and Kramer BH. (1989) The development of factor-based level of functioning scales from a level of care instrument. *Journal of Clinical Psychology* 45: 903-909.
- McClelland R, Trimble P, Fox ML, et al. (2000) Validation of an outcome scale for use in adult psychiatric practice. *Quality in Health Care* 9: 98-105.
- McDevitt J, Wilbur J, Kogan J, et al. (2005) A walking program for outpatients in psychiatric rehabilitation: pilot study. *Biological Research for Nursing* 7: 87-97.
- McGuire AB, Kean J, Bonfils K, et al. (2014) Rasch analysis of the illness management and recovery scale-clinician version. *Journal of Evaluation in Clinical Practice* 20: 383-389.
- Moher D, Liberati A, Tetzlaff J, et al. (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. *PLoS Medicine* 6: e1000097.
- Molodynski A, Linden M, Juckel G, et al. (2013) The reliability, validity, and applicability of an English language version of the Mini-ICF-APP. *Social Psychiatry and Psychiatric Epidemiology* 48: 1347-1354.
- Morosini PL, Magliano L, Brambilla L, et al. (2000) Development, reliability and acceptability of a new version of the DSM-IV Social and Occupational Functioning Assessment Scale (SOFAS) to assess routine social functioning. *Acta Psychiatrica Scandinavica* 101: 323-329.
- Mueser K, Gingerich S, Salyers M, et al. (2004) The Illness Management and Recovery Scale (IMR). Lebanon, NH: New Hampshire-Dartmouth Psychiatric Research Center.
- Mueser K, Meyer P, Penn D, et al. (2006) The Illness Management and Recovery program: Rationale, development, and preliminary findings. *Schizophrenia Bulletin* 32: 32-43.

- Mulkern V and Manderscheid R. (1989) Characteristics of community support program clients in 1980 and 1984. *Hospital and Community Psychiatry* 40: 165-172.
- Nafees B, van Hanswijck de Jonge P, Stull D, et al. (2012) Reliability and validity of the Personal and Social Performance scale in patients with schizophrenia. *Schizophrenia Research* 140: 71-76.
- Nasrallah H, Morosini P and Gagnon DD. (2008) Reliability, validity and ability to detect change of the Personal and Social Performance scale in patients with stable schizophrenia. *Psychiatry Research* 161: 213-224.
- National Mental Health Information Development Expert Advisory Panel (NHMIDEAP). (2013) Mental Health National Outcomes and Casemix Collection: NOCC Strategic Directions 2014-2024. Canberra: Commonwealth of Australia.
- National Mental Health Information Development Expert Advisory Panel (NHMIDEAP). (2015) National Outcomes and Casemix Collection (NOCC): Domain Framework. Canberra: Commonwealth of Australia.
- Newnham E, Harwood K and Page A. (2009) The subscale structure of the Health of the Nation Outcome Scales. *Journal of Mental Health* 18: 326-334.
- Nicholl D, Nasrallah H, Nuamah I, et al. (2010) Personal and social functioning in schizophrenia: defining a clinically meaningful measure of maintenance in relapse prevention. *Current Medical Research and Opinion* 26: 1471-1484.
- Norman RMG, Malla AK, McLean T, et al. (2000) The relationship of symptoms and level of functioning in schizophrenia to general wellbeing and the Quality of Life Scale. *Acta Psychiatrica Scandinavica* 102: 303-309.
- Oiesvold T, Bakkejord T and Sexton JA. (2011) Concurrent validity of the Health of the Nation Outcome Scales compared with a patient-derived measure, the Symptom Checklist-90-Revised in out-patient clinics. *Psychiatry Research* 187: 297-300.
- Orrell M, Yard P, Handysides J, et al. (1999) Validity and reliability of the Health of the Nation Outcome Scales in psychiatric patients in the community. *British Journal of Psychiatry* 174: 409-412.
- Page AC, Hooke GR and Rutherford EM. (2001) Measuring mental health outcomes in a private psychiatric clinic: Health of the Nation Outcome Scales and Medical Outcomes Short Form SF-36. *Australian and New Zealand Journal of Psychiatry* 35: 377-381.
- Parker G and Hadzi-Pavlovic D. (1995) The capacity of a measure of disability (the LSP) to predict hospital readmission in those with schizophrenia. *Psychological Medicine* 25: 157-163.
- Parker G, O'Donnell M, Hadzi-Pavlovic D, et al. (2002) Assessing outcome in community mental health patients: A comparative analysis of measures. *International Journal of Social Psychiatry* 48: 11-19.
- Parker G, Rosen A, Emdur N, et al. (1991) The Life Skills Profile: Psychometric properties of a measure assessing function and disability in schizophrenia. *Acta Psychiatrica Scandinavica* 83: 145-152.
- Parker G, Rosen A, Trauer T, et al. (2007) Disability associated with mood states and comparator conditions: Application of the Life Skills Profile measure of disability. *Bipolar Disorders* 9: 11-15.
- Patrick DL, Burns T, Morosini P, et al. (2009) Reliability, validity and ability to detect change of the clinician-rated Personal and Social Performance scale in patients with acute symptoms of schizophrenia. *Current Medical Research and Opinion* 25: 325-338.
- Peterson L. (1983) Social functioning assessment of aftercare psychiatric patients in socialization therapy. *Psychological Reports* 53: 1123-1130.
- Phelan M, Slade M, Thornicroft G, et al. (1995) The Camberwell Assessment of Need: the validity and reliability of an instrument to assess the needs of people with severe mental illness. *British Journal of Psychiatry* 167: 589-595.

- Phuaphanprasert B, Srisurapanont M, Silpakit C, et al. (2007) Reliability and validity of the Thai version of the Health of the Nation Outcome Scales (HoNOS). *Journal of the Medical Association of Thailand* 90: 2487-2493.
- Pirkis JE, Burgess PM, Kirk PK, et al. (2005) A review of the psychometric properties of the Health of the Nation Outcome Scales (HoNOS) family of measures. *Health and Quality of Life Outcomes* 3: 76.
- Preston NJ. (2000a) The Health of the Nation Outcome Scales: Validating factorial structure and invariance across two health services. *Australian and New Zealand Journal of Psychiatry* 34: 512-519; discussion 520-511.
- Preston NJ. (2000b) Predicting community survival in early psychosis and schizophrenia populations after receiving intensive case management. *Australian and New Zealand Journal of Psychiatry* 34: 122-128.
- Preti A, Pisano A, Cascio MT, et al. (2012) Validation of the Health of the Nation Outcome Scales as a routine measure of outcome in early intervention programmes. *Early Intervention in Psychiatry* 6: 423-431.
- Prouteau A, Verdoux H, Briand C, et al. (2005) Cognitive predictors of psychosocial functioning outcome in schizophrenia: A follow-up study of subjects participating in a rehabilitation program. *Schizophrenia Research* 77: 343-353.
- Prowse L and Coombs T. (2009) The use of the Health of the Nation Outcome Scales (HoNOS) to inform discharge and transfer decisions in community mental health services. *Australian Health Review* 33: 13-18.
- Rosen A, Hadzi-Pavlovic D and Parker G. (1989) The life skills profile: A measure assessing function and disability in schizophrenia. *Schizophrenia Bulletin* 15: 325-337.
- Rosen A and Teesson M. (2001) Does case management work? The evidence and the abuse of evidence-based medicine. *Australian and New Zealand Journal of Psychiatry* 35: 731-746.
- Rosen A, Trauer T, Hadzi-Pavlovic D, et al. (2001) Development of a brief form of the Life Skills Profile: The LSP-20. *Australian and New Zealand Journal of Psychiatry* 35: 677-683.
- Salyers M, Godfrey J, Mueser K, et al. (2007) Measuring illness management outcomes: A psychometric study of clinician and consumer rating scales for illness self management and recovery. *Community Mental Health Journal* 43: 459-480.
- Salyers MP, Godfrey JL, McGuire AB, et al. (2009) Implementing the illness management and recovery program for consumers with severe mental illness. *Psychiatric Services* 60: 483-490.
- Sanderson K, Issakidis C, Johnston S, et al. (1996) Cost-effectiveness of Intensive Case Management for People with Serious Mental Illness. Sydney: South Eastern Sydney Area Mental Health Service.
- Schaub D, Brune M, Bierhoff H-W, et al. (2012) Comparison of self- and clinician's ratings of personal and social performance in patients with schizophrenia: The role of insight. *Psychopathology* 45: 109-116.
- Schaub D, Brune M, Jaspert E, et al. (2011) The illness and everyday living: Close interplay of psychopathological syndromes and psychosocial functioning in chronic schizophrenia. *European Archives of Psychiatry and Clinical Neuroscience* 261: 85-93.
- Schneider J, Wooff D, Carpenter J, et al. (2002) Service organisation, service use and costs of community mental health care. *Journal of Mental Health Policy and Economics* 5: 79-87.
- Shergill SS, Shankar KK, Seneviratna K, et al. (1999) The validity and reliability of the Health of the Nation Outcome Scales (HoNOS) in the elderly. *Journal of Mental Health* 8: 511-521.
- Simon AE, Giacomini V, Ferrero F, et al. (2003) Dysexecutive syndrome and social adjustment in schizophrenia. *Australian and New Zealand Journal of Psychiatry* 37: 340-346.
- Sklar M, Sarkin A, Gilmer T, et al. (2012) The psychometric properties of the Illness Management and Recovery scale in a large American public mental health system. *Psychiatry Research* 199: 220-227.

- Speak B and Muncer S. (2015) The structure and reliability of the Health of the Nation Outcome Scales. *Australasian Psychiatry* 23: 66-68.
- Speak BL and Muncer SJ. (2016) Factorial structure of the Health of the Nation Outcome Scales: An ordinal confirmatory factor analysis using a national sample of clinician ratings in England. *International Journal of Mental Health Nursing* 25: 87-98.
- Srisurapanont M, Arunpongpaissal S, Chuntaruchikapong S, et al. (2008) Cross-cultural validation and inter-rater reliability of the Personal and Social Performance scale, Thai version. *Journal of the Medical Association of Thailand* 91: 1603-1608.
- Stedman T, Yellowlees P, Mellsop G, et al. (1997) Measuring Consumer Outcomes In Mental Health: Field Testing of Selected Measures of Consumer Outcome in Mental Health. Canberra: Department of Health and Family Services.
- Suzuki T, Uchida H, Sakurai H, et al. (2015) Relationships between global assessment of functioning and other rating scales in clinical trials for schizophrenia. *Psychiatry Research* 227: 265-269.
- Taylor JR and Wilkinson G. (1997) HoNOS v. GP opinion in a shifted out-patient clinic. *Psychiatric Bulletin* 21: 483-485.
- Tessler R and Goldman H. (1982) *The Chronically Mentally Ill: Assessing Community Support Programs*, Cambridge, Massachusetts: Ballinger.
- Tianmei S, Liang S, Yun'ai S, et al. (2011) The Chinese version of the Personal and Social Performance Scale (PSP): Validity and reliability. *Psychiatry Research* 185: 275-279.
- Trauer T. (1999) The subscale structure of the Health of the Nation Outcome Scales (HoNOS). *Journal of Mental Health* 8: 499-509.
- Trauer T. (2001) Symptom severity and personal functioning among patients with schizophrenia discharged from long-term hospital care into the community. *Community Mental Health Journal* 37: 145-155.
- Trauer T. (2003) Analysis of Outcome Measurement Data from the Four Victorian 'Round One' Agencies. Mental Health Branch, Department of Human Services, Victoria.
- Trauer T, Callaly T, Hantz P, et al. (1999) Health of the Nation Outcome Scales. Results of the Victorian field trial. *British Journal of Psychiatry* 174: 380-388.
- Trauer T, Duckmanton RA and Chiu E. (1995) The Life Skills Profile: a study of its psychometric properties. *Australian and New Zealand Journal of Psychiatry* 29: 492-499.
- Trauer T, Duckmanton RA and Chiu E. (1997) The assessment of clinically significant change using the Life Skills Profile. *Australian and New Zealand Journal of Psychiatry* 31: 257-263.
- Trauer T, Duckmanton RA and Chiu E. (1998) Estimation of costs of public psychiatric treatment. *Psychiatric Services* 49: 440-442.
- Trauer T and Eagar K. (2004) New Zealand Mental Health Consumers and their Outcomes. Auckland: Health Research Council of New Zealand.
- Waikato Evaluation Team. (2005) Functional Outcome Measures for Children and Youth, Adults and Older Adult Consumers in New Zealand Mental Health Services. Hamilton, NZ: University of Waikato.
- Webster J, Bretherton F, Goulter NS, et al. (2013) Does an educational intervention improve the usefulness of the Health of the Nation Outcome Scales in an acute mental health setting? *International Journal of Mental Health Nursing* 22: 322-328.
- Weiner HR. (1993) Multi-Function Needs Assessment: The development of a functional assessment instrument. *Psychosocial Rehabilitation Journal* 16: 51-61.
- Widlak PA, McKee D, Greenberg JR, et al. (1992) An assessment of client function scales in the Uniform Client Data Instrument (UCDI). *Psychosocial Rehabilitation Journal* 15: 19-35.
- Wing J, and Lelliott, P. (1999a) Reliability and validity of HoNOS [Correspondence]. *Psychiatric Bulletin* 23: 375.
- Wing J, and Lelliott, P. (1999b) Reliability and validity of HoNOS. *Psychiatric Bulletin* 23: 375.
- Wing J, Curtis RH and Beevor A. (1999) Health of the Nation Outcome Scales (HoNOS). Glossary for HoNOS score sheet. *British Journal of Psychiatry* 174: 432-434.

- Wing JK, Beevor AS, Curtis RH, et al. (1998) Health of the Nation Outcome Scales (HoNOS). Research and development. *British Journal of Psychiatry* 172: 11-18.
- Wooff D, Schneider J, Carpenter J, et al. (2003) Correlates of stress in carers. *Journal of Mental Health* 12: 29-40.
- World Health Organization. (2001) International Classification of Functioning, Disability and Health (ICF). Geneva: World Health Organization.
- Zani B, McFarland B, Wachal M, et al. (1999) Statewide replication of predictive validation for the Multnomah Community Ability Scale. *Community Mental Health Journal* 35: 223-229.

Table 1: Profile of clinician-rated instruments designed to assess functioning

INSTRUMENT	DATE^a	COUNTRY	DESCRIPTION
Disability Rating Form (DRF)	1992	United States	Designed to rate disabilities associated with severe mental illness. Five items measure five areas of disability: <i>activity of daily living; social functioning; concentration and task performance; adaptation to change; and impulse control</i> (Hoyle et al., 1992; 1993).
FACE Core Assessment	1994	United Kingdom	Developed for use in adult mental health services. 50 items cover domains of: behaviour; cognitive; mental health; physical wellbeing; activities of daily living; interpersonal relationships; social circumstances; and response to care. Also includes a global rating of the impact of consumers' problems on their quality of life and/or functioning in the past month (Clifford; 1994; 1997b; 1997a; Clifford et al., 1999).
Global Assessment of Functioning (GAF)	1987	United States	Introduced in Diagnostic and Statistical Manual, revised third edition (DSM-III-R) as a means of assessing 'adaptive functioning' (American Psychiatric Association, 1987). A 100-point, single item scale assesses three dimensions of functioning (<i>psychological; social; and occupational</i>). Can yield a single score (where only the most severe of the symptom and functioning values are recorded) or separate scores for symptoms (GAF-S) and functioning (GAF-F).
Health of the Nation Outcomes Scales (HoNOS)	1998	United Kingdom	Designed for routine use by clinicians to measure consumer outcomes (Wing, 1999b; 1999a; Wing et al., 1999; Wing et al., 1998). Can be regarded a general measure of mental health and social functioning in people with a mental illness (Wing et al., 1998). 12 items roll up into four subscales (<i>behaviour, impairment, symptoms, and social</i>). The <i>social</i> subscale contains four items (Items 9-12) covering: <i>relationships; activities of daily living; living conditions; and occupation and activities</i> .
Illness Management and Recovery Scale - Clinician Version (IMRS-C)	2004	United States	Developed to assess outcomes for participants the widely-used IMR program (teaches people with schizophrenia illness self-management strategies) (Gingerich and Mueser, 2002; Mueser et al., 2006). 15 items aggregate into three scales: <i>recovery (progress towards goals; knowledge; contact with people outside of family; relapse prevention planning; involvement with self-help activities); management (symptom distress; impairment of functioning; relapse of symptoms; psychiatric hospitalisations; coping); and biology (using medication effectively; alcohol use; drug use)</i> (Mueser et al., 2004).
Level of Functioning Scale (LFS)	1989	United States	Developed through a factor analysis of 73 items on the 79-item Missouri Level of Care (MLC) that measure functioning. Factors are: <i>community skills; self-care skills; nuisance behaviour; sociability; skilled nursing; proclivity for violence; and control of anger</i> (Massey et al., 1989).
Life Functioning Assessment Inventory (L-FAI)	2013	United States	Designed to measure functioning in consumers with psychosis. Assesses four life domains: <i>work; social; relationships; leisure; and homemaking</i> . Each domain is given a status score (reflecting general performance) and a grade score (reflecting a more specific performance level within the grade) (Hui et al., 2013).
Life Skills Profile (LSP-16)	1989	Australia	A short-form of the LSP-39 which was designed to measure constructs relevant to survival and adaptation in the community for individuals with schizophrenia and chronic mental illness (Parker et al., 1991; Rosen et al., 1989). LSP-16 was created to minimise the rating burden on clinicians participating in the Australian Mental Health Classification and Service Costs (MH-CASC) Project (Buckingham et al., 1998b; 1998a; Burgess et al., 1999). 16 items measure: <i>withdrawal; self-care; compliance; and anti-social behaviour</i> .
Mini-ICF-APP	2009	Germany	Created with reference to the International Classification of Functioning, Disability and Health (ICF), initially in German (Linden et al., 2009) and then in several other languages, including English (Molodynski et al., 2013). Assesses 13 domains: <i>adherence to regulations; planning and structuring of tasks; flexibility; competency; endurance; assertiveness; contact with others; public exposure; intimacy; non-work activities; self-maintenance; mobility; and competence to judge and decide</i> .
Multi-Function Needs	1982	United States	Developed to assess the service needs and general functional performance of consumers in a single psychiatric hospital (Angelini, 1982). 118 items cover 13 areas of functioning: <i>physical self-maintenance; physical health; substance</i>

INSTRUMENT	DATE ^a	COUNTRY	DESCRIPTION
Assessment (MFNA)			<i>abuse; motor behaviour; psychiatric symptoms; attitude and motivation; attention and memory; verbal communication; family interaction; social interaction; independent living skills; public behaviour; and work/school/leisure</i> (Weiner, 1993).
Multnomah Community Ability Scale (MCAS)	1994	United States	Assesses the level of functioning of consumers with chronic mental illness living in the community. 17 items aggregate into four subscales: <i>interference with functioning; adjustment to living; social competence; and behaviour problems</i> (Barker et al., 1994a; Barker et al., 1994b; Barker et al., 1994c).
Need of Support and Service Questionnaire (NSSQ)	2005	Sweden	Developed for a study examining whether mental health staff and social services staff were consistent in their judgements of consumers' needs. Of 33 items, 23 relate to three domains of need: <i>need of support in activities of daily living; need of service provided by the public health and social service sectors; and need of assisted care living and need of work</i> (Jansson et al., 2005).
Personal and Social Performance Scale (PSP)	2000	Italy	Developed as part of a package for planning and evaluating psychiatric rehabilitation. A 100-point, single item rating scale yields a rating based on four main areas: <i>socially useful activities, including work and study; personal and social relationships; self-care; and disturbing and aggressive behaviours</i> (Morosini et al., 2000).
Profile of Community Psychiatry Clients (PCPC)	1998	Australia	Developed to measure common problems and probable needs experienced by consumers in the community. Designed for use in screening, service quality assurance, and research. 35 items measure domains of: <i>coping limitations; behavioural problems; levels of social support; and organic problems</i> (Cheah et al., 1998).
Rehabilitation Evaluation Hall and Baker (REHAB)	1984	United Kingdom	A measure of socially appropriate or adaptive behaviour, designed for use with people with chronic mental illness. Of 23 items, seven form a <i>deviant behaviour</i> subscale and 16 form a general behaviour subscale. Covers the following areas: <i>social activity; disturbed speech; communication skills; self-care skills; and community skills</i> (Baker and Hall, 1984; 1988).
Residential Competency Scale (RCS)	1989	Canada	Designed to assess the community living skills of consumers residing in community residential support services following deinstitutionalisation. 82 items measure skills in: <i>community skills; self-care; friendship; consideration; social competence; clarifying communication; money management; self-control; meal preparation; leisure; time routine; independence; and time planning</i> (Kazarian et al., 1989).
Social Adjustment Behavior Rating Scale (SABRS)	1962	United States	Designed to measure two aspects of consumers' social adjustment – work level and socialisation level. Of 61 items, 29 relate to <i>work</i> and 33 to <i>socialisation</i> , with one overlapping item (Aumack, 1962).
Social Functioning Index (SFI)	1983	United States	Designed to measure social skills in consumers receiving community-based care, post-hospitalisation. 51 items measure: <i>energy; self-control; hygiene; communication; and awareness of the environment</i> (Peterson, 1983).
Social Occupational Functioning Assessment Scale (SOFAS)	1994	United States	Intended to assess social and occupational functioning independently of the severity of psychological symptoms. Provided in DSM-IV as an Axis V measure (American Psychiatric Association, 1994). A one-item rating of consumer functioning scored 0-100.
Uniform Client Data Instrument (UCDI)	1982	United States	Designed to enable standardised information to be collected for consumers of mental health services in the United States (Mulkern and Manderscheid, 1989; Tessler and Goldman, 1982). Covers multiple domains, including two explicit measures of functioning: <i>community living skills</i> (10 items); and <i>social activities</i> (8 items) (Widlak et al., 1992).

a. Refers to the date of published information on the original version of the instrument.

Table 2: Psychometric properties of instruments meeting Criteria 1-5

INSTRUMENT	PSYCHOMETRIC PROPERTIES	
Health of the Nation Outcome Scales (HoNOS) ^a	Construct validity	Some studies support the four factor model defined by the original HoNOS subscales (Preston, 2000a), while others suggest alternative structures (Andreas et al., 2010; Lovaglio and Monzani, 2011; Lovaglio and Monzani, 2012; McClelland et al., 2000; Newnham et al., 2009; Speak and Muncer, 2015; Speak and Muncer, 2016; Trauer, 1999). Generally, the structure of the <i>social</i> subscale is supported, however Items 11 and 12 make a lesser contribution to the HoNOS total score than other items (Andreas et al., 2010; Lovaglio and Monzani, 2011; McClelland et al., 2000).
	Concurrent validity	In some studies, the HoNOS has correlated well with other clinician-rated functioning measures, including the Camberwell Assessment of Need Short Appraisal Schedule (CANSAS) (Issakidis and Teesson, 1999), Role Functioning Scale (RFS) (Wing et al., 1998), Global Assessment of Functioning (GAF) (Amin et al., 1999; Browne et al., 2000; McClelland et al., 2000; Orrell et al., 1999; Parker et al., 2002; Phuaphanprasert et al., 2007; Preti et al., 2012; Shergill et al., 1999), Life Skills Profile (LSP) (Parker et al., 2002), and Disability Assessment Schedule (DAS) (Amin et al., 1999). Some studies report low correlations between the HoNOS functioning items and other measures of functioning (e.g., one study found no correlation between Item 9 (<i>relationships</i>) and the social communication scale on the Beeinträchtigungs-Schwere-Score (BSS), and no correlation between Item 10 (<i>activities of daily living</i>) and the GAF) (Andreas et al., 2010). The HoNOS has been found to discriminate between groups of consumers on the basis of treatment received, which may reflect levels of functioning - for example, standard case management versus assertive case management (Gallagher and Teesson, 2000), residential/nursing home, day patient, outpatient and inpatient settings (Hope et al., 1998; Kisely et al., 2007; Lovaglio and Monzani, 2011; Orrell et al., 1999; Shergill et al., 1999), acute and sub-acute settings (Phuaphanprasert et al., 2007), and consumers in long-stay settings with low, medium and high expectations of discharge (Allan and McGonagle, 1997).
	Predictive validity	In several studies the HoNOS has demonstrated good predictive validity, explaining a significant proportion of the variance in resource use (e.g., service contacts, length of stay, and costs) and treatment outcome (e.g., readmission rates, retention in the community, treatment response, and death) (Broadbent, 2001; Hope et al., 1998; Kisely et al., 2010; Parker et al., 2002; Schneider et al., 2002), and/or being an indicator of discharge or transfer decisions (Prowse and Coombs, 2009). In other studies, little or no correspondence reported between the HoNOS (total score and <i>social</i> subscale) and treatment outcome (Preti et al., 2012) or resource use (Goldney et al., 1998).
	Internal consistency	The HoNOS has demonstrated moderately high internal consistency and low levels of item redundancy (Cronbach's alphas 0.59 to 0.76), however the <i>social</i> subscale does not always perform as strongly as the <i>behaviour</i> and <i>symptoms</i> subscales (Andreas et al., 2010; Lovaglio and Monzani, 2011; McClelland et al., 2000; Oiesvold et al., 2011; Orrell et al., 1999; Page et al., 2001; Phuaphanprasert et al., 2007; Shergill et al., 1999; Stedman et al., 1997; Trauer, 1999; Wing et al., 1998).
	Inter-rater reliability^b	The HoNOS has shown fair to moderate (Bebbington et al., 1999; Brooks, 2000; Shergill et al., 1999) or moderate to good inter-rater reliability (Amin et al., 1999; Andreas et al., 2007; Hope et al., 1998; Orrell et al., 1999; Phuaphanprasert et al., 2007; Wing et al., 1998). There are exceptions (Webster et al., 2013), however, and agreement tends to be poor on particular items, including Item 9 (<i>relationships</i>) (Orrell et al., 1999), Item 11 (<i>living conditions</i>) (Orrell et al., 1999; Trauer et al., 1999), and Item 12 (<i>occupation and activities</i>) (Bebbington et al., 1999; Trauer et al., 1999; Wing et al., 1998).
	Test-retest reliability^b	The HoNOS has generally shown fair to moderate, or good to very good, test-retest reliability (Andreas et al., 2007; Brooks, 2000; Orrell et al., 1999; Preti et al., 2012; Shergill et al., 1999) although some items perform less well than others, including Item 10 (<i>activities of daily living</i>).
	Sensitivity to change	Some studies have examined change in HoNOS over time in given settings, hypothesising that there should be a decrease in severity as the consumer nears the end of an episode. This hypothesis has generally been supported (Andreas et al., 2007; Andreas et al., 2010; Audin et al., 2001; Egger et al., 2015; Goldney et al., 1998; Kisely et al., 2007; Kisely et al., 2010; McClelland et al., 2000; Preti et al., 2012; Trauer et al., 1999), although some studies have shed doubt on the ability of Item 11 (<i>living conditions</i>) and Item 12 (<i>occupation and activities</i>) to measure change (Andreas et al., 2007; Andreas et al., 2010; Kisely et al., 2010; McClelland et al., 2000). Other studies have judged HoNOS sensitivity to change against 'gold standards'. These studies have

INSTRUMENT	PSYCHOMETRIC PROPERTIES	
		found correlations between HoNOS and clinicians' assessments of improvement or deterioration (Gallagher and Teesson, 2000; Taylor and Wilkinson, 1997) and consumers' self-reported goal attainment (Hunter et al., 2004). Still other studies have compared the HoNOS's capacity to detect change against established measures, including the GAF (McClelland et al., 2000) predominantly for the <i>behaviour</i> and <i>symptoms</i> subscales (Bech et al., 2003). Finally, one study sought clinical expert opinion about the ability of the HoNOS to detect change; clinicians believed that the HoNOS was sensitive to change, but that the <i>social</i> and <i>impairment</i> subscales were less responsive than the <i>behaviour</i> and <i>symptoms</i> subscales (Burgess et al., 2009).
Illness Management and Recovery Scale - Clinician Version (IMRS-C)	Construct validity	Several studies using Rasch analysis or confirmatory factor analysis support the IMRS-C factors of <i>recovery</i> , <i>management</i> and <i>biological vulnerability</i> (Hasson-Ohayon et al., 2008), although one study suggests that the first two of these perform better than the third (McGuire et al., 2014) and another suggests some modifications, including replacement of the <i>biological vulnerability</i> factor with a <i>substance use</i> factor (Sklar et al., 2012).
	Concurrent validity	The IMRS-C has demonstrated good concurrent validity, showing strong negative correlations with clinicians' assessments of symptomatology (Färdig et al., 2011), and strong positive correlations with clinicians' contemporaneous assessments of progress towards employment, education and housing goals (Sklar et al., 2012) and other clinician-rated instruments, including the Multnomah Community Ability Scale (MCAS) (Salyers et al., 2007), Psychosis Evaluation Tool for Common Use by Caregivers (PECC) (Färdig et al., 2011), and Substance Abuse Treatment Scale-Revised (SATS-R) (Salyers et al., 2007; Sklar et al., 2012). IMRS-C also correlates well with the consumer-rated IMRS (Färdig et al., 2011; Hasson-Ohayon et al., 2008; Salyers et al., 2007), and with other consumer-rated instruments like the Coping Efficacy Scale (CES) (Hasson-Ohayon et al., 2008), Multidimensional Scale for Perceived Social Support (MSPSS) (Hasson-Ohayon et al., 2008), Manchester Short Assessment of Quality of Life (MANSA) (Färdig et al., 2011), Recovery Assessment Scale (RAS) (Färdig et al., 2011), and Modified Colorado Symptom Index (MCSI) (Färdig et al., 2011).
	Predictive validity	The IMRS-C has demonstrated good predictive validity, showing strong positive correlations with clinicians' subsequent assessments of progress towards employment, education and housing goals (Sklar et al., 2012).
	Internal consistency	The IMRS-C has generally demonstrated good internal consistency (Cronbach's alpha 0.70 to 0.82) (Färdig et al., 2011; Hasson-Ohayon et al., 2008; Salyers et al., 2007; Sklar et al., 2012).
	Inter-rater reliability^b	No information available.
	Test-retest reliability^b	The IMRS-C has demonstrated strong test-retest reliability over a period of two weeks ($r=0.81-0.88$) (Färdig et al., 2011; Salyers et al., 2007).
	Sensitivity to change	Limited information about the IMRS-C's sensitivity to change, but evaluations of the IMR program have demonstrated outcomes in the expected direction (Hasson-Ohayon et al., 2007; Salyers et al., 2009).
Life Skills Profile (LSP-16) ^c	Construct validity	The original five subscales of the LSP-39 identified by the instrument's developers (<i>communication</i> , <i>social contact</i> , <i>non-turbulence</i> , <i>self-care</i> and <i>responsibility</i>) have undergone subsequent testing using principal components analysis and confirmatory factor analysis, resulting in proposals for alternative subscale structures. One study, for example, suggested alternative subscales of <i>bizarre</i> , <i>withdrawal</i> , <i>self-care</i> , <i>compliance</i> and <i>anti-social behaviour</i> (Trauer et al., 1995), and another recommended that the subscales be further divided into the two dimensions of <i>general impairment</i> and <i>difficulty</i> (Andrews et al., 1990). The LSP-16's four subscales (<i>self-care</i> , <i>antisocial</i> , <i>withdrawal</i> and <i>compliance</i>) have been tested; one study using multilevel confirmatory factor analysis demonstrated that the four-factor model was imperfect, and that a 15-item version fit the data better (Little, 2013).
	Concurrent validity	The LSP-39 has been shown to perform well against the HoNOS (Parker et al., 2002; Stedman et al., 1997; Trauer and Eagar, 2004; Wooff et al., 2003; Eagar et al., 2005), Katz Adjustment Scale (KAS) (Parker et al., 1991), MCAS (Dickinson and Coursey, 2002), Strauss-Carpenter Levels of Functioning Scale (LOF) (Dickinson and Coursey, 2002), GAF (Dickinson and Coursey, 2002; Parker et al., 2002; Simon et al., 2003), RFS (Stedman et al., 1997), Quality of Life Scale (QOL) (Norman et al., 2000), Interviewer-rated Quality of Life Scale (IQL) (Wooff et al., 2003), Social Behaviour Schedule (SBS) (Simon et al., 2003), Resource Associated Functional Level Scale (RAFLS) (Trauer et al., 1995), and Global Assessment Scale (GAS) (Wooff et al., 2003). However, it has demonstrated poor or mixed performance against the BASIS-32 [®] (Stedman et al., 1997), Mental Health Inventory (MHI) (Stedman et al., 1997), Short Form-36 (SF-36) (Stedman et al.,

INSTRUMENT	PSYCHOMETRIC PROPERTIES	
		1997), General Wellbeing Scale (GWB) (Norman et al., 2000; Trauer et al., 1998), Brief Psychiatric Rating Scales (BPRS) (Rosen et al., 1989; Trauer et al., 1995; Wooff et al., 2003), Dysexecutive Questionnaire (DEX) (Simon et al., 2003; Wooff et al., 2003), Cantril's Ladder (Wooff et al., 2003), and Affect Balance Scale (ABS) (Wooff et al., 2003). LSP-16 shown to correlate well with the LSP-39 (Buckingham et al., 1998b; Rosen et al., 2001), and HoNOS (Trauer, 2003) but has demonstrated poor or mixed performance against the Behaviour and Symptom Identification Scale (BASIS-32 [®]) (Trauer, 2003). The LSP-39 and LSP-16 shown to discriminate between consumers on the basis of the stability and independence of their living situations (Andrews et al., 1990; Browne and Courtney, 2004; Keller and Hayes, 1998; Kirkby et al., 1997; Trauer et al., 1998; Trauer et al., 1997), levels of social functioning (e.g., unstable employment, welfare dependency, police contact, and complaints by neighbours) (Andrews et al., 1990; Parker and Hadzi-Pavlovic, 1995; Rosen et al., 1989), legal status (Eagar et al., 2005), and diagnosis (Eagar et al., 2005; Parker et al., 2007).
	Predictive validity	Generally, studies have shown the LSP-39 to predict outcomes relating to community tenure (Preston, 2000b), hospital readmission (Andrews et al., 1990; Parker and Hadzi-Pavlovic, 1995), change in locus of care (Trauer et al., 1997), and overall costs (Trauer et al., 1998), although one study reported discrepant findings (Parker et al., 2002). LSP-16 shown to predict outcomes related to length of inpatient stay and overall costs (Kisely et al., 2000).
	Internal consistency	The LSP-39 has demonstrated moderately high internal consistency, with subscale and total score Cronbach's alphas 0.64 to 0.88 (Parker et al., 1991; Rosen et al., 1989; Stedman et al., 1997; Trauer et al., 1995), and 0.93 to 0.94 (Dickinson and Coursey, 2002; Stedman et al., 1997; Trauer et al., 1995), respectively.
	Inter-rater reliability^b	Inter-rater reliability has been shown to be fair to moderate (Andrews et al., 1994; Stedman et al., 1997; Trauer et al., 1995) or moderate to good (Parker et al., 1991; Rosen et al., 1989; Trauer et al., 1995) for the LSP-39, and moderate to good for the LSP-16 (Rosen et al., 2001).
	Test-retest reliability^b	LSP-39 has shown high test-retest reliability (Andrews et al., 1994; Parker et al., 1991; Stedman et al., 1997). One study established high test-retest reliability for the LSP-16 (Rosen et al., 2001).
	Sensitivity to change	Some studies have reported significant associations between changes on the LSP-39 and changes on established measures such as the Global Change Ratings Scale (GCRS) (Stedman et al., 1997), the Modified Clinical Global Impressions Scale (CGI) (Stedman et al., 1997), the RFS (Stedman et al., 1997), the HoNOS (Parker et al., 2002; Stedman et al., 1997), and the GAF (Parker et al., 2002). Other studies have compared changes in LSP-39 scores for different consumer groups that would be expected to show greater or lesser improvement depending on their treatment circumstances. Typically, the LSP-39 demonstrates greater improvement for those in intensive case management versus those in routine case management (Craig et al., 2004; Hambridge and Rosen, 1994; Hamernik and Pakenham, 1999; Johnston et al., 1998; Rosen and Teesson, 2001; Sanderson et al., 1996), but there have been some exceptions (Ford et al., 2001; Ford et al., 1997). Still other studies have compared changes in LSP-39 scores with consumer self-reported improvement or deterioration as the 'gold standard'. One study found that LSP-39 scores worsened in the group who reported a decline in their levels of functioning, but no association in groups with other levels and directions of self-reported change (Stedman et al., 1997).
Multnomah Community Ability Scale (MCAS)	Construct validity	Tests of the factor structure of the MCAS have suggested that the developers' four factor model (<i>interference with functioning; adjustment to living; social competence; and behaviour problems</i>) (Barker et al., 1994a; Barker et al., 1994b; Barker et al., 1994c) may not be the best solution, and alternative structures have been proposed (Bassani et al., 2009; Corbiere et al., 2002; Hendryx et al., 2001).
	Concurrent validity	The MCAS has been shown to perform well against clinicians' global assessments (Barker et al., 1994b) and consumers' levels of resource use (Barker et al., 1994b; Barker et al., 1994c), and to discriminate between consumers in meaningful ways (e.g., on age-related factors, severity of symptoms and cognitive functioning) (Barker et al., 1994b; Barker et al., 1994c; Prouteau et al., 2005). MCAS shown to correlate with similar measures, including the Client Satisfaction Questionnaire (CSQ) (Hendryx et al., 2001), the physical health scale of the SF-36 (Hendryx et al., 2001), Lehman Quality of Life Scale (LQLS) (Hendryx et al., 2001), Social Occupational Functioning Assessment Scale (SOFAS) (Hendryx et al., 2001), BPRS (Brown et al., 2003), and Positive and Negative Syndrome Scale (PANSS) (Trauer, 2001), but not the RAS (Lavin and Ryan, 2012).
	Predictive validity	The MCAS has demonstrated good predictive validity, with poorer scores associated with subsequent hospitalisations (Barker et al., 1994b; Zani et al., 1999).

INSTRUMENT	PSYCHOMETRIC PROPERTIES	
	Internal consistency	Good internal consistency reported by the MCAS developers (Cronbach's alphas up to 0.90) (Barker et al., 1994b), but lower levels reported by others using both the original structure and revised structures (Bassani et al., 2009; Corbiere et al., 2002; Hendryx et al., 2001).
	Inter-rater reliability^b	The MCAS has shown good inter-rater reliability, with intra-class correlation coefficients of 0.85 during development (Barker et al., 1994b) and 0.62 to 0.99 in subsequent studies (Dickerson, 1997; Dickerson et al., 2003; Trauer, 2001).
	Test-retest reliability^b	MCAS shown to perform well on inter-rater reliability, with an intra-class correlation coefficient of 0.83 (Barker et al., 1994b).
	Sensitivity to change	The MCAS's sensitivity to change has not been formally tested. In several evaluation studies, the MCAS has shown improvements in the expected direction (Hopkins and Ramsundar, 2006; McDevitt et al., 2005) and has been associated with commensurate levels of change on other relevant instruments, including the BPRS (Adair et al., 2005).
Personal and Social Performance Scale (PSP)	Construct validity	The factor structure proposed by the PSP developers (<i>socially useful activities, including work and study; personal and social relationships; self-care; and disturbing and aggressive behaviours</i>) has been confirmed in subsequent independent analysis (Kawata and Revicki, 2008).
	Concurrent validity	The PSP has demonstrated strong correlations with clinician-rated and consumer-rated instruments assessing functioning and related constructs, including the GAF (Apiquian et al., 2009; Brissos et al., 2012; Juckel et al., 2008; Nafees et al., 2012; Schaub et al., 2011; Tianmei et al., 2011), SOFAS (Garcia-Portilla et al., 2011; Juckel et al., 2008; Schaub et al., 2011), Mini-ICF-APP (Juckel et al., 2008; Schaub et al., 2011), Strauss-Carpenter Level of Functioning (SCLF) (Nasrallah et al., 2008), Activities of Daily Living Rating Scale II (ADLRS-II) (Hsieh et al., 2011), and Quality of Life Scale (QLS) (Kawata and Revicki, 2008; Nafees et al., 2012). Correlations with a consumer-rated version of the PSP were marginal, but mediated by consumers' levels of insight (Schaub et al., 2012). The PSP has been shown to discriminate between consumers on the basis of treatment setting (e.g., inpatient versus community) (Apiquian et al., 2009; Patrick et al., 2009; Schaub et al., 2011), diagnosis (Garcia-Portilla et al., 2011), and neurocognitive capacity and symptom levels, as measured by the Wechsler Memory Scale-Revised (WMS-R), Continuous Performance Test (CPT), Wisconsin Card Sorting Test (WCST), PANSS, and Clinical Global Impressions-Severity (CGI-S) (Apiquian et al., 2009; Brissos et al., 2012; Garcia-Portilla et al., 2011; Hsieh et al., 2011; Jelastopulu et al., 2014; Juckel et al., 2008; Kawata and Revicki, 2008; Nafees et al., 2012; Nasrallah et al., 2008; Patrick et al., 2009; Schaub et al., 2011; Tianmei et al., 2011).
	Predictive validity	The PSP has been shown to predict relapse in consumers with schizophrenia (Nicholl et al., 2010).
	Internal consistency	The PSP has been found to have moderate to high internal consistency (Cronbach's alphas 0.64 to 0.87) (Apiquian et al., 2009; Brissos et al., 2012; Garcia-Portilla et al., 2011; Juckel et al., 2008; Kawata and Revicki, 2008; Nicholl et al., 2010; Schaub et al., 2011; Tianmei et al., 2011).
	Inter-rater reliability^b	The PSP's inter-rater reliability has varied across studies, with intra-class coefficients ranging from 0.43 to 1.0 (Brissos et al., 2012; Juckel et al., 2008; Morosini et al., 2000; Patrick et al., 2009; Schaub et al., 2011; Srisurapanont et al., 2008; Tianmei et al., 2011).
	Test-retest reliability^b	The PSP's test-retest reliability has varied, with intra-class coefficients between 0.61 and 0.98 (Garcia-Portilla et al., 2011; Juckel et al., 2008; Nasrallah et al., 2008; Patrick et al., 2009; Tianmei et al., 2011).
	Sensitivity to change	The PSP has been shown to detect change alongside established instruments like the PANSS and the CGI-S (Garcia-Portilla et al., 2011; Jelastopulu et al., 2014; Nafees et al., 2012; Nasrallah et al., 2008; Patrick et al., 2009; Tianmei et al., 2011). Increases in total scores of 7-9 may indicate clinically significant improvement (Nasrallah et al., 2008; Patrick et al., 2009), and decreases of 10 may indicate clinically significant decline (Nicholl et al., 2010).

- More detail on the psychometric properties of the HoNOS can be found elsewhere (Pirkis et al., 2005). The information presented in this table relates primarily to the HoNOS items that are concerned with functioning (Items 9-12).
- The level of reliability of an instrument is traditionally measured by a kappa value. Kappas of ≤ 0.20 are regarded as poor, 0.21-0.40 as fair, 0.41-0.60 as moderate, 0.61-0.80 as good, and ≥ 0.81 as very good.
- Some information on the LSP-16 is drawn from studies using the 20-item LSP-20 which includes all LSP-16 items.

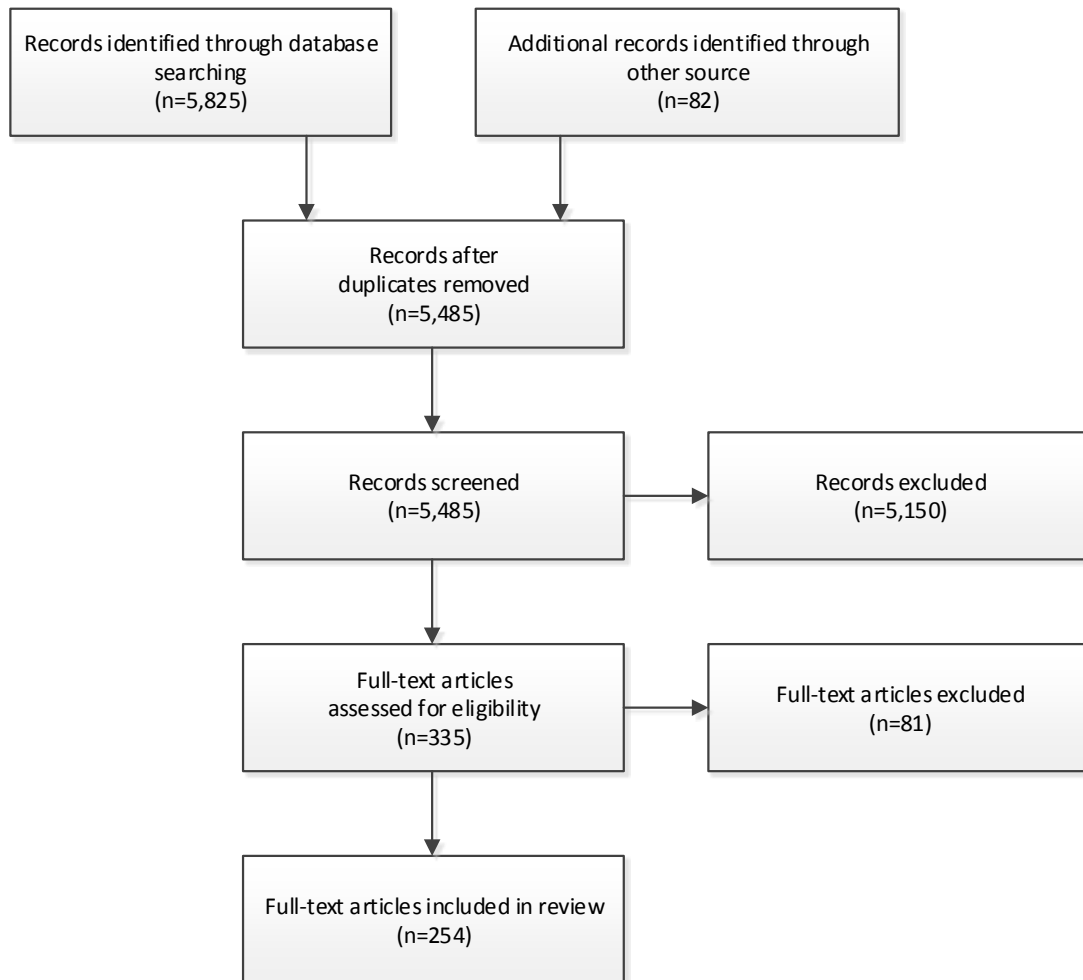


Figure 1: Article selection

All instruments	Disability Rating Form (DRF) FACE Core Assessment Global Assessment of Functioning (GAF) Health of the Nation Outcomes Scales (HoNOS) Illness Management and Recovery Scale - Clinician Version (IMRS-C) Level of Functioning Scale (LFS) Life Functioning Assessment Inventory (LFAI) Life Skills Profile 16 (LSP-16) Mini-ICF-APP Multi-Function Needs Assessment (MFNA)	Multnomah Community Ability Scale (MCAS) Need of Support and Service Questionnaire (NSSQ) Personal and Social Performance Scale (PSP) Profile of Community Psychiatry Clients (PCPC) Rehabilitation Evaluation Hall and Baker (REHAB) Residential Competency Scale (RCS) Social Adjustment Behavior Rating Scale (SABRS) Social and Occupational Functioning Assessment Scale (SOFAS) Social Functioning Index (SFI) Uniform Client Data Instrument (UCDI)
1. Is brief (<50 items) and simple to score	Disability Rating Form (DRF) Global Assessment of Functioning (GAF) Health of the Nation Outcomes Scales (HoNOS) Illness Management and Recovery Scale - Clinician Version (IMRS-C) Life Skills Profile 16 (LSP-16) Mini-ICF-APP	Multnomah Community Ability Scale (MCAS) Need of Support and Service Questionnaire (NSSQ) Personal and Social Performance Scale (PSP) Profile of Community Psychiatry Clients (PCPC) Rehabilitation Evaluation Hall and Baker (REHAB) Social and Occupational Functioning Assessment Scale (SOFAS) Uniform Client Data Instrument (UCDI)
2. Is not made redundant by more recent instruments	Disability Rating Form (DRF) Health of the Nation Outcomes Scales (HoNOS) Illness Management and Recovery Scale - Clinician Version (IMRS-C) Life Skills Profile 16 (LSP-16) Mini-ICF-APP	Multnomah Community Ability Scale (MCAS) Need of Support and Service Questionnaire (NSSQ) Personal and Social Performance Scale (PSP) Profile of Community Psychiatry Clients (PCPC) Rehabilitation Evaluation Hall and Baker (REHAB) Uniform Client Data Instrument (UCDI)
3. Relevant version has been scientifically scrutinised	Health of the Nation Outcomes Scales (HoNOS) Illness Management and Recovery Scale - Clinician Version (IMRS-C) Life Skills Profile 16 (LSP-16)	Multnomah Community Ability Scale (MCAS) Personal and Social Performance Scale (PSP) Rehabilitation Evaluation Hall and Baker (REHAB)
4. Considers functioning in a contemporary way	Health of the Nation Outcomes Scales (HoNOS) Illness Management and Recovery Scale - Clinician Version (IMRS-C) Life Skills Profile 16 (LSP-16)	Multnomah Community Ability Scale (MCAS) Personal and Social Performance Scale (PSP)
5. Demonstrates sound psychometric properties	Health of the Nation Outcomes Scales (HoNOS) Illness Management and Recovery Scale - Clinician Version (IMRS-C) Life Skills Profile 16 (LSP-16)	Multnomah Community Ability Scale (MCAS) Personal and Social Performance Scale (PSP)

Figure 2: Summary of instruments meeting criteria at each level of the hierarchy