

## International research

# Measuring mental health outcomes in primary care: the psychometric properties of a new patient-generated outcome measure, 'PSYCHLOPS' ('psychological outcome profiles')

Mark Ashworth DM MRCP MRCPsych

Honorary Senior Lecturer, Department of General Practice and Primary Care, King's College London, London, UK

Susan I Robinson PhD

Research Fellow, Lewisham Research Unit, London, UK

Emma Godfrey PhD

Lecturer, Department of Psychology, Institute of Psychiatry, King's College London, London, UK

Melanie Shepherd Dip Clin Psychol

Principal Clinical Psychologist, Ladywell Unit, University Hospital Lewisham, London, UK

Chris Evans MRCPsych

Consultant Psychiatrist in Psychotherapy and Trust Research Programmes Director, Nottinghamshire Healthcare NHS Trust, Rampton Hospital, Nottinghamshire, UK

Paul Seed MSc

Medical Statistician, Division of Reproductive Health, Endocrinology and Development, King's College London, London, UK

Henk Parmentier MD

STaRNet London Lead GP, Selsdon Park Medical Practice, Selsdon, UK

Andre Tylee MD FRCPsych

Editor, *Primary Care Mental Health*; Professor of Primary Care Mental Health, Institute of Psychiatry, King's College London, London, UK

### ABSTRACT

**Background** Patient-generated outcome measures are rarely used in evaluating talking therapies in primary care but in other contexts they show high sensitivity to change. We have devised a novel patient-generated measure called 'PSYCHLOPS' ('Psychological Outcome Profiles'). This paper describes the psychometric properties of PSYCHLOPS.

**Method** Standardised responses to PSYCHLOPS, pre- and post-therapy, were compared with responses to an established measure, CORE-OM.

**Results** Data were obtained from 110 patients. The effect size, a measure of sensitivity to change, was  $-1.53$  for PSYCHLOPS and  $-1.06$  for CORE-

OM ( $t = 5.10$ ,  $P < 0.001$ ). Pre-therapy alpha scores were  $0.79$  and  $0.94$ , respectively. Change scores of both instruments correlated strongly (Spearman's  $\rho = 0.61$ ;  $P < 0.001$ ). Detailed validity testing is reported.

**Conclusion** PSYCHLOPS is a sensitive measure of change after therapy. Alpha scores suggest satisfactory internal reliability. Evidence of convergent, concurrent and construct validity has been obtained. Further work is required to establish test-retest reliability.

**Keywords:** patient-generated outcome measures, psychometrics, validation study

## Background

Patient-generated outcome measures quantify aspects of health-related quality of life of most importance to the patient. Their development has been greatest in the field of rehabilitation.<sup>1</sup> In the psychological literature, they have been described as 'idiographic measures'. Examples include the Personal Questionnaire (used to determine treatment goals) and Goal Attainment Scaling (used to elicit definitions of recovery).<sup>2,3</sup> Use of these instruments has been hampered by their length, complexity or requirement for interviewer guidance to enable completion.

The increasing popularity of patient-based measures in other disciplines has not been reflected in psychiatric research.<sup>4</sup> Commonly used mental health outcome measures in primary care generally consist of a mix of symptom scores and quality of life dimensions. Examples include CORE-OM (Clinical Outcomes Routine Evaluation – Outcome Measure), HoNOS (Health of the Nation Outcome Score), and HRU (How Are You?).<sup>5–7</sup> These self-completed quality of life assessments allow patients to participate in measuring outcomes, but are these the outcomes that matter to patients? The majority of instruments are expert derived and consist of checklists of symptoms. Many items on a checklist questionnaire might have little personal relevance to an individual patient. We wanted to develop an outcome measure that would measure just those items considered to have personal meaning and importance to patients.

### Description of PSYCHLOPS

We have devised a patient-generated outcome measure termed 'PSYCHLOPS' (Psychological Outcome Profiles).<sup>8</sup> It was adapted from MYMOP (Measure Yourself Medical Outcome Profile), an instrument used to assess physical illness.<sup>9</sup> Like its precursor, PSYCHLOPS is short (one page) and can readily be self-administered. PSYCHLOPS consists of four questions. The first question asks: 'Choose the problem that troubles you most. Please write it in the box below'. Having described the problem, the respondent is invited to score it on a six-point score ranging from 'not at all affected' to 'severely affected'. The duration of the problem is also elicited. Further questions ask for a description and scoring of any other problems and a description and scoring of any consequent functional impairment. Finally, the respondent is asked to score their wellbeing using a six-point score.

At the end of therapy, the patient is given a similar questionnaire. The therapist is required to transcribe

the original responses to the freetext questions (two problem domain questions and one function domain question), copying them into identical boxes on the post-therapy questionnaire. The patient is then asked to score these same items again, using identical scales. The wellbeing score is repeated. A fifth, validation question, concludes the questionnaire and asks how the patient feels in themselves, now that therapy has finished; responses range from 'much better' to 'much worse' on a five-point scale.

Total scores are obtained by adding the individual six-point scores for the four questions. The sum is divided by four to give the mean score per question (or by three if only one of the two problem questions has been answered). Outcome scores are the difference between the total pre- and post-therapy questionnaire scores. The validation question is not included in the total scoring.

We aimed to validate PSYCHLOPS by comparing it to an established measure, CORE-OM (Clinical Outcomes Routine Evaluation – Outcome Measure).<sup>5</sup> CORE-OM has been well validated in primary care populations and has become a widely used self report psychological outcome measure.

## Methods

### Setting and therapists

We recruited four primary care therapists, all routinely collecting CORE-OM data, to participate in our trial. All were based in south London or the south-east of England. Between them, they represented a range of professional groups offering talking therapy in primary care: counsellors, clinical psychologists, counselling psychologists and psychotherapists.

### Study design

We conducted a longitudinal survey, in which patients entering therapy in primary care were asked to fill in the CORE-OM and PSYCHLOPS measures at initial assessment and at the final therapy session. Ethical approval for the study was obtained from the London Multi-centre Research Ethics Committee (MREC).

### Sample size

A sample size of at least 100 with both CORE-OM and PSYCHLOPS completed pre- and post-therapy

was chosen to provide tight confidence intervals on correlations between the measures, on change and for internal consistency (the 95% confidence interval (CI) on an observed alpha of 0.8 for  $n = 100$  and for three items is from 0.72 to 0.86).<sup>10</sup> This is well in excess of the numbers necessary when not looking for high factorial complexity.<sup>11</sup> Recruitment was terminated when it was clear that this number would be reached. Inevitably, some patients subsequently attended a final session and completed their post-therapy measures. The final dataset contained 110 sets of data from both measures on both occasions.

## Statistical methods

Questionnaire responses were compared in the longitudinal dataset, consisting of all clients completing both sets of questionnaires. The two instruments had three domains in common (problems, function and wellbeing). CORE-OM data were analysed as described in the CORE System Handbook.<sup>12</sup> With occasional exceptions, all CORE-OM analyses were based on the 28 questions covering the three domains in common with PSYCHLOPS (a few analyses consider the scores across all 34 items including risk, the fourth domain, and are marked 'CORE-OM – including risk items'). The two instruments have differing numerical scales so comparison was based on standardised or  $z$  scores produced by subtracting the pre-therapy mean from the score from each observation and dividing by the standard deviation (SD) of pre-therapy values. Standardisation results in all cases showing a pre-therapy mean  $z$  score of zero and SD of 1.0. Since standardisation is to the pre-therapy scores, change and post-therapy scores generally show negative means, and the SD of change can be above or below one.

Sensitivity to change was explored by calculating Cohen's 'effect size', which is the change score divided by the SD of the baseline measure.<sup>13</sup> The effect size is the most commonly reported standardised method for expressing responsiveness.<sup>14</sup> An effect size of  $-1.0$  indicates that the mean score after therapy has reduced by one SD. An effect size of magnitude 0.8 or greater is considered large.<sup>15</sup>

Reliability was tested by calculating Cronbach's alpha score for each instrument. A score above 0.70 is generally considered as demonstrating satisfactory internal consistency, whereas a score in excess of 0.90 suggests some redundancy of individual items within the questionnaire.<sup>16</sup>

Validity was tested in a number of ways. Convergent validity was assessed by exploring the correlation between comparable domains in the two measures. Spearman's rho correlation coefficients were obtained since distributions were not Gaussian.

Concurrent validity was also assessed by comparing change scores in those reporting differing degrees of recovery. One aspect of construct validity was assessed by relating change to problem duration, on the basis that less change is generally expected if problems are long-standing.

CORE-OM may be used to define clinically significant change and reliable change.<sup>17</sup> In order to further define the characteristics of PSYCHLOPS, we compared change scores in our study sample using this methodology.

Results were analysed using SPSS for Windows 11.0 and STATA version 8.2.

## Results

### Sample size and non-completers

A total of 235 patients completed the pre-therapy questionnaires and 110 (47%) subsequently completed post-therapy questionnaires. The mean age of those completing both sets of questionnaires was 32 years (range 15–64 years), 71% were female and the median number of therapy sessions was 5.5 (range 1–28).

The 125 patients who failed to complete therapy were similar in age and sex to those completing both sets of questionnaires but had significantly lower mean pre-therapy PSYCHLOPS and CORE-OM scores ( $t = -2.63$ ,  $P = 0.01$ ;  $t = -2.71$ ,  $P = 0.01$ , respectively) suggesting that they had less psychological distress. All subsequent results refer only to the 110 patients for whom pre- and post-therapy questionnaire data were available.

### Acceptability

Completion rates for the domains in PSYCHLOPS pre- and post-therapy, respectively, were: problems: 100%, 100%; functioning: 100%, 100%; wellbeing: 97%, 94%. CORE-OM individual item completion rates were 99.8% pre-therapy and 99.6% post-therapy. Just three patients (3%) did not describe and score a second problem.

### Internal reliability

Pre- and post-therapy values of Cronbach's alpha for CORE-OM were 0.94 and 0.95, respectively. For PSYCHLOPS the values were 0.79 and 0.87 respectively. These values indicate strong internal reliability for both measures. The correlation coefficients

between the two PSYCHLOPS problem domain scores were 0.62,  $P < 0.001$  pre-therapy and 0.66,  $P < 0.001$  post-therapy (Spearman's rho), indicating the internal consistency of PSYCHLOPS.

### CORE-OM scores

Prior to therapy, the mean CORE-OM score was 1.81 and on completion of therapy, the mean fell to 1.05 (see Table 1). Applying the Jacobson clinically significant change cut-offs for caseness, 65% were cases pre-therapy and 26% were cases post-therapy.<sup>18</sup> Using the observed initial SD (0.72) and alpha coefficient (0.94) for these data gave a reliable change criterion of 0.50, 74 (67%) patients showed reliable change and 2 (2%) showed reliable deterioration; 44 (40%) showed both clinically significant and reliable change.

### PSYCHLOPS scores

Improvements in PSYCHLOPS scores were also observed. Scores fell from a pre-therapy mean of 3.22 to a post-therapy mean of 1.73 (see Table 1).

### Comparison of PSYCHLOPS and CORE-OM scores

Initial CORE-OM and PSYCHLOPS scores correlated strongly ( $\rho = 0.65$ ) as did end of therapy scores ( $\rho = 0.74$ ) and change scores ( $\rho = 0.57$ ). The relationships between the standardised scores are shown in Figures 1–3. Patients reaching the score

level of caseness on CORE-OM had a mean initial PSYCHLOPS score of 3.62; non-cases of 2.50 ( $t = -6.67$ ;  $P < 0.0001$ ). Reductions in PSYCHLOPS scores were greater in those whose CORE-OM scores had demonstrated clinically significant and reliable change:  $-2.1$  compared to  $-1.1$ , respectively ( $t = 4.85$ ;  $P < 0.001$ ).

### Sensitivity to change

PSYCHLOPS change scores demonstrated a larger effect size than CORE-OM (see Table 2):  $-1.53$  (95% CI  $-1.30$  to  $-1.76$ ) compared to  $-1.06$  (95% CI  $-0.90$  to  $-1.23$ ), respectively,  $t = 5.10$ ,  $P < 0.001$ .

### Convergent validity – domain correlates

Three domain scores were common to both instruments, though for neither instrument were well-being, functioning or problems expected to be distinct factors of interpersonal difference. Change scores were highly correlated (see Table 3), although correlations between parallel domains were not consistently greater than between non-parallel domains.

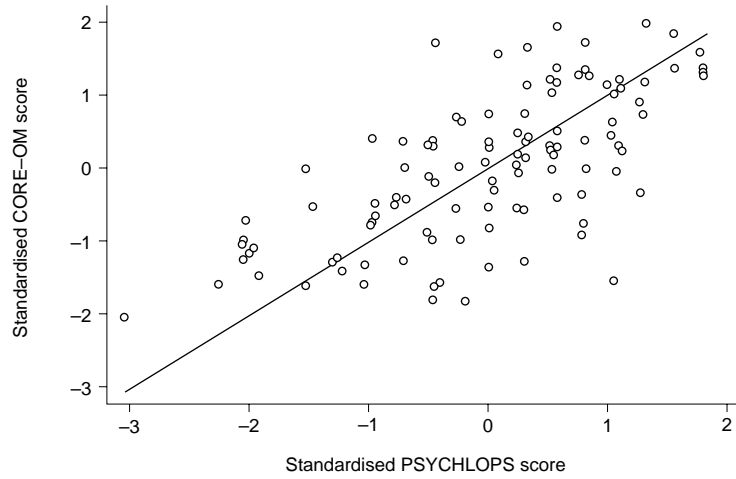
### Concurrent validity – change scores and self-report improvement scores

Change scores were compared, using effect sizes, for those reporting themselves 'a little better' ( $n = 32$ ) or 'much better' ( $n = 41$ ). Other responses to the five-point scale were ignored because no patients reported

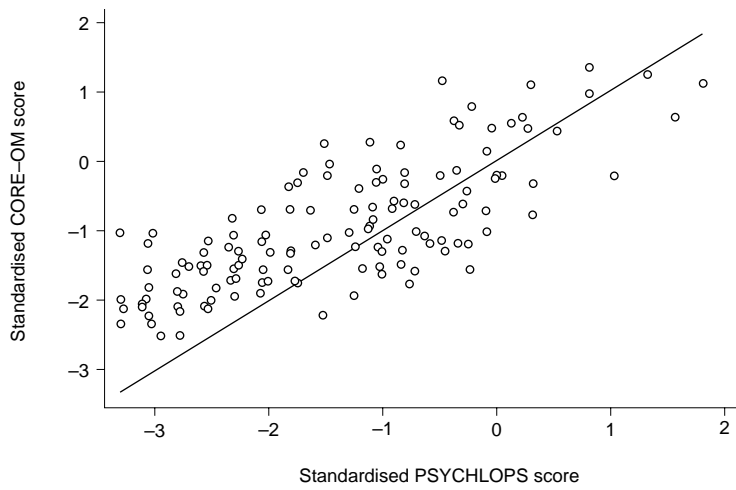
**Table 1** PSYCHLOPS and CORE-OM parameters: unstandardised means and standard deviations<sup>a</sup> ( $n = 110$ )

Domain	PSYCHLOPS			CORE-OM		
	Pre-therapy	Post-therapy	Change	Pre-therapy	Post-therapy	Change
Problems	6.69 (2.14)	3.69 (2.65)	-3.00 (2.78)	1.98 (0.77)	1.09 (0.70)	-0.89 (0.75)
Functioning	3.43 (1.40)	1.67 (1.50)	-1.75 (1.53)	1.55 (0.74)	0.98 (0.68)	-0.57 (0.60)
Wellbeing	2.77 (1.13)	1.59 (1.22)	-1.16 (1.40)	2.09 (0.97)	1.13 (0.83)	-0.96 (0.89)
Risk	n/a	n/a	n/a	0.29 (0.39)	0.13 (0.28)	-0.16 (0.36)
Sum of all non-risk items	3.22 (0.97)	1.73 (1.19)	-1.49 (1.19)	1.81 (0.72)	1.05 (0.66)	-0.77 (0.63)

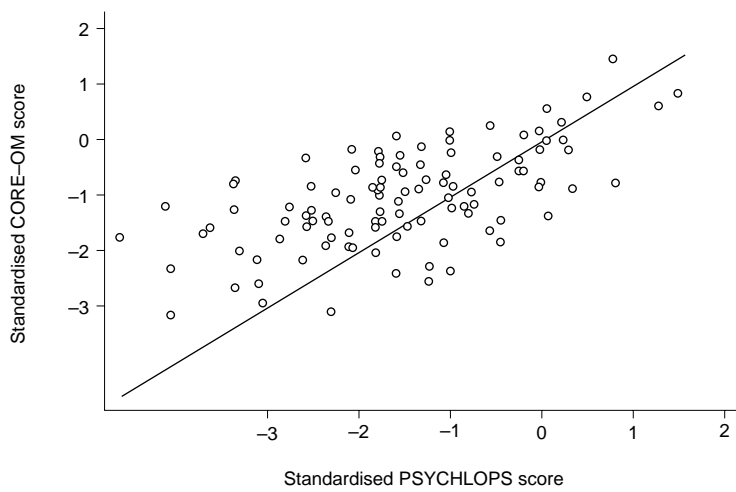
<sup>a</sup>Figures are means and, in brackets, standard deviations  
n/a: not available.



**Figure 1** Comparison of standardised PSYCHLOPS and CORE-OM scores, pre-therapy



**Figure 2** Comparison of standardised PSYCHLOPS and CORE-OM scores, post-therapy



**Figure 3** Comparison of change in standardised PSYCHLOPS and CORE-OM scores

**Table 2** Comparing change after therapy: the effect size<sup>a</sup>

Domain	PSYCHLOPS: mean effect size, (SD)	CORE-OM: mean effect size, (SD)	Difference between PSYCHLOPS and CORE-OM	
			Mean (95% CI)	Significance
Problems	-1.40 (1.30)	-1.16 (0.98)	-0.24 (-0.46, -0.02)	$t = -2.15; P = 0.03$
Functioning	-1.25 (1.09)	-0.78 (0.82)	-0.47 (-0.67, -0.27)	$t = -4.57; P < 0.001$
Wellbeing	-1.02 (1.23)	-0.96 (0.93)	-0.06 (-0.27, 0.15)	$t = -0.55; P = 0.58$
Risk	n/a	-0.41 (0.73)	n/a	n/a
Sum of all non-risk items	-1.53 (1.22)	-1.06 (0.87)	-0.47 (-0.65, -0.29)	$t = -5.10; P < 0.001$

<sup>a</sup> The effect size (change scores/baseline SD) equals the standardised post-therapy score, since standardisation resulted in a pre-therapy mean score of zero and SD of 1.0  
n/a, not available

**Table 3** Comparing PSYCHLOPS and CORE-OM: correlates of standardised change scores ( $n=104$ )

PSYCHLOPS	CORE-OM			
	Problems	Functioning	Wellbeing	Sum of all non-risk items
Problems	<b>0.48</b>	0.43	0.39	0.50
Functioning	0.46	<b>0.40</b>	0.40	0.48
Well-being	0.51	0.46	<b>0.52</b>	0.56
Sum of all items	0.58	0.51	0.51	<b>0.61</b>

Figures are Spearman's rho correlation coefficients; significance of all values,  $P < 0.001$   
Figures in bold represent correlations between the same domains of PSYCHLOPS and CORE-OM

that they were 'a little worse' or 'much worse', and just five reported that they were 'about the same'. The effect sizes for responses to PSYCHLOPS were -1.20 for those reporting 'a little better' and -2.16 for those reporting 'much better' ( $t = -3.78; P < 0.001$ ). The effect sizes for responses to CORE-OM were -1.37 and -0.83, respectively ( $t = -2.81; P = 0.01$ ).

from one month to over five years. The longer the problems had lasted, the smaller the PSYCHLOPS change score (Spearman's rho = 0.24;  $P = 0.01$ ) and the CORE-OM change score (Spearman's rho = 0.19;  $P = 0.05$ ).

### Construct validity – change scores and symptom duration

Questions in the problem domain of PSYCHLOPS elicited information about the duration of the main problems. Results were scored on a five-point scale

## Discussion

### Main findings

PSYCHLOPS was a sensitive measure of change after therapy. Measuring the responsiveness of both instruments by using Cohen's effect size parameter

showed PSYCHLOPS to be significantly more responsive than the CORE-OM. High sensitivity to change has been found in other studies of patient-generated measures and appears to be a characteristic of such instruments, which aim to address the concerns of the patient rather than eliciting responses to predetermined questions that are probably not experienced as similarly personally relevant.<sup>9,19</sup>

The chief drawback of patient-generated measures is feasibility, as they are normally complex, requiring administration by trained interviewers.<sup>14</sup> We have produced an individualised outcome measure that is brief and self-completed. Completion rates were high in our survey, suggesting that feasibility was similar to that of CORE-OM.

Parameters for the CORE-OM were not markedly dissimilar from other similar samples.<sup>20</sup> Both measures showed good reliability; the value found in our study for the CORE-OM was almost identical to that found in the original development samples.<sup>18</sup> The relatively high internal reliability for the four questions in PSYCHLOPS, both pre- and post-therapy raises no concerns about scoring such a small number of items as a combined indicator.

## Validity of PSYCHLOPS

We have already reported on high face validity in a group of therapists developing and using PSYCHLOPS.<sup>21</sup>

Convergent validity with CORE-OM has been demonstrated by the strong overall correlations between change scores on the two instruments and by strong correlations between similar domains. However, these correlations showed no discriminant validation between separate domains since a number of correlations were greater for non-corresponding domains than for corresponding domains. This is congruent with the findings of Evans *et al*, and the more detailed psychometric work of Lyne *et al*, which both show that domain separation within the CORE-OM is weak, a finding in common with that in many self-report measures.<sup>18,22</sup>

Unsurprisingly, patients identified as cases had significantly higher PSYCHLOPS scores. Patients demonstrating a reliable change from caseness to non-caseness ('clinically significant and reliable change') had significantly greater PSYCHLOPS change scores.

Concurrent validity is supported by significantly greater change scores in those self-reporting greater recovery after therapy. Although not a strong test of construct validity, the finding of significantly greater change in PSYCHLOPS scores in those whose problems were of more recent onset fits with the

expectation that more chronic problems would show less change over therapy. The same relationship was just statistically significant for CORE-OM change too, though less so than for PSYCHLOPS, probably because the duration questions related directly to the PSYCHLOPS problems rather than to the problem domain questions of CORE-OM.

## Differences between PSYCHLOPS and CORE-OM

Although overall change scores were significantly correlated, the scatter plots demonstrate a degree of disagreement between the instruments, more apparent post-therapy than pre-therapy. The effects summate, as shown in Figure 3, such that the greatest divergence between the measures occurs at the bottom left corner. Dots in this section of the plot represent patients whose scores have improved the most. For these patients, PSYCHLOPS scores tended to show rather more improvement than CORE-OM scores. Put differently, the more a patient had improved after therapy, the more this improvement could be demonstrated by changes in PSYCHLOPS scores rather than CORE-OM scores.

Further differences emerged when analysing PSYCHLOPS scores according to the criteria of caseness, as defined by CORE-OM scores. Some highly-scoring PSYCHLOPS cases failed to reach the level of caseness on CORE-OM. Some cases demonstrated substantial improvement on PSYCHLOPS, but the improvement did not cross the threshold of clinically significant and reliable change on CORE-OM. In further work, we intend to analyse the similarities and differences based on a qualitative analysis of the freetext responses made by patients using PSYCHLOPS. We hope that such analysis will enable us to determine whether some high-initial- or high-improvement-scoring PSYCHLOPS cases were reporting areas of psychological distress not elicited by CORE-OM, thus accounting for their apparent invisibility on this measure.

## Limitations of the present study and of PSYCHLOPS

Conceptual differences have arisen between patient-generated measures and nomothetic measures. Although tuned to what matters to individual patients, idiographic measures are not as obviously suited to comparative analyses between patients as nomothetic measures. At one extreme, it has been argued that patient-generated measures have no external frame of reference and can only be used to compare

change within an individual rather than comparing an individual with known clinical or non-clinical populations.<sup>23</sup> However, the strong correlations between the measures does suggest that there is some covariance across patients and hence, presumably, some empirical legitimacy to making comparisons between different patients' PSYCHLOPS scores, even though each patient is likely to be rating different aspects of psychological distress.

In reality, patient-generated measures are more likely to represent a trade-off between two fundamental applications of health status instruments: evaluation and discrimination. The evaluative instrument measures change over time, whereas the discriminative instrument distinguishes between two groups of patients. Maximising the evaluative attributes is likely to be achieved at the cost of diminishing the discriminative attributes. Similarly, evaluative instruments are more likely to detect real clinical change in patients but they tend to have lower test–retest reliability and lower internal reliability.

The lower reliability of patient-generated questionnaires does not diminish their ability to measure within-patient change over time. But instruments such as PSYCHLOPS are unlikely to be able to discriminate well between individuals since there is no theoretical reason why baseline scores should be comparable. When patients are asked to select their main problems, those with less-severe psychological distress will select different problems from those who are more severely distressed. The resulting scores cannot be compared since like is not being compared with like. Thus an analysis of PSYCHLOPS based on artificially derived cut-off scores denoting a pathological level of distress will be misleading. Instead, initial scores should be interpreted in the light of subsequent within-patient change scores.

## Further work

In addition to the need for determining test–retest reliability, further work needs to explore construct validity by independent questioning of the therapists to determine convergence between therapists' impression of recovery and that of patients themselves.

Fitzpatrick *et al* have suggested eight criteria for selecting an ideal patient-based outcomes measure.<sup>14</sup> These are: appropriateness, reliability, validity, responsiveness, precision, interpretability, acceptability and feasibility. Further work is needed to explore interpretability since PSYCHLOPS change scores in a clinical population cannot be compared to change scores based on normative population

data, although we are currently conducting a population survey to explore discriminative validity. Such change scores would act as a control group providing data on the usual patterns of, for example, high initial scores over time. The interpretation of change requires study across varying populations.

Finally, CORE-OM and PSYCHLOPS are asking fundamentally different questions within similar domains. For each question, CORE-OM asks 'how often' and PSYCHLOPS asks 'how severe'. Without further qualitative study, we do not know if this distinction will have influenced comparability.

## Implications of findings

There is increasing pressure on all therapists to evaluate their work. This pressure is driven by commissioners and by the Department of Health. In spite of this, one recent survey found that only 8% of commissioners received routine outcome measures.<sup>12</sup> Several complex factors probably hinder attempts to measure the outcomes of therapy, as evidenced by the recent finding that just 15% of counsellors use validated outcome measures.<sup>24</sup> We have devised and tested an instrument that is likely to be acceptable to therapists, and highly sensitive to change after therapy.<sup>21</sup> Both these features may encourage more widespread adoption of PSYCHLOPS as an outcome measure.

Others have noted that idiographic measures frequently elicit information not covered by standard patient-based measures,<sup>19</sup> and we consider that the real potential of PSYCHLOPS is in the richness of the information offered by patients as they describe the problem that affects them most and how this has impaired their functioning. We are now embarking on a qualitative thematic evaluation of these self-generated reports of distress elicited from patients completing PSYCHLOPS during the course of this study. We hope shortly to publish findings detailing the nature of problems described by patients using PSYCHLOPS.

The importance of service user involvement in the management of healthcare has grown over the last decade, and in particular, the concept of incorporating the service user perspective into more global evaluations of mental healthcare.<sup>25</sup> The momentum of change is likely to shift the focus of evaluation toward a more user-centred approach so that service users themselves set the criteria by which the effectiveness and quality of mental healthcare provision is judged. PSYCHLOPS is an easily administered, patient-generated measure that, because of its high sensitivity to change, is suitable for evaluating mental health outcomes.



## ACKNOWLEDGEMENTS

We would like to thank all the therapists who recruited patients for this study and the patients who completed the psychometric measures. We would also like to thank Charlotte Paterson who generously allowed us to adapt MYMOP which formed the starting point for our attempts to create a new mental health outcome measure.

This work was funded through a grant made by the King's Fund. SR's post has been funded exclusively from this source. MA has been part funded by the South Thames Research Network (STaRNet), London.

## REFERENCES

- 1 Donnelly C and Carswell A. Individualized outcome measures: a review of the literature. *Canadian Journal of Occupational Therapy* 2002;69:84–94.
- 2 Shapiro M. A method of measuring psychological changes specific to the individual psychiatric patient. *British Journal of Medical Psychology* 1961; 34:151–5.
- 3 Kiresuk T and Sherman R. Goal attainment scaling: A general method for evaluating community mental health programs. *Community Mental Health Journal* 1968;4:443–53.
- 4 Gilbody S, House A and Sheldon T. *Outcome Measurement in Psychiatry*. Report no. 23. University of York: NHS Centre for Reviews and Dissemination, 2003.
- 5 Evans C, Mellor-Clark J, Margison F *et al.* CORE: Clinical Outcomes in Routine Evaluation. *Journal of Mental Health* 2000;9:247–55.
- 6 Wing JK, Beevor AS, Curtis RH *et al.* Health of the Nation Outcome Scales (HoNOS). Research and development. *British Journal of Psychiatry* 1998; 172:11–18.
- 7 Katsavdakis KA, Clifford PI, Evans RB *et al.* The How Are You? scale: a quality-of-life outcomes measure for routine practice. *Bulletin of the Menninger Clinic* 1999;63:366–87.
- 8 Ashworth M, Shepherd M, Christey J *et al.* A client-centred psychometric instrument: the development of 'PSYCHLOPS' ('Psychological Outcome Profiles'). *Counselling and Psychotherapy Research* 2004;4:27–33.
- 9 Paterson C. Measuring outcome in primary care: a patient-generated measure, MYMOP, compared with the SF-36 health survey. *British Medical Journal* 1996;312:1016–20.
- 10 Feldt L, Woodruff D and Salih F. Statistical inference for coefficient alpha. *Applied Psychological Measurement* 1987;11:93–103.
- 11 Guadagnoli E and Velicer W. Relation of sample size to the stability of component patterns. *Psychological Bulletin* 1988;103:265–75.
- 12 CORE System Group. *CORE System (Information Management) Handbook*. Leeds: CORE System Group, 1998.
- 13 Cohen J. *Statistical Power Analysis for the Behavioural Sciences*. New York: Academic Press, 1977.
- 14 Fitzpatrick R, Davey C and Buxton M *et al.* Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment* 1998;2:1–74.
- 15 Kazis L, Anderson J and Meenan R. Effect sizes for interpreting changes in health status. *Medical Care* 1989;27:S178–89.
- 16 Nunnally J and Bernstein I. *Psychometric Theory* (3e). New York: McGraw-Hill, 1994.
- 17 Evans C, Margison F and Barkham M. The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence Based Mental Health* 1998;1:70–2.
- 18 Evans C, Connell J, Barkham M *et al.* Towards a standardised brief outcome measure: psychometric properties and utility of the CORE-OM. *British Journal of Psychiatry* 2002;180:51–60.
- 19 Hunter R, McLean J, Peck D *et al.* 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* 2004;13:93–105.
- 20 Evans C, Connell J, Barkham M *et al.* Practice-based evidence: benchmarking NHS primary care counselling services at national and local levels. *Clinical Psychology and Psychotherapy* 2003;10:374–88.
- 21 Ashworth M, Robinson S, Godfrey E *et al.* The experiences of therapists using a new client-centred psychometric instrument, 'PSYCHLOPS' ('Psychological Outcome Profiles'). *Counselling and Psychotherapy Research* 2005;5:37–41.
- 22 Lyne KD, Barrett P, Evans C *et al.* Dimensions of variation on the CORE-OM. *British Journal of Clinical Psychology* 2006;in press.
- 23 Lacasse Y, Wong E and Guyatt, G. Individualising questionnaires. In: Joyce CRB, O'Boyle CA and McGee H (eds). *Individual Quality of Life. Approaches to conceptualisation and assessment*. Amsterdam: Harwood Academic Publishers, 1999, pp. 87–103.
- 24 Mellor-Clark J and Barkham M (2000) Quality evaluation: methods, measures and meaning. In: Feltham, C and Horton I (eds). *The Handbook of Counselling and Psychotherapy*. London: Sage Publications.
- 25 Rea D. Towards routine user assessment of mental health quality performance. *International Journal of Health Care Quality Assurance* 1999;12:169–76.

## CONFLICTS OF INTEREST

Although the final named author is the editor of *Primary Care Mental Health*, this paper was reviewed by two independent reviewers and amended in the light of their comments.

## ADDRESS FOR CORRESPONDENCE

Mark Ashworth, GKT Department of General Practice and Primary Care, King's College London, 5 Lambeth Walk, London SE11 6SP, UK. email: [mark.ashworth@gp-G85053.nhs.uk](mailto:mark.ashworth@gp-G85053.nhs.uk)

*Received ??????*

*Accepted ??????*