ORIGINAL ARTICLE

Personalising Outcome Measurement in Substance Misuse Treatment: the Feasibility and Psychometrics of Two Individualised Outcome Measures

Paula Alves[1] · Luís Faísca[2] · Célia M. D. Sales[3] · Mark Ashworth[4] 🔵

Abstract

Individualised information in substance misuse treatment complements standardised outcome measures. However, few studies investigate the use of individualised measures and their robustness in terms of quantifying outcomes. In this study, we analysed the psychometrics and feasibility of two individualised outcome measures (PQ and PSYCHLOPS). We followed a cross-sectional methodology, administering the individualised measures and three additional standardised measures (TOP, a measure of psychological health within addiction services; PHQ-9; CORE-OM) to a sample of 93 patients entering substance misuse treatment in four clinical services. The results showed high levels of patient acceptability of the two individualised measures (response rates > 95%). The internal reliability was good for both PQ and PSYCHLOPS (Cronbach's alpha, .79 and .72, respectively). Convergent validity of PQ with standardised measures was weak: Pearson's r values for TOP (psychological health), PHQ-9 and CORE-OM were .21, .22 and .27, respectively. In contrast, convergent validity of PSYCHLOPS was moderate: $r = .40$, .39 and .50, respectively. Convergence between PQ and PSYCHLOPS was weak ($r = .28$). Experience of previous treatment episodes was associated with higher PQ and PSYCHLOPS scores; PSYCHLOPS but not PQ scores were higher among those opting to complete the questionnaires in written rather than verbal format. Our findings demonstrated that PQ and PSYCHLOPS are reliable and feasible individualised outcome measures for use in substance misuse treatment units, although the lack of strong convergent validity indicates that they may be measuring different underlying constructs. Optimal outcome measurement may involve combining individualised and standardised measures.

Keywords Substance misuse treatment · Individualised outcome measures · Personalised assessment · PQ · PSYCHLOPS

✉ Mark Ashworth
mark.ashworth@kcl.ac.uk

Extended author information available on the last page of the article

The evaluation of treatment outcomes has, until recently, been dominated by a nomothetic approach. However, a growing body of expert opinion has recommended that idiographic, or individualised, information should also be used for outcome measurement purposes (Ashworth et al. 2019; Alves et al. 2015; Neale and Strang 2015). Combining nomothetic and idiographic approaches has been described as an 'optimal way of understanding and measuring patient reported outcomes of addiction and its related constructs' (Neale and Strang 2015, p. 1216). However, little is known about the psychometric properties of individualised outcome measures when tested within the context of substance misuse treatment services and the validity of self-reported outcomes in this setting.

Broadly speaking, outcome measures aim to quantify changes that occur in a patient's clinical status over a course of treatment. The commonest strategy involves the use of standardised outcome measures, based on pre-set questions administered to a group of patients. The standardised structure of these measures allows for mechanisation of the outcome measurement process, since all individuals are presented with the same pre-determined questions (OECD 2013). The generation of large-scale data from standardised instruments has given these measures a golden status in the field of outcome evaluation (Evans et al. 2000).

In recent years, individualised outcome measures have gained ground in health care (Fitzpatrick et al. 1998) for a number of reasons. First, they provide patients with an opportunity to indicate items to evaluate the outcome of their treatment (Barkham 2016); second, they allow patients to rate the intensity or level of distress caused by their own problem, corresponding to their priorities (Elliott et al. 2016; Sales and Alves 2012, 2016); third, they have a greater sensitivity to clinical change than standardised outcome measures (Ashworth et al. 2009; Elliott et al. 2016); fourth, they are likely to capture information overlooked by standardised measures (Alves et al. 2018); and fifth, patients value the opportunity to define and express their own range of concerns (Alves et al. 2016); sixth, there is recognition of the clinical utility of PQ and PSYCHLOPS, since both contain items of high relevance for patients (Ashworth et al. 2005; Sales et al. 2007); and seventh, therapist satisfaction has also been reported for both measures for a variety of clinical tasks including outcome assessment (Sales et al. 2007; Sales and Alves 2012).

For Lambert and Hawkins (2004), to be appropriate, outcome measures must be 'psychometrically strong, applicable and practical' (p. 493). A growing body of literature has explored these properties among individualised outcome measures. These studies have focused mostly on the Simplified Personal Questionnaire (PQ; Elliott et al. 2016) and the Psychological Outcome Profiles (PSYCHLOPS; Ashworth et al. 2004). PQ and PSYCHLOPS are individualised measures targeting general psychological distress, based on items generated by patients, and widely used in mental health care (Sales and Alves 2016).

Regarding psychometric properties, both PQ and PSYCHLOPS have demonstrated reliability and validity within a range of mental health settings (e.g. university counselling units, mental health primary care) (Ashworth et al. 2009; Elliott et al. 2016). However, further studies to extend psychometric testing are required to diminish scepticism about the use of individualised measures (Elliott et al. 2016).

Substance misuse treatment is one of the fields where individualised measures have been little explored. In this client group, outcome measurement tends to focus on specific drug-related issues, as well as psychosocial variables, including general psychological distress and specific problems such as depression or suicidal thoughts (Darke et al. 2007; Gossop et al. 2003; Jones et al. 2009; Wanigaratne et al. 2005). Outcomes tend to be assessed with standardised instruments and it is unknown whether individualised outcome measures in this

context are feasible and psychometrically robust for outcome measurement purposes, as observed in other fields (Alves et al. 2015).

The first recorded use of PQ and PSYCHLOPS in a substance misuse treatment setting was in Portugal in 2013 and has formed the basis of a longer term study which includes the present study (Alves et al. 2013). Preliminary findings show that patients value these measures for being easy to understand; helping them to reflect about personal difficulties; providing them with freedom to talk about problems, related or not to substance misuse; and making them 'feel like normal people' (Alves et al. 2016). As a downside, patients reported that the level of self-exposure required by these instruments was sometimes uncomfortable and might contribute to non-completion (Alves et al. 2016). This project also revealed that PQ and PSYCHLOPS elicit problems overlooked by standardised measures (Alves et al. 2018). For instance, more than two-thirds of themes indicated by patients in these individualised measures were absent from the Treatment Outcomes Profile (TOP; Marsden et al. 2008), including problems such as 'money worries' and 'self-image/self-worth' (Alves et al. 2018). The present study builds on these findings by exploring the feasibility of PQ and PSYCHLOPS, as well as their psychometric properties in this sample.

Study Rationale

The current study focussed mainly on the feasibility and psychometric properties of PQ and PSYCHLOPS and standardised comparators. Additionally, we explored potential factors that may influence PQ and PSYCHLOPS scores (e.g. socio-demographic, substance misuse, treatment history and other variables of interest), to gain further knowledge about the psychometric value of these instruments. We also aimed to contrast the scorings obtained with PQ and PSYCHLOPS, since both are individualised measures albeit with different characteristics (e.g. mode of administration). To our knowledge, this is the first study to explore these three questions within the context of substance misuse services.

Methods

This was a cross-sectional study with a sample of adult patients entering treatment in four services for drug and alcohol substance misuse in Portugal, between July 2013 and May 2015. Approval was granted by the Committee for Health of Lisbon and surrounding areas (ARSLVT, Ref. 8251/CES/2012).

Participants

The sample comprised of patients aged 18 years and older, who were starting a treatment episode (i.e. first episode or treatment of a relapse) and were fluent in Portuguese. We excluded patients who were not motivated to participate or were too distressed to complete the evaluation protocol. This decision was made by researchers in consultation with the administrative/clinical staff available. We calculated our proposed sample size based on previously reported alpha scores of .80 for PSYCHLOPS, with 114 patients, and .86 for PQ, with 72 patients (Ashworth et al. 2009; Elliott et al. 2016). Based on these findings, we sought to recruit a sample of 100 patients. Of the 102 individuals who met the inclusion

criteria, 94 (92%) consented to participate. One was excluded because of non-completion of PQ and PSYCHLOPS, resulting in a final sample of 93 patients.

Instruments

Simplified Personal Questionnaire (PQ; Elliott et al. 2016) This is a semi-structured interview-based individualised outcome measure that asks patients to describe the main problem(s) that resulted in them seeking treatment. The intention is to prompt patients to elicit an unlimited number of personalised items, without any limitations on subject, where each item represents a unique problem of the patient expressed in their own words. After the item generation process, the patient is asked to rate each item/ problem according to how much it troubled him/her over the last week using a 7-point scale (ranging from '1 = not at all' to '7 = maximum possible') and for how long the problem has been of concern, also recorded on a 7-point rating scale (from '1 = less than 1 month' to '7 = more than 10 years'). PQ total score corresponds to the sum of the scorings of all items elicited by each patient with a possible score range depending on the number of items reported with no ceiling on the permissible number of items (e.g. for two items reported, the score will range from 2 to 14).

Psychological Outcome Profiles (PSYCHLOPS; Ashworth et al. 2004) This is a self-report individualised measure in which patients are invited to write down one or two personalised problems in free-text boxes, in response to the statements, "choose the problem that troubles you most" (P1) and "choose another problem that troubles you" (P2). Each problem is rated according to how much it has affected him/her over the last week, on a 6-point scale (from '0 = not at all affected' to '5 = severely affected'), and for duration on a 5-point scale (from '0 = under one month' to '4 = over five years'). PSYCHLOPS includes a third personalised item prompted by the statement, "choose one thing that is hard to do because of your problem or problems" (functioning), also rated on a 6-point scale (from '0 = not at all hard' to '5 = very hard'). There is a final standardised item where patients are asked to rate the question, "how have you felt in yourself this last week" on a 6-point scale (from '0 = very good' to '5 = very bad') (well-being). PSYCHLOPS total score corresponds to the sum of the scores of the two problems, the functioning and well-being items. In case P2 is missing, the score of P1 should be doubled ('pro-rated'); if either P1 or the functioning or well-being items are missing, the total score cannot be computed. Scores range from 0 to 20.

Clinical Outcomes in Routine Evaluation–Outcome Measure (CORE-OM; Evans et al. 2002) This is a self-report standardised 34-item tool to assess the level of psychological distress in four domains: problems (depression, anxiety, physical problems and trauma), functioning (general daily functioning, relationships), subjective well-being (feelings about self and optimism about the future) and risk behaviours (risk to self and to others). All items are scored on a 5-point scale (from '0 = not at all' to '4 = most or all the time', or the opposite, for reversed items) based on the number of times that each item was experienced by patients in the previous week. CORE-OM total score corresponds to the sum of the 34 items and ranges from 0 to 136. If more than three items are missing, the total score cannot be computed. Regarding the four domains, if more than one item is missing from problems and well-being, these total scores cannot be computed as well.

Patient Health Questionnaire-9 (PHQ-9; Kroenke et al. 2001) This is a self-report standardised 9-item questionnaire to evaluate depression. In this measure, patients are asked to score each item in a 4-point scale (from '0 = not at all' to '3 = nearly every day'), according to the frequency that each problem troubled the patient in the last 2 weeks. PHQ-9 total score corresponds to the sum of the 9 items, ranging from 0 to 27.

Treatment Outcomes Profile (TOP; Marsden et al. 2008) This is a standardised questionnaire orally administered by researchers, in an interview, that includes 20 questions about key aspects of substance misuse treatment: substance use, injecting risk behaviour, crime, health and quality of life. In TOP, items are rated in number of days, yes/no questions and scaling items ranging from 1 to 20 points. Three items within TOP can be used for comparison with other psychometric instruments: psychological health, physical health and quality of life scales.

Procedure

Researchers invited patients attending the treatment centres to take part in the study prior to commencing treatment. Written consent was obtained and participants completed the evaluation protocol immediately prior to the first treatment session.

Patients were presented with the evaluation protocol containing the five outcome measures, followed by a socio-demographic and treatment history form. Patients were offered assistance if they needed additional help with self-completion questionnaires. We randomised the order in which instruments were presented to patients, to control for sequencing effects and avoid bias arising from patient responses in one instrument influencing responses in a subsequent instrument. The presentation order of TOP was not randomised. TOP was invariably presented at the end of the protocol because it focusses specifically on drug-related topics and little or no influence was expected.

Data Analysis

Feasibility We focussed on three main parameters to evaluate the feasibility of individualised measures: administration duration in minutes (note: we opted for timing the entire protocol, instead of each measure separately, to interfere as little as possible with the evaluation process, considering that only one researcher was present in the room); acceptability, i.e. proportion of patients who completed each instrument; and the preferred mode of administration, i.e. proportion of patients who were able to complete by self-report or who required assistance from a research administrator (upon patient's request).

Descriptive Analyses of PQ and PSYCHLOPS PQ and PSYCHLOPS data were analysed in terms of mean number of patient-reported items, mean severity of distress and mean problem duration since onset. Because of the unpredictability of the number of patient-reported items in PQ, overall mean severity and duration were based on the individual PQ score mean for each patient. Data on problem duration were also recorded based on the longest reported duration in PQ and PSYCHLOPS.

Internal Reliability of PQ and PSYCHLOPS For PQ, we computed consecutive Cronbach's alpha values, starting with 2 items (PQ item #1 and PQ item #2) and then consecutively calculating

new alphas by adding the next PQ item (one at a time) to the calculation. A total of four alphas were computed until the last item mentioned by at least 15 patients (i.e. PQ item #5) was added. The mean (and respective standard deviation) of the alpha for PQ represents the mean of the four alphas that resulted from this procedure, following the procedure described by Elliott et al. (2016). For PSYCHLOPS, we opted for two alpha values, one for the three individualised items and one for the total score.

Convergent Validity of PQ and PSYCHLOPS Correlations between instrument scores were calculated using Pearson's r. The mean scores of PQ and PSYCHLOPS were correlated with CORE-OM, PHQ-9 and TOP (psychological health, physical health and quality of life scale items). As in Elliott et al. (2016), we expected moderate to strong correlations between individualised and standardised measures. Unlike CORE-OM and PHQ-9, the scale items of TOP have their scores reversed (i.e. $0 = $ very bad and $20 = $ very good) in contrast to PQ and PSYCHLOPS; hence, convergent validity would be indicated by a negative correlation.

In addition to feasibility and psychometrics of the individualised measures, we also explored the association between PQ and PSYCHLOPS scores with socio-demographic information, substance misuse and treatment history variables. The objective of these further analyses was to extend our knowledge about the scoring outcome data generated by individualised instruments, as well as investigating potential variables that might influence data interpretation. PQ and PSYCHLOPS scores were also contrasted according to their mode of administration (interview vs. self-report) and order of instrument administration (severity of scores in the first and second individualised completed by patients). For binary variables, the Student's t test was used; for nominal variables with more than two values, one-way ANOVAs (followed by Bonferroni post hoc analyses) were computed; for continuous data, Pearson's correlation was calculated.

All analyses were conducted in IBM SPSS Statistics 22.

Results

The 93 patients who participated in our study consisted of a higher proportion of men (59%); the mean age was 43 years (SD = 11). Tables 1, 2 and 3 show further characteristics of the sample, as well as their substance use and treatment history, respectively.

Feasibility of Individualised vs. Standardised Outcome Measures

The mean time required for completion of the five instruments within the evaluation protocol was 30.1 min (SD = 11.4, range 10–61 min). Regarding the acceptability, the number and proportion of patients who responded to each instrument were as follows: PQ, 92 (99%); PSYCHLOPS, 88 (95%); CORE-OM, 92 (99%); PHQ-9, 93 (100%); TOP, 90 (97%). Concerning the mode of administration, PQ and TOP were delivered in an interview-based format, as required by their instructions. The self-report tools often required research administrator completion. PSYCHLOPS was administered orally to 37 (41%) patients, CORE-OM to 36 (40%) patients and PHQ-9 to 35 (39%) patients. Patients requested assistance because of literacy issues, self-reported anxiety and poor sight.

Table 1 Patient characteristics ($N = 93$)

Variable	Description	N (%)
Relationship status	Single	45 (48.4%)
	Married/cohabiting	19 (20.4)
	Divorced/separated	29 (31.2%)
Educational level	Illiterate	3 (3.2%)
	Literate, but not completed primary school	1 (1.1%)
	Basic school (years 1–9)	64 (68.8%)
	Secondary school (years 10–12)	15 (16.1%)
	University (years > 12)	10 (10.8%)
Employment status	Paid work (full-time/part-time)	26 (28.0%)
	Unemployed	52 (55.9%)
	Retired	8 (8.6%)
	Other (e.g. housewife/househusband)	7 (7.5%)
Accommodation status	Rented/owned house	84 (90.3%)
	Temporary accommodation (e.g. guesthouse)	1 (1.1%)
	Homeless	2 (2.2%)
	Shared supported accommodation (e.g. community shelters)	6 (6.5%)

Descriptive Analyses of PQ and PSYCHLOPS Mean Scores

The mean number of items in PQ was 2.9 (SD = 2.1), range 1 to 12. The mean severity score for PQ items was 5.3 (SD = 1.7). Eighty-two (89%) patients had mean PQ severity scores above the clinical cutoff value of 3.25 points, as defined in Elliott et al. (2016). Problems described in PQ had lasted, on average, 1 to 2 years; however, the longest duration problem was 3 to 5 years.

For PSYCHLOPS, the mean number of patient-generated items (P1 + P2 + function) was 2.3 (SD = .8), range 0 to 3 items. The mean item severity scores were as follows: problems, 7.1 (SD = 2.9, range 0–10); function, 3.4 (SD = 1.7, range 0–5); well-being, 2.3 (SD = 1.8). The mean total PSYCHLOPS score was 13.8 (SD = 4.9, range 0–20). Even though we collected 88 PSYCHLOPS, the total score could only be computed for 58 (66%) because of insufficient data (i.e. two problems missing and/or function, and/or well-being item missing). The duration

Table 2 Substance misuse—age of onset and usage in the last 12 months ($N = 93$)

Type of substance		Years: mean (standard deviation)	n (%)
Alcohol	Age of onset	16.5 (7.7)	
	Used in the last 12 months (yes)		58 (62.4%)
Heroin	Age of onset	20.2 (6.7)	
	Used in the last 12 months (yes)		10 (10.8%)
Cocaine (powder)	Age of onset	20.3 (6.7)	
	Used in the last 12 months (yes)		4 (4.3%)
Crack cocaine	Age of onset	28.0 (12.8)	
	Used in the last 12 months (yes)		1 (1.1%)
Cannabis	Age of onset	14.7 (2.5)	
	Used in the last 12 months (yes)		14 (15.1%)
Ecstasy	Age of onset	16.3 (1.7)	
	Used in the last 12 months (yes)		1 (1.1%)
Tobacco	Age of onset	14.1 (3.3)	
	Used in the last 12 months (yes)		58 (62.4%)

Table 3 Substance misuse treatment history

Substance misuse treatment history		n (%)
First treatment episode		49 (52.7%)
Type of previous treatment	Psychological treatment	15 (17.4%)
	Opioid substitution treatment	7 (8.1%)
	Alcohol dependence medication	11 (12.8%)
	Inpatient therapeutic community	13 (15.1%)
	Detoxification unit	17 (19.8%)
	Self-help groups	4 (4.7%)

of problems reported in PSYCHLOPS was 3 months to 1 year. In this sample, the PSYCHLOPS problem of the longest duration was 1 to 5 years.

Internal Reliability of PQ and PSYCHLOPS Mean Scores

Both individualised measures showed good internal reliability, with alpha scores for PQ of .79 (mean) and for PSYCHLOPS .72 (Table 4).

Convergent Validity of PQ and PSYCHLOPS vs. Standardised Outcome Measures

Table 5 shows the correlation matrix between the five instruments. We found a weak and significant convergence between PQ and PSYCHLOPS ($r = .28$) total scores. Correlations between PQ and PSYCHLOPS problem scores were similar when self-completed ($r = .25$) or when researcher administered ($r = .25$). The correlations between PQ and the total scores of CORE-OM, PHQ-9 and TOP (psychological health) were weak at .27, .22 and .21 respectively (see Table 5). In contrast, correlations between CORE-OM, PHQ-9 and TOP (psychological health) with PSYCHLOPS were moderate at .50, .39 and .40, respectively.

Association Between PQ/PSYCHLOPS Scores and Demographic and Clinical Variables

Socio-Demographic Data There were no significant differences in PQ and PSYCHLOPS scores according to gender, educational level, type of accommodation and employment status. No correlation was found between item severity and patient age.

Substance Use and Treatment History There was no significant correlation between age when patients started to use certain substances and their distress scores on individualised

Table 4 Internal reliability (Cronbach's alpha) of PQ and PSYCHLOPS

Cronbach's alpha values				
PQ	PQ items 1–2 .69	PQ items 1–3 .75	PQ items 1–4 .83	PQ items 1–5 .88	Mean (SD) .79 (.09)
PSYCHLOPS	Problem items 1–2	Problem + functioning items 1–3 (individualised component)		All items 1–4 (individualised items + standardised item)
	.42	.69		.72

In PQ, item #1 corresponds to the first item in the PQ item list, i.e. the item identified by the patient as being the problem he/she is most concerned about; item #2 is the second problem that most concerns the patient, etc.

Table 5 Convergent validity of PQ and PSYCHLOPS with CORE-OM, PHQ-9 and TOP

Measures	Domains	PQ (n = 92)	PSYCHLOPS (n = 58)			
			Problems	Functioning	Well-being	Total
PSYCHLOPS	Problems	.35**				
	Functioning	.23	.53**			
	Well-being	.17	.39**	.27*		
	Total score	.28**	.89**	.73**	.70**	
CORE-OM	Problems	.31**	.39**	.38**	.49**	.48**
	Functioning	.12	.20	.31*	.53**	.39**
	Well-being	.30**	.41**	.35**	.52**	.49**
	Risk	.19	.32**	.16	.37**	.37**
	Total score	.27**	.38**	.37**	.54**	.50**
PHQ-9		.22*	.39**	.28*	.52**	.39**
TOP	Psychological health	− .21	− .33**	− .12	− .60**	− .40**
	Physical health	− .12	− .15	− .01	− .26*	− .07*
	Quality of life	− .24*	− .30**	− .30*	− .48	− .44**

*$p \le .05$; **$p \le .01$

outcome measures. PQ and PSYCHLOPS score correlations with age of first substance misuse for alcohol were $r = .11$, $p = .33$; $r = .27$, $p = .06$ (respectively) and for tobacco were $r = .01$, $p = .95$; $r = .02$, $p = .89$ (respectively); numbers were too small to determine age correlations for other substances (see Table 2).

Patients who had received previous treatment for substance misuse were more likely to report a higher level of psychological distress both in PQ (t (90) $= − 1.9$, $p < .05$) and PSYCHLOPS (t (56) $= − 2.5$, $p < .02$). Regarding the type of treatments received, patients with previous psychological outpatient treatment (t (50) $= 2.0$, $p < .04$) and in therapeutic communities (t (50) $= 2.4$, $p < .02$) had significantly higher problem scores in PSYCHLOPS than those who did not. These effects were not observed in PQ. There were no significant differences in mean scores of PQ and PSYCHLOPS problems comparing those receiving and not receiving the remaining treatment modalities.

Mode of Administration PSYCHLOPS scores were significantly higher among those who completed free-text descriptive items about problems in addition to scoring these items (t (84) $= 2.5$, $p < .01$) and functioning (t (56) $= 2.4$, $p < .02$) in self-completed format.

Order of Instrument Administration There were no significant differences in mean total severity scores relating to the order of individualised instrument completion. Comparing the sequence of administration for PSYCHLOPS, the ANOVA F value $= .32$, $p = .81$; for PQ, $F = .71$, $p = .55$.

Discussion

Previous findings have demonstrated the importance of PQ and PSYCHLOPS as sources of qualitative, personalised information about patients undergoing treatment for substance misuse (Alves et al. 2018). The present study adds to that body of literature by providing estimates

about the practicability, reliability and validity of such individualised outcome measures in this context.

Are Individualised Measures Feasible for Use in Substance Misuse Treatment?

We found that the administration of the whole evaluation protocol took approximately 30 min (i.e. approximately half the typical duration of a clinical session). This meant that individualised measures were completed relatively quickly, suggesting that adopting one or other of these measures might not be a burden in real clinical settings. Replicating previous findings in a primary care mental health setting (Ashworth et al. 2005), the completion rates of individualised measures were above 90%, demonstrating high levels of patient acceptability and comparable to completion rates of the standardised measures included in our study, showing that both types of measures are equally received by patients. However, nearly half our sample missed at least two responses in PSYCHLOPS, hindering the calculation of total scores and raising concerns about its feasibility as a self-report instrument. This was a low number in comparison with previous studies, where more than 90% of patients' responses to PSYCHLOPS allowed total score calculation (Ashworth et al. 2004, 2005, 2009). In other studies, PSYCHLOPS has been administered as an interview measure, to ensure no missing data in populations with literacy issues (Harper Shehadeh et al. 2019).

Regarding the mode of administration, although designed to be self-completed, almost half the sample asked for PSYCHLOPS to be researcher administered. This may have been due to low literacy skills, high levels of patient distress on entering treatment and patients may have felt more supported by being able to talk about their problems during assessment (Harper Shehadeh et al. 2019). Patients reported higher severity scores in the written format, implying that they might feel less comfortable verbalising their distress or that responses may be moderated by perceptions of social disapproved, or even illegal, behaviours. Although contradictory, our results were consistent with Bowling's observation (Bowling 2005) that interviewer-based methods are likely to increase response accuracy, but self-report questionnaires may encourage the sharing of more sensitive information. Conversely, no difference was observed between self-report or researcher administered scores on the standardised measures CORE-OM and PHQ-9.

Overall, our findings indicate that PQ and PSYCHLOPS are practicable in substance misuse treatment settings. Nearly all patients invited to participate in the study were motivated to complete the individualised measures, even though not all responded to them in the expected format. In particular, there were many patients requesting the instruments to be administered verbally by the researcher. This suggests that, even when self-report measures such as PSYCHLOPS are selected, researchers and clinicians may have to enhance their interview skills to ensure that data collection is not negatively affected for patients requesting assistance.

Do Individualised Measures Generate Robust Psychometric Data?

Both PQ and PSYCHLOPS elicited fewer items than when used in other clinical populations. For instance, in Elliott et al. (2016), the mean number of items created across five samples of patients in general mental health treatment was 10, compared with three items in the present study. Similar findings were reported in the previous section, when we observed that patients in our sample tended to omit PSYCHLOPS items more often than other samples (Ashworth

et al. 2004, 2007, 2009). When combined, these results indicate that substance misuse patients provide less information in individualised measures than other clinical populations. Further research is needed to explore why this occurred. However, we hypothesise that being admitted for a specialist treatment, targeting a specific behaviour, might have led patients to think that it was inappropriate, or unnecessary, to disclose too much information about their lives. On the other hand, item scores and mean problem duration were similar to those reported in previous studies (Ashworth et al. 2005; Elliott et al. 2016), indicating that patients experience similar levels of psychological distress as in other generalist mental health samples, emphasising the importance of addressing psychological health in this population (Wanigaratne et al. 2005).

Concerning psychometric properties, our analyses revealed good internal reliability for PQ and PSYCHLOPS in the context of substance misuse therapy. Such reliability values were similar to those previously reported in Elliott et al. (2016) (.86) and Ashworth et al. (2005) (.79). On the other hand, the convergence between individualised and standardised outcome measures was weak to moderate. This might reflect the diversity of problems experienced by our sample of patients, often unrelated to health and including concerns such as housing or finances not covered by the standardised comparators used (Alves et al. 2007, 2018). PSYCHLOPS correlated more strongly with standardised measures than did PQ. This might be a feature of the structural characteristics of PSYCHLOPS which is a hybrid instrument containing one standardised well-being item which correlated more strongly with the standardised measures, whilst PQ is a pure patient-generated instrument, only containing personalised items.

A weak correlation was found between the two individualised measures. This was expected for several reasons: the different structures of PQ and PSYCHLOPS may have prompted patients to respond differently to both measures; a previous study (Alves et al. 2018) had already found that patients tend not to repeat items/problems when answering both instruments as part of the same evaluation protocol. Moreover, the original self-report format of PSYCHLOPS was altered to an interview format in approximately half of the sample. Considering that the information provided in a questionnaire depends on its administration mode (Bowling 2005), it is unknown how much PSYCHLOPS responses were affected by researcher completion. It remains possible that differences also arose because these instruments were measuring differing underlying constructs, in spite of both focussing on patient-reported 'problems'.

To sum up, even though individualised measures produce reliable measurements, their weak to moderate correlation with standardised measures suggests that they should complement standardised instruments rather than being used alone. Combining individualised and standardised measures has been suggested by authors working in other treatment contexts and may enhance their potential for outcome measurement (Barkham 2016; Sales and Alves 2012; Sales et al. 2014). As Möller (2009) put it, whilst standardised measures can objectively quantify psychological constructs, individualised information is able to capture "distinctive characteristics of individuals" (p. 2016), giving a more meaningful context to the numbers obtained with standardised tools.

Which Individual or Treatment Characteristics May Influence PQ and PSYCHLOPS Scores?

Patients with past experience of substance misuse treatment were more likely to present higher distress in PQ and PSYCHLOPS, in comparison with those starting their first treatment episode. On the one hand, relapse may be particularly associated with distress and disruptive

situations; also, patients who dropped out of previous treatment episode(s) without a successful therapeutic result may be more likely to have unresolved personal issues reported as problems on the individualised measures. Ramos and Brown (2008) identified that, among substance misusers, the main reasons leading to substance misuse and relapse are social situations, coping with negative emotions and interpersonal conflicts. These types of concerns were reported by the same sample in PQ and PSYCHLOPS (Alves et al. 2018), which is in line with our findings.

Regarding the age of onset of drug use, our study was partly in line with the literature, which has reported that the younger individuals start to use drugs, the more severe their drug-related problems (Hser et al. 2007). In support of this was the negative correlation between age of first use of cannabis and PQ; and first use of LSD and PSYCHLOPS. However, for crack and LSD in PQ and ecstasy in PSYCHLOPS, the reverse was true. For these patients, it is likely that other bio-psycho-social factors may have contributed for their dependence disorder, reinforcing the importance of a multidimensional evaluation of clinical status (Alves et al. 2018). Divergence in total PQ and PSYCHLOPS scores according to age of first use for different substances indicates that they may be measuring different aspects of patient distress.

Limitations

The main limitation of our study is that it has not included post-treatment scores. Contrasting pre-post treatment data would allow an estimation of further psychometric parameters of PQ and PSYCHLOPS, such as sensitivity to change and temporal structure. High sensitivity to change is a central characteristic of patient-generated instruments and requires further evaluation within the context of dependence disorders (Ashworth et al. 2009; Elliott et al. 2016). Feasibility could be explored further by comparing patient and staff preferences for different combinations of instruments. Further feasibility comparisons could be made between pre-therapy and during-therapy administration of both patient-generated instruments since each adopts a different format after the start of therapy, with PQ only rating severity and PSYCHLOPS changing to a during-therapy version. We only collected the administration time for the entire protocol. Separate times would have provided a better estimate of individualised measures' length and how they compare to their standardised counterparts. Moreover, we did not evaluate the quality of the items produced by patients in PQ and PSYCHLOPS which may have contributed to divergent results.

Conclusion

With this study, we have confirmed that individualised measures are feasible and reliable to use as outcome measures in substance misuse treatment, in line with previous studies on the psychometrics of PQ and PSYCHLOPS in other settings (Ashworth et al. 2005, 2009; Elliott et al. 2016). Convergent validity between individualised and standardised measures was weak to moderate, suggesting that standardised and individualised measures may be measuring different aspects of patient distress.

A further finding was the weak correlation between PQ and PSYCHLOPS and the variability of results produced by them, when in fact these measures were expected to measure distress due to similar constructs—patient's individual concerns. Researcher and self-administered modes of completion may have influenced item generation, potentially

leading to the differences between the scorings of PQ and PSYCHLOPS. The possibility remains that the two individualised measures were measuring different aspects of patient distress.

In summary, our study suggests that PQ and PSYCHLOPS contribute to the measurement of patient distress in substance misuse treatment. Both individualised measures are reliable and have high levels of patient acceptability. Low to moderate convergence with standardised instruments suggests that optimal outcome measurement may involve combining individualised and standardised measures since this may broaden the capture of underlying constructs.

Compliance with Ethical Standards

Conflict of Interest   All four authors declare that they have no conflict of interest. MA led the team which devised the mental health outcome measure, PSYCHLOPS. However, he has no financial interest in this outcome measure which is available free of charge to all visitors to the website.

Ethical Approval   All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

References

Alves, P., Sales, C., & Ashworth, M. (2013). Enhancing the patient involvement in outcomes: a study protocol of personalised outcome measurement in the treatment of substance misuse. *BMC Psychiatry, 13*, 337–349. https://doi.org/10.1186/1471-244X-13-337.

Alves, P., Sales, C., & Ashworth, M. (2015). Personalising the evaluation of substance misuse treatment: a new approach to outcome measurement. *International Journal of Drug Policy, 26*, 333–335. https://doi.org/10.1016/j.drugpo.2014.11.014.

Alves, P., Sales, C., & Ashworth, M. (2016). "It is not just about the alcohol": a focus group to explore service users' views about outcome measurement. *Substance Abuse Treatment, Prevention, and Policy, 11*, 1–7. https://doi.org/10.1186/s13011-016-0070-5.

Alves, P., Sales, C., & Ashworth, M. (2017). Does outcome measurement of treatment for substance use disorder reflect the personal concerns of patients? A scoping review of measures recommended in Europe. *Drug and Alcohol Dependence, 179*, 299–308. https://doi.org/10.1016/j.drugalcdep.2017.05.049.

Alves, P., Sales, C., Ashworth, M., & Faisca, L. (2018). "There are things I want to say but you do not ask": a comparison between standardised and individualised evaluations in substance use treatment. *International Journal of Mental Health and Addiction*, 1–14. https://doi.org/10.1007/s11469-018-9985-6.

Ashworth, M., Guerra, D., & Kordowicz, M. (2019). Individualised or standardised outcome measures: a co-habitation? *Administration and Policy in Mental Health, 46*, 425–428. https://doi.org/10.1007/s10488-019-00928-z.

Ashworth, M., Shepherd, M., Christey, J., Matthews, V., Wright, K., Parmentier, H., Robinson, S., & Godfrey, E. (2004). A client-generated psychometric instrument: the development of "PSYCHLOPS". *Counselling and Psychotherapy Research, 4*, 27–31. https://doi.org/10.1080/14733140412331383913.

Ashworth, M., Robinson, S., Godfrey, E., Shepherd, M., Evans, C., Seed, P., Parmentier, H., & Tylee, A. (2005). Measuring mental health outcomes in primary care: the psychometric properties of a new patient-generated outcome measure, 'PSYCHLOPS' ('psychological outcome profiles'). *Primary Care Mental Health, 3*, 261–270.

Ashworth, M., Evans, C., & Clement, S. (2009). Measuring psychological outcomes after cognitive behaviour therapy in primary care: A comparison between a new patient generated measure 'PSYCHLOPS' and 'HADS'. *Journal of Mental Health, 18*, 169–177. https://doi.org/10.1080/09638230701879144.

Barkham, M. (2016). Patient-centered assessment in psychotherapy: toward a greater bandwidth of evidence. *Clinical Psychology: Science and Practice, 23*, 284–287. https://doi.org/10.1111/cpsp.12163.

Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health, 27*, 281–291. https://doi.org/10.1093/pubmed/fdi031.

Darke, S., Ross, J., & Teesson, M. (2007). The Australian Treatment Outcome Study (ATOS): what have we learnt about treatment for heroin dependence? *Drug and Alcohol Review, 26*, 49–54.

Elliott, R., Wagner, J., Sales, C., Rodgers, B., Alves, P., & Café, M. (2016). Psychometrics of the personal questionnaire: a client-generated outcome measure. *Psychological Assessment, 28*, 263–278. https://doi.org/10.1037/pas000017.

Evans, S., Greenhalg, J., & Connelly, J. (2000). Selecting a mental health needs assessment scale: guidance on the critical appraisal of standardized measures. *Journal of Evaluation in Clinical Practice, 6*, 379–393. https://doi.org/10.1046/j.1365-2753.2000.00269x.

Evans, C., Connell, J., Barkham, M., Mellor-Clark, J., & Audin, K. (2002). Towards a standardised brief outcome measure: psychometric properties and utility of the CORE-OM. *British Journal of Psychiatry, 180*, 51–60. https://doi.org/10.1192/bjp.180.1.51.

Fitzpatrick, R., Davey, C., Buxton, M. J., & Jones, D. R. (1998). Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment, 2*, 1–86. https://doi.org/10.3310/hta2140.

Hser, Y., Longshore, D., & Anglin, D. (2007). The life course perspective on drug use: a conceptual framework for understanding drug use trajectories. *Evaluation Review, 31*, 515–547.

Gossop, M., Marsden, J., Stewart, D., & Kidd, T. (2003). The National Treatment Outcome Research Study (NTORS): 4-5 year follow-up results. *Addiction, 98*, 291–303.

Harper Shehadeh, M., van't Hof, E., van Ommeren, M., Farooq, S., Hamdani, S. U., Koyiet, P., Akhtar, P., Masood, A., Nazir, H., Dawson, K., & Albanese, E. (2019). Using a person-generated mental health outcome measure in large clinical trials in Kenya and Pakistan: self-perceived problem responses in diverse communities. *Transcultural Psychiatry*. https://doi.org/10.1177/1363461519854831.

Jones, A., Donmall, M., Millar, T., Moody, A., Weston, S., Anderson, T., Gittins, M., Abeywardana, V., & D'Souza, J. (2009). *The drug treatment outcomes research study: final outcomes report*. London: Home Office.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*, 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

Lambert, M., & Hawkins, E. (2004). Measuring outcome in professional practice: considerations in selecting and using brief outcome instruments. *Professional Psychology: Research and Practice, 35*, 492–499. https://doi.org/10.1037/0735-7028.35.5.492.

Marsden, J., Farrell, M., Bradbury, C., Dale-Perera, A., Eastwood, B., Roxburgh, M., & Taylor, S. (2008). Development of the treatment outcomes profile. *Addiction, 103*, 1450–1460. https://doi.org/10.1111/j.1360-0443.2008.02284.x.

Möller, H. (2009). Standardised rating scales in psychiatry: methodological basis, their possibilities and limitations and descriptions of important rating scales. *The World Journal of Biological Psychiatry, 10*, 6–26. https://doi.org/10.1080/15622970802264606.

Neale, J., & Strang, J. (2015). Blending qualitative and quantitative research methods to optimize patient reported outcome measures (PROMs). *Addiction, 10*, 1215–1216. https://doi.org/10.1111/add.12896.

OECD. (2013). *OECD guidelines on measuring subjective well-being*. OECD Publishing.. https://doi.org/10.1787/9789264191655-en.

Ramos, D., & Brown, S. (2008). Classes of substance abuse relapse situations: A comparison of adolescents and adults. *Psychology of Addictive Behaviors, 22*, 372–9.

Sales, C., Goncalves, S., Fragoeiro, A., Noronha, S., & Elliott, R. (2007). Psychotherapists' openness to routine naturalistic idiographic research. *Mental Health and Learning Disabilities Research and Practice, 4*, 145–161. https://doi.org/10.5920/mhldrp.2007.42145.

Sales, C., & Alves, P. (2012). Individualised patient-progress systems: why we need to move towards a personalized evaluation of psychological treatments. *Canadian Psychology, 53*, 115–121. https://doi.org/10.1037/a00280537.

Sales, C. M. D., Alves, P. C. G., Evans, C., & Elliott, R. (2014). The Individualised Patient Progress System (IPPS): a decade of international collaborative networking. *Counselling and Psychotherapy Research, 14*, 181–191. https://doi.org/10.1080/14733145.2014.929417.

Sales, C., & Alves, P. (2016). Patient centred assessment in psychotherapy: a review of individualised tools. *Clinical Psychology: Science and Practice, 23*, 265–283. https://doi.org/10.1111/cpsp.12162.

Wanigaratne, S., Davis, P., Pryce, K., & Brotchie, J. (2005). *The effectiveness of psychological therapies on drug misusing clients*. London: NHS National Treatment Agency for Substance Misuse.

Affiliations

Paula Alves[1] · Luís Faísca[2] · Célia M. D. Sales[3] · Mark Ashworth[4]

[1]   Research Department of Primary Care & Population Health, University College London, Upper 3rd Floor, Royal Free Campus, Rowland Hill Street, London NW3 2PF, UK

[2]   Faculty of Human and Social Sciences, Department of Psychology and Educational Sciences, Algarve University, Estr. da Penha 139, 8005-139 Faro, Portugal

[3]   Faculty of Psychology and Education Sciences at the University of Porto (FPCEUP), Center for Psychology at the University of Porto (CPUP), University of Porto, R. Alfredo Allen, 4200-135 Porto, Portugal

[4]   School of Population Health and Environmental Sciences, King's College London, Guy's Campus, Addison House, London SE1 1UL, UK