Drug and Alcohol REVIEW

APSAD

Drug and Alcohol Review (2020)
DOI: 10.1111/dar.13050

Validation of the Australian Treatment Outcomes Profile for use in clients with cannabis dependence

LLEWELLYN MILLS[1,2,3], NICHOLAS LINTZERIS[1,2,3], RAIMONDO BRUNO[4], MARK MONTEBELLO[2,3,5,6], ADRIAN DUNLOP[3,6,7,8], RACHEL M. DEACON[1,2,3], JAN COPELAND[9], MERYEM JEFFERIES[3,10], CONSUELO RIVAS[1,3] & KRISTIE MAMMEN[1,3]

[1]Drug and Alcohol Services, South Eastern Sydney Local Health District, Sydney, Australia, [2]Discipline of Addiction Medicine, Faculty Medicine and Health, University of Sydney, Sydney, Australia, [3]NSW Drug and Alcohol Clinical Research and Improvement Network, Sydney, Australia, [4]School of Medicine, University of Tasmania, Hobart, Australia, [5]Drug and Alcohol Services, North Sydney Local Health District, Sydney, Australia, [6]National Drug and Alcohol Research Centre, UNSW Sydney, Sydney, Australia, [7]Drug and Alcohol Services, Hunter New England Local Health District, Newcastle, Australia, [8]Hunter Medical Research Institute, University of Newcastle, Newcastle, Australia, [9]Mind and Neuroscience Thompson Institute, University of the Sunshine Coast, Sunshine Coast, Australia, and [10]Drug Health, Western Sydney Local Health District, Sydney, Australia

Abstract

Introduction and Aims. *The Australian Treatment Outcomes Profile (ATOP) was developed as a clinical tool for monitoring the substance use, health and wellbeing of clients in alcohol and other drug treatment. This is the first psychometric validation of the ATOP in a cannabis-dependent treatment population.* ***Design and Methods.*** *A total of 128 individuals with cannabis dependence enrolled in an outpatient randomised controlled trial were administered the ATOP and gold-standard health and wellbeing questionnaires once by clinicians and once by researchers at baseline. Concurrent validity was assessed by testing ATOP Psychological Health, Physical Health and Quality of Life questions against concurrently administered gold-standard questionnaires: the Short Form 36 Health Survey (SF-36), the 21-item Depression, Anxiety and Stress Scale (DASS-21) and the Sheehan Disability Scale (SDS). Interrater reliability was tested by comparing clinician-administered ATOP items at the medical screening interview to the same ATOP items administered by researchers at baseline.* ***Results.*** *ATOP Psychological Health showed moderate to strong correlations with SF-36 Mental Components, SF-36 Mental Health and DASS-21 scores (r = 0.40–0.52) and ATOP Physical Health with SF-36 Physical Components and SF-36 General Health scores (r = 0.36–0.67). The ATOP Quality of Life scale showed moderate agreement with the SDS and six-dimensional health state short form scales (r = 0.38–0.40). ATOP substance use, employment, education and child care items showed good to excellent interrater reliability (Krippendorff's α = 0.62–0.81), and tobacco use, Psychological Health, Physical Health and Quality of Life showed fair to moderate interrater reliability (Krippendorff's α = 0.42–0.53).* ***Discussion and Conclusions.*** *The ATOP appears to be valid and reliable when tested in a population with cannabis-dependence, justifying its widespread use in clinical settings.* [Mills L, Lintzeris N, Bruno R, Montebello M, Dunlop A, Deacon RM, Copeland J, Jefferies M, Rivas C, Mammen K. Validation of the Australian Treatment Outcomes Profile for use in clients with cannabis dependence. *Drug Alcohol Rev* 2020]

Key words: clinical outcome monitoring, health service evaluation, psychometric validation, cannabis dependence, addiction treatment.

Introduction

There is a need in the alcohol and other drug (AoD) treatment sector for standardised instruments that assess substance use, health and general wellbeing in order to monitor clinical outcomes and facilitate individual client care planning over time. The 22-item Australian Treatment Outcomes Profile (ATOP) [1] is a brief clinical assessment tool adapted for Australian conditions from the Treatment Outcomes Profile

Llewellyn Mills PhD, Research Fellow, Nicholas Lintzeris MBBS, PhD, Director, Raimondo Bruno PhD, Associate Professor, Mark Montebello MBBS, PhD, Director, Adrian Dunlop MBBS, PhD, Director, Rachel M. Deacon PhD, Research Associate, Jan Copeland PhD, Adjunct Professor, Meryem Jefferies PhD, Research Officer, Consuelo Rivas BNurs, Research Nurse, Kristie Mammen BPsych (Hons), Project Manager. Dr Llewellyn Mills, The Langton Centre, Cnr Nobbs and South Dowling Streets, Surry Hills, NSW 2006, Australia. Tel: +61 2 9332 8734; E-mail: llew.mills@sydney.edu.au.

Received 25 September 2019; accepted for publication 11 February 2020.

(TOP) [2], developed in the United Kingdom. Created for the purposes of monitoring clinical outcomes in key domains of substance use, injecting behaviour, social functioning (education, employment, housing), criminal activity, psychological health, physical health and quality of life, the TOP has demonstrated good validity, reliability and sensitivity to change in validation studies in opioid- and cocaine-using populations [2]. Widespread use of the TOP across the British AoD treatment sector has allowed system-wide clinical outcome monitoring to take place [3,4].

The ATOP has also demonstrated good psychometric properties in studies validating its use in alcohol dependent and opioid-dependent treatment seekers. When originally tested in an opioid treatment population [1] the ATOP showed good interrater reliability (i.e. reliability coefficients ≥0.6) across all variables except crime and sharing injecting equipment, and good concurrent validity, with ATOP variables correlating strongly (i.e. $r \geq 0.5$) with gold standard questionnaires that measured similar constructs such as the Patient Health Questionnaire-15 [5], Kessler Psychological Distress Scale [6] and World Health Organization Quality of Life scale (WHOQOL-BREF) [7]. A more recent study assessed the concurrent validity of ATOP frequency of substance use, psychological health, physical health and quality of life variables in a population of older clients in AoD treatment for either opioids or alcohol dependence [8]. These variables correlated strongly with scales measuring similar constructs such as the Alcohol Use Disorders Identification Test [9] (ATOP alcohol frequency of use), the Geriatric Depression Scale [10] (ATOP Psychological Health and Quality of Life) and the 12-Item Short Form Health Survey [11] (ATOP Psychological Health and Physical Health).

The ATOP has proven to be an instrument that can be integrated into routine clinical care, evidenced by: (i) positive clinician and client ratings concerning its clinical utility and ease of administration [1]; and (ii) its incorporation into the electronic medical records system of government AoD services across Australia. The ATOP has also proven to be a useful research tool, with multiple studies using it to measure outcomes related to substance use, health and wellbeing [12–17].

For a standardised measurement tool to be useful it must be able to accurately measure what it claims to measure across different settings and people. Neither the TOP nor the ATOP has been psychometrically validated for use with treatment-seekers whose primary drug of concern is cannabis. Cannabis dependence is the second most prevalent drug of dependence after alcohol, with approximately 1 in 100 Australians (1%) meeting criteria for either harmful use or dependence [18] and is the third most common principal drug of concern by proportion of closed treatment episodes in Australia in 2017–2018 [19]. It is important that the ATOP be psychometrically validated for use in this large group of treatment seekers.

Our clinical experience suggests that the profile of clients attending AoD services for treatment of a primary cannabis use disorder (primary drug of concern) is often quite different to clients attending for alcohol or opioid use disorders—not only in their patterns of substance use, but also in their age, prevalence of social (e.g. employment, housing), mental and physical health concerns [19]. Given these differences in client profiles, it is important to ensure that a scale used in AoD services has been validated in clients attending primarily for cannabis related problems, rather than assume that what applies to heavy drinkers and heroin users will also apply to this different client group. This study continues the tradition of psychometrically validating scales in clients with different substance use profiles [20–22] by testing the validity and reliability of the ATOP in a population of individuals seeking treatment for cannabis dependence.

Methods

Design

This paper is a secondary analysis of data using a sample of participants enrolled in a 24-week double-blinded, outpatient randomised trial testing the efficacy of a cannabinoid-agonist medication (nabiximols) against placebo for the treatment of cannabis dependence (described in detail elsewhere [23,24]). The data used in this study were collected prior to any medication (placebo or nabiximols) being administered and all clients, clinicians and researchers were blinded to treatment allocation. Clients were screened by study medical officers (addiction medicine specialists) for eligibility, and interviewed by research staff at baseline (day 1; usually within 14 days of medical screening), during which the ATOP and a number of research instruments were administered (see below). This investigation examines concurrent validity (comparing ATOP against validated 'gold standard' instruments that measure the same constructs) and interrater reliability (comparing clinician and researcher ATOP scores) for participants who completed both the pre-trial medical screen and baseline research interview administered on day 1 of the trial phase of the study.

Participants

Participants were included in the study if they: (i) were between 18 and 65 years of age; (ii) were seeking

treatment for their cannabis use; (iii) fulfilled criteria for International Classification of Diseases, 10th edition (ICD-10) [25] cannabis dependence, as diagnosed by a specialist in addiction medicine; (iv) did not meet criteria for dependence on any other substance (other than nicotine or caffeine); (v) did not meet criteria for a severe and active medical or psychiatric disorder; and (vi) had not received treatment for cannabis dependence in the 4 weeks prior to study commencement [23].

Measures

The ATOP is a 22-item questionnaire designed to measure respondents' substance use and general health and wellbeing over the previous 28 days. For substance use variables [cannabis, alcohol, amphetamine-type substances (including methamphetamines), benzodiazepines, heroin, cocaine, other opioids and 'other substances'] respondents indicate how many days in the previous 28 they used, using a modified Timeline Follow Back technique [26]. Items also include the number of days in the previous 28 when participants injected drugs, participated in paid employment or undertook education or training. Binary yes/no questions measure whether or not respondents have consumed tobacco daily, shared injecting equipment, been homeless, at risk of eviction, the primary caregiver for any children under 16 years old, been arrested or been either the victim or the perpetrator of violence. Respondents' Psychological Health, Physical Health and Quality of Life over the previous 28 days are measured via a 0–10 Visual Analog Scale with higher scores indicating more positive outcomes.

The Short Form-36 (SF-36 [27]) is a health and wellbeing questionnaire used widely across many health and research settings. Its 36 items are grouped into eight domains representing a core set of generic health outcomes: physical function, role limitations due to physical problems, social functioning, bodily pain, general mental health, role limitations due to emotional problems, vitality and general health. Scores for these eight domains can also be used to calculate the composite Mental Components and Physical Components scores, which measure broad mental and physical health [28]. The factors of the SF-36 have well-established reliability [29] and construct validity, with relevant SF-36 factor scores showing sensitivity to differences in severity of clinically verified psychopathology and physical distress [30]. A composite quality of life measure, the six-dimensional health state short form (SF-6D) can also be calculated from a subset of 11-items within the SF-36 [31]. Like the SF-36, the SF-6D has well-established reliability and is sensitive to known differences in severity of health-related conditions [32,33]. SF-6D scores range from 0 to 1. Higher scores on SF-36 items and factors, and all composite measures derived from these, indicate more positive outcomes.

The Depression, Anxiety and Stress Scale (DASS) [34] is designed to measure severity of respondents' depression, anxiety and stress. The scale has proven reliability and concurrent validity in both clinical [35] and non-clinical [36] samples. A total DASS score can be calculated by summing all 21 items, yielding a score out of 63. This total score, which can be thought of as a 'composite measure of negative emotional symptoms' [37], has been shown to be sensitive to reductions in psychopathology experienced between admission and discharge in patients admitted to inpatient treatment programs in a psychiatric hospital [38]. Higher scores on the DASS indicate negative outcomes, that is, more severe psychopathology.

The Sheehan Disability Scale [39] uses three items to measure respondents' impairment of function in relation to work, social and family life. These scores can be summed to yield a composite score representing overall social and interpersonal functioning. The Sheehan Disability scale has well-established reliability and has the ability to detect clinician-rated differences in functioning [40–42]. Higher scores indicate more negative outcomes, that is, greater dysfunction.

All clinical and research staff involved in delivering the ATOP were trained in its administration in accordance with formal guidelines for its use, contained in the ATOP user manual written when the tool was developed [43]. All research staff involved in administering the research scales (e.g. SF-36, DASS-21) were all trained in their administration prior to commencement of the trial.

Statistical analysis

Concurrent validity. For ATOP Psychological Health, Physical Health and Quality of Life questions, concurrent validity was assessed by comparing scores for each of these variables to scores on gold standard questionnaires that measure equivalent constructs and have demonstrated the ability to detect either changes in these constructs within individuals (e.g. between admission and discharge to a psychiatric hospital) or differences in severity of the construct between groups of individuals (e.g. those diagnosed with a condition versus those without the condition). Scores for ATOP Psychological Health were compared to the SF-36 Mental Components score, SF-36 Mental Health

domain score and DASS total score. A DASS total score was chosen for comparison as it is a non-specific measure of negative emotional state and hence is more similar conceptually to the single ATOP item than any of the three condition-specific DASS factor scores. Scores for ATOP Physical Health were compared to SF-36 Physical Components Scores and to the SF-36 General Health factor (which was chosen over the SF-36 Physical Functioning factor score because, similar to the ATOP Physical Health question, items are phrased as more general statements about physical health). ATOP Quality of Life was compared to total scores for the Sheehan Disability Scale and the SF-6D. Pearson's correlation coefficients with 95% confidence intervals were used to measure the level of agreement between pairs of variables. All comparisons were made using questions within the test battery administered at the researcher baseline. Cohen's benchmarks were used to interpret Pearson's correlation coefficients (≥0.50 strong, 0.30–0.49 moderate, <0.30 weak) [44].

Interrater reliability. For all ATOP questions interrater reliability was assessed by comparing scores on ATOPs administered by clinicians during pre-trial screening to ATOPs administered by researchers at baseline. All study raters were blinded to treatment condition. Participants were excluded from this analysis if there was a greater than 14-day (i.e. more than 2 week) gap between their screening session and their baseline session. Krippendorff's alpha (α) test [45] was used as the reliability statistic due to its ability to yield accurate estimates of agreement when there are many raters; robustness to missing values; and flexibility of use for both continuous and categorical variables [46]. Krippendorff's α values range from 0 to 1, where 0 is perfect disagreement and 1 is perfect agreement. Cichetti and Sparrow's benchmarks for reliability coefficients were used to interpret α's (≥0.75 excellent, 0.60–0.74 good, 0.40–0.59 fair to moderate, < 0.40 poor) [47,48]. Alpha (α) values were not considered to be dependable if the number of participants who engaged in the outcome in question was less than 10. For all categorical variables, percentage agreement—percentage of total instances where both clinician and researchers gave the same rating—was also calculated.

All statistical analyses were performed in R version 3.4.1 [49], using tidyverse [50] and irr [51] packages.

Validity and reliability criteria. As the ATOP's purpose is a clinical review tool rather than a research tool, with single items measuring each construct, we decided that it was reasonable to impose less stringent criteria for judging variables to be valid and reliable than one might impose on a less wide-ranging research tool. Hence, based on similar minimum reliability thresholds used for clinical review tools in this sector [52,53], if variables met Cohen's moderate criteria (0.30–0.40) they were considered valid and if they met Cichetti and Sparrow's fair to moderate criteria (0.40–0.59) they were considered reliable.

Results

Participants

The medical screening and the researcher baseline battery of tests were each completed by all 128 participants who took part in the study. Demographic, substance use and psychosocial characteristics are presented in Table 1. All patients met ICD-10 criteria for cannabis dependence as a condition of study eligibility, with participants reporting a mean number of ICD-10 dependence symptoms of 7.8 ± 1.3 (out of a possible 9).

Concurrent validity

Results of the concurrent validity analyses are presented in Table 2. Based on Cohen's rules of thumb (≥0.50 strong, 0.30–0.49 moderate, <0.30 weak) there was moderate to strong agreement (r = 0.36–0.67) between the ATOP Psychological Health, Physical Health and Quality of Life scales and all comparison scales.

Interrater reliability

Twenty-three separate raters administered ATOPs for this study, with each participant completing an ATOP with two different raters. Nine participants were excluded from the interrater reliability analysis due to there being a greater than 14-day gap between their Medical Screen ATOP and their Researcher ATOP, leaving 119 participants. The average number of days' lag between the Medical Screen and Researcher Baseline was 6.7 ± 2.8 days. A histogram of distribution of lag times is included in Figure S1. The results of the interrater reliability analysis are presented in Tables 3 and 4. Variables where less than 10 participants engaged in the behaviour in question have been removed from these tables as reliability coefficients for these variables cannot be considered accurate. For full tables with all variables included see Tables S1 and S2.

After excluding the variables where the numbers of participants who engaged in the outcome in question

Table 1. *Demographic, substance use and psychosocial characteristics at baseline (*n = *128)*

	Placebo (*n* = 67)	Nabiximols (*n* = 61)	Total (*n* = 128)
Demographic variables			
Age, mean (SD), years	33.8 (10.3)	36.2 (11.5)	35.0 (10.9)
Female sex, *n* (%)	14 (20.9)	16 (26.2)	30 (23.4)
Born in Australia, *n* (%)	56 (83.6)	51 (83.6)	107 (83.6)
Aboriginal/Torres Strait Islander, *n* (%)	6 (9.0)	4 (6.6)	10 (7.8)
Tertiary educated, *n* (%)	22 (32.8)	27 (44.3)	49 (38.3)
Employment as main source of income, *n* (%)	36 (53.7)	35 (57.4)	71 (55.5)
In a relationship, *n* (%)	27 (40.3)	18 (29.5)	45 (35.2)
Have ≥1 child, *n* (%)	23 (34.3)	21 (34.4)	44 (34.4)
Current legal problems, *n* (%)	6 (9.0)	2 (3.3)	8 (6.2)
Cannabis use variables			
Number of days used in last 28, mean (SD)	25.6 (4.5)	25.9 (4.6)	25.7 (4.5)
Average grams per day used, mean (SD)	2.6 (2.5)	2.0 (1.4)	2.3 (2.1)
Age of first use, mean (SD); (range: 5–40)	15.0 (4.3)	16.0 (3.4)	15.5 (3.9)
Duration since first regular use, mean (SD); years	15.2 (9.8)	16.2 (9.9)	15.7 (9.8)
Psychosocial variables			
Sheehan Disability Scale, mean (SD), (Range: 0–30)	7.2 (1.1)	6.9 (1.2)	7.1 (1.2)
SF-36, mean (SD)			
SF-36 factor scores (range: 0–100)			
Physical functioning	87.2 (18.4)	86.2 (20.7)	86.7 (19.4)
Role limitations due to physical health	36.6 (42.3)	34.4 (41.4)	35.6 (41.7)
Role limitations due to emotional problems	48.3 (44.7)	48.1 (45.8)	48.2 (45.0)
Energy/fatigue	43.4 (18.3)	40.8 (19.3)	42.2 (18.8)
Mental health	56.1 (20.2)	54.7 (18.1)	55.1 (19.2)
Social functioning	54.9 (30.6)	59.1 (26.7)	56.6 (28.8)
Pain	69.8 (26.2)	70.5 (25.6)	70.1 (25.8)
General health	49.3 (19.7)	53.9 (21.6)	51.5 (20.6)
SF-36 composite scores			
Mental components (range: 0–100)	40.7 (6.9)	40.6 (6.6)	40.7 (6.7)
Physical components (range: 0–100)	52.4 (5.0)	52.7 (6.1)	52.5 (5.5)
SF-6D (range: 0–1)	0.68 (0.13)	0.66 (0.12)	0.67 (0.12)
DASS total score, mean (SD), (range: 0–63)	28.0 (13.5)	29.0 (14.4)	28.5 (13.9)

DASS, Depression, Anxiety and Stress Scale; SF-36, Short Form 36 Health Survey; SF-6D, six-dimensional health state short form.

were less than 10, the ATOP demonstrates good to excellent interrater reliability (Krippendorff's α = 0.62–0.81) except daily tobacco use, psychological health, physical health and quality of life, whose reliability was fair to moderate (Krippendorff's α = 0.42–0.53). It should be noted that though the interrater reliability α of tobacco use was low the percentage agreement was high, with both clinicians and researchers giving the same rating in 87% of cases.

We conducted a post-hoc sensitivity analysis examining how α's changed when different lag criterion were imposed, reported in full in Tables S3 and S4. When we imposed no restrictions on lag time α's were slightly lower, due perhaps to the inclusion of data where there were lags of between 30 and 77 days. Lags of this size increase the possibility that differences between ratings are due to actual change in the variables of interest rather than simply change in the way participants responded to the different raters. When we used a more stringent lag criteria of 7 days between the Medical Screen and the Researcher Baseline α's were also mostly smaller, which could be due to loss of power from the halving in sample size that resulted from imposing this restriction [54].

Discussion

The ATOP was designed as a brief clinical review tool that can measure multiple domains relevant to the health and wellbeing of clients in AoD treatment as

Table 2. *Baseline comparison of scores for Australian Treatment Outcomes Profile (ATOP) Psychological Health, Physical Health and Quality of Life to researcher gold standards (*n = *128)*

	ATOP, mean (SD)	Comparator scale	Comparator, mean (SD)	PCC[a]	95% confidence interval
Psychological Health	5.7 (2.1)	SF-36: Mental Components	40.7 (6.7)	0.40	0.24, 0.54
		SF-36: Mental Health	43.2 (8.7)	0.52	0.38, 0.64
		DASS: Total Score	28.5 (13.9)	−0.52	−0.64, −0.38
Physical Health	6.0 (2.0)	SF-36: Physical Components	52.5 (5.5)	0.36	0.20, 0.50
		SF-36: General Health	47.3 (9.8)	0.67	0.56, 0.75
Quality of Life	5.9 (2.3)	Sheehan Disability Scale	13.5 (7.8)	−0.38	−0.52, −0.22
		SF-6D	0.66 (0.11)	0.43	0.26, 0.57

[a]PCC, Pearson's correlation coefficient (≥0.50 strong, 0.30–0.49 moderate, <0.30 weak). Short Form 36 Health Survey (SF-36): Scores for the Mental and Physical Components and the Mental Health and General Health factors of the SF-36 are t-scores, with a mean of 50 and an SD of 10 (normed on US population). Higher scores indicate more positive outcomes. Six-dimensional health state short form (SF-6D): scores on the SF-6D range from 0 to 1. Australian population norm for age 18–30 = 0.79, 31–40 = 0.788, 41–50 = 0.77. Higher scores indicate more positive outcomes. Depression, Anxiety, and Stress Scale (DASS): DASS total score obtained by adding all 21 item scores. Higher scores indicate more severe psychopathology. Sheehan Disability Scale: total score summed across three 0–10 items, max = 30. Higher scores indicate more negative outcomes.

Table 3. *Interrater reliability estimates for Australian Treatment Outcomes Profile continuous variables*

	Any use, *n*[a]	Researcher, mean (SD)	Clinician, mean (SD)	α[b]
Cannabis, days used	119	25.6 (4.7)	25.9 (4.4)	0.624
Alcohol, days used	87	5.4 (7.7)	5.6 (7.6)	0.667
ATS, days used	28	0.3 (0.9)	0.4 (0.8)	0.688
Benzodiazepenes, days used	40	1.7 (5.6)	2.3 (6.0)	0.697
Cocaine, days used	16	0.2 (0.9)	0.2 (0.9)	0.780
Employment, days worked	73	10.0 (9.5)	9.2 (9.7)	0.814
Education, days in education	23	1.3 (3.7)	1.8 (5.0)	0.657
Psychological Health	119	5.7 (2.0)	6.0 (1.9)	0.419
Physical Health	119	6.0 (2.0)	6.4 (2.0)	0.422
Quality of Life	119	6.0 (2.2)	6.2 (2.0)	0.533

[a]*n* = number who indicated to either rater that they had engaged in the activity once or more during the previous 28 days. [b]Krippendorff's α reliability statistic (≥0.75 excellent, 0.60–0.74 good, 0.40–0.59 fair to moderate, < 0.40 poor). Variables where fewer than 10 participants engaged in the outcome in question (heroin use, other opioid use and injecting drug use) in the previous 28 days have been omitted (for full table see Table S1). ATS, amphetamine-type substances.

part of routine care, and to assist in treatment monitoring, care planning and outcome assessment. Our study shows that the ATOP has good validity and reliability for treatment-seeking clients dependent on cannabis. The ATOP has been validated for use in an opioid-dependent [1] and older client population [8] and now in a population of cannabis-dependent treatment seekers.

As with previous studies in populations with opioid dependence and alcohol dependence [1], ATOP Psychological Health, Physical Health and Quality of Life items showed moderate to strong agreement with all of the gold standard comparison measures. Correlations between ATOP Psychological Health and SF-36 Mental Components, SF-36 Mental Health factor and DASS total scores ranged from r = 0.40–0.52. ATOP Physical Health correlated more strongly with the SF-36 General Health Factor (r = 0.67) than with the SF-36 Physical Components score (r = 0.36–0.67), perhaps reflecting the fact that, similar to the single ATOP Physical Health questions, the items making up the General Health factor are phrased in quite general terms, whereas in the Physical Components scale there are more items and their content pertained more to specific health-related domains. The agreement between the ATOP scales and validated scales that measure similar constructs suggests that it is able to gauge, with an acceptable degree of accuracy, the health and overall quality of life of people with cannabis dependence.

Table 4. *Interrater reliability estimates for Australian Treatment Outcomes Profile (ATOP) dichotomous variables*

		Clinician ATOP				
	Researcher ATOP	No, *n*	Yes, *n*	Total, *n* (%)	α^{b}	% Agreement[a]
Daily tobacco use	No, *n*	8	10	18 (15.4)	0.445	87.2%
	Yes, *n*	5	94	99 (84.6)		
	Total, *n* (%)	13 (11.1)	104 (88.9)	117		
Caregiver of child(ren) < 5 years	No, *n*	93	3	96 (86.5)	0.813	95.5%
	Yes, *n*	2	13	15 (13.5)		
	Total, *n* (%)	95 (85.6)	16 (14.4)	111		
Caregiver of child(ren) (≥5 and < 15 years)	No, *n*	92	3	95 (85.6)	0.782	94.6%
	Yes, *n*	3	13	16 (14.4)		
	Total, *n* (%)	95 (85.6)	16 (14.4)	111		

[a]Percentage of total observations where clinicians and researchers agreed. Variables where fewer than 10 participants engaged in the outcome in question (shared injecting equipment, homelessness, at risk of eviction, arrested, victim of violence and perpetrated violence) in the previous 28 days have been omitted (for full table see Table S2). [b]Krippendorff's α reliability statistic (≥0.75 excellent, 0.60–0.74 good, 0.40–0.59 fair to moderate, <0.40 poor).

For substances that were used by a sufficient number of participants (cannabis, alcohol, benzodiazepines, heroin, amphetamine-type substances, cocaine) interrater reliability was good to excellent, as it was for employment, education and caring for children. Several variables—ATOP Tobacco Use, Psychological Health, Physical Health and Quality of Life—were less reliable. In Ryan and colleagues' study [1], where the participants' primary drug of concern was opioids, reliability estimates for ATOP Psychological Health, Physical Health and Quality of Life variables range from 0.68 to 0.76. By comparison reliability estimates were lower for these variables in our sample of individuals with cannabis dependence (Psychological Health = 0.53; Physical Health = 0.42, Quality of Life = 0.42). It is difficult to tell if the source of the difference in reliability was due to: (i) an actual difference in the stability of the health and wellbeing constructs themselves – less stable in individuals with addiction to cannabis and more stable in patients seeking treatment for opioids; or (ii) the longer duration between raters' measurements in our study (up to 14 days) than in Ryan and colleagues' where there was a maximum 72-h interval between ratings [1].

There were several limitations to this study. First, the sample of patients used in this study were participating in a clinical trial in four specialist AoD treatment services, thus our results may not generalise to other AoD treatment settings, or to patients with less severe cannabis dependence. Second, there was a larger than ideal gap in time between the two ratings that were the basis of the interrater reliability analysis. The more time that elapses between raters' measurements, the greater the risk that the construct being measured actually changes, and is confounded with between-rater variability [55], which may explain why our reliability estimates were lower than previous validation studies, where there was a maximum 72-h gap between ratings [1]. Unfortunately in this case the procedure of this validation study was dictated by the procedure of the larger nabiximols randomised controlled trial. Future studies testing the reliability of these ATOP variables should aim for a smaller duration between ratings.

The idea of a clinically useful AoD outcomes measure has proven appealing to health services around the world. As well as the ATOP there are Chilean [56] and Chinese [57] versions of the TOP currently in use, with Hellenic and youth versions reportedly under development. Although there are minor differences across scales, the Substance Use, Psychological Health, Physical Health and Quality of Life items are extremely similar. Thus validating the ATOP for use in individuals with cannabis dependence simultaneously supports the use of the other versions of the TOP for monitoring outcomes in individuals seeking treatment for cannabis dependence.

Conclusion

For an instrument to be taken up and used widely by clinicians it must be useful. This means clinicians must believe it: (i) measures the construct it aims to measure; (ii) measures it reliably; (iii) is easy to use and score; (iv) can be completed quickly (i.e. 5–10 min); and (v) is applicable across a wide range of conditions, clients and treatment settings. The ATOP is already

widely used across Australian AoD treatment services, however until now it has not been validated for use with cannabis-dependent individuals. Our investigation shows that the ATOP is valid, reliable and suitable for use in this population. Use of the ATOP as a part of routine clinical care will enable monitoring of a wide range of patient reported outcomes and clinical risk factors, and will allow the drug and alcohol treatment sector to use its own clinical data to improve treatment effectiveness and patient safety.

Conflict of Interest

NL reported receiving grants from National Health and Medical Research Council of Australia during the conduct of the study; grants from Camurus, personal fees from Indivior and Mundipharma outside the submitted work; and being the Clinical Director of the Lambert Initiative in Cannabinoid Therapeutics at University of Sydney from 2015–2017, involved in a number of studies of medical cannabis, unrelated to this study. No other disclosures were reported.

References

[1] Ryan A, Holmes J, Hunt V *et al*. Validation and implementation of the Australian Treatment Outcomes Profile in specialist drug and alcohol settings. Drug Alcohol Rev 2014;33:33–42.

[2] Marsden J, Farrell M, Bradbury C *et al*. Development of the treatment outcomes profile. Addiction 2008;103:1450–60.

[3] Marsden J, Eastwood B, Jones H *et al*. Risk adjustment of heroin treatment outcomes for comparative performance assessment in England. Addiction 2012;107:2161–72.

[4] Marsden JD, Eastwood BM, Bradbury CB *et al*. Effectiveness of community treatments for heroin and crack cocaine addiction in England: a prospective, in-treatment cohort study. Lancet 2009;374:1262–70.

[5] Kroenke K, Spitzer RL, Williams JBW. The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. Psychosom Med 2002;64:258–66.

[6] Kessler RC, Andrews G, Colpe LJ *et al*. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol Med 2002;32:959–76.

[7] The WHOQOL Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. Psychol Med 1998;28: 551–8.

[8] Lintzeris N, Monds LA, Rivas G, Leung S, Withall A, Draper B. The Australian Treatment Outcomes Profile instrument as a clinical tool for older alcohol and other drug clients: a validation study. Drug Alcohol Rev 2016;35:673–7.

[9] Saunders JB, Aasland OG, Babor TF, De La Fuente JR, Grant M. Development of the alcohol use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption-II. Addiction 1993;88:791–804.

[10] Sheikh JI, Yesavage JA. Geriatric depression scale (GDS): recent evidence and development of a shorter version. Clin Gerontol 1986;5: 165–73.

[11] Ware JE Jr, Kosinski M, Keller SD. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996;34:220–33.

[12] Allsop DJ, Rooney K, Arnold JC *et al*. Randomised controlled trial (RCT) of daily aerobic exercise for inpatient cannabis withdrawal: a study protocol. Ment Health Phys Act 2017;13:57–67.

[13] Barker SF, Best D, Manning V, Savic M, Lubman DI, Rush B. A tiered model of substance use severity and life complexity: potential for application to needs-based planning. Subst Abus 2016;37:526–33.

[14] Bathish R, Best D, Savic M, Beckwith M, Mackenzie J, Lubman DI. 'Is it me or should my friends take the credit?' The role of social networks and social identity in recovery from addiction. J Appl Soc Psychol 2017; 47:35–46.

[15] Dore G, Sinclair B, Murray R. Treatment resistant and resistant to treatment? Evaluation of 40 alcohol dependent patients admitted for involuntary treatment. Alcohol Alcohol 2015;51:291–5.

[16] Mitchell PF, Kutin JJ, Daley K, Best D, Bruun AJ. Gender differences in psychosocial complexity for a cohort of adolescents attending youth-specific substance abuse services. Child Youth Serv Rev 2016;68:34–43.

[17] Savic M, Barker SF, Best D, Lubman DI. Alcohol problems among migrants in substance use treatment: the role of drinking patterns in countries of birth. Aust J Prim Health 2014;20:220–1.

[18] Slade T, Johnston A, Teesson M *et al*. The mental health of Australians 2. Report on the 2007, national survey of mental health and wellbeing. Contract No. 07. Canberra: Department of Health and Aging, 2009.

[19] Australian Institute of Health and Welfare. Alcohol and other drug treatment services in Australia 2017–18: key findings. Canberra, Australia: Australian Institute of Health and Welfare, 2019.

[20] Daeppen J-B, Krieg M-A, Burnand B, Yersin B. MOS-SF-36 in evaluating health-related quality of life in alcohol-dependent patients. Am J Drug Alcohol Abuse 1998;24:685–94.

[21] Chiu E-C, Hsueh IP, Hsieh C-H, Hsieh C-L. Tests of data quality, scaling assumptions, reliability, and construct validity of the SF-36 health survey in people who abuse heroin. J Formos Med Assoc 2014;113:234–41.

[22] Yan S, Lian Z, Sun G, Bao Y, Ge Y, Liu Z. Assessment of the Chinese-version SF-36 in the Chinese opiate addicts. Subst Use Misuse 2011;46: 1561–8.

[23] Bhardwaj AK, Allsop DJ, Copeland J *et al*. Randomised controlled trial (RCT) of cannabinoid replacement therapy (nabiximols) for the management of treatment-resistant cannabis dependent patients: a study protocol. BMC Psychiatry 2018;18:140.

[24] Lintzeris N, Bhardwaj A, Mills L *et al*. Nabiximols for the treatment of cannabis dependence: a randomized clinical trial. JAMA Intern Med 2019;179:1242.

[25] WHO. ICD-10: international statistical classification of diseases and related health problems. 10th rev., 2nd edn. Geneva: World Health Organization, 2005.

[26] Sobell LC, Sobell MB. Timeline follow-back. In: Litten RZ, Allen JP, eds. Measuring alcohol consumption. Totowa, NJ: Springer, 1992: 41–72.

[27] Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. Med Care 1992;30:473–83.

[28] Ware JE, Kosinski M. Interpreting SF&-36 summary health measures: a response. Qual Life Res 2001;10:405–13.

[29] McHorney CA, Ware JE Jr, Lu JR, Sherbourne CDJM. The MOS 36-item short-form health survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. Med Care 1994;32:40–66.

[30] McHorney CA, Ware JE, Raczek AE. The MOS 36-item short-form health survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care 1993;31:247–63.

[31] Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ 2002;21:271–92.

[32] Abdin E, Chong SA, Seow E *et al*. A comparison of the reliability and validity of SF-6D, EQ-5D and HUI3 utility measures in patients with schizophrenia and patients with depression in Singapore. Psychiatry Res 2019;274:400–8.

[33] Stavem K, Frøland S, Hellum K. Comparison of preference-based utilities of the 15D, EQ-5D and SF-6D in patients with HIV/AIDS. Qual Life Res 2005;14:971–80.

[34] Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the depression anxiety stress scales (DASS) with the Beck depression and anxiety inventories. Behav Res Ther 1995;33:335–43.

[35] Brown TA, Chorpita BF, Korotitsch W, Barlow DH. Psychometric properties of the depression anxiety stress scales (DASS) in clinical samples. Behav Res Ther 1997;35:79–89.

[36] Crawford JR, Henry JD. The depression anxiety stress scales (DASS): normative data and latent structure in a large non-clinical sample. Br J Clin Psychol 2003;42:111–31.

[37] Lovibond P. DASS website – frequently asked questions. Psychology Foundation of Australia, 2014. Available at: http://www2.psy.unsw.edu.au/dass/DASSFAQ.htm.
[38] Ng F, Trauer T, Dodd S, Callaly T, Campbell S, Berk M. The validity of the 21-item version of the depression anxiety stress scales as a routine clinical outcome measure. Acta Neuropsychiatr 2007;19:304–10.
[39] Sheehan D, Harnett-Sheehan K, Raj B. The measurement of disability. Int Clin Psychopharmacol 1996;11 (Suppl 3):89–95.
[40] Arbuckle R, Frye MA, Brecher M *et al*. The psychometric validation of the Sheehan disability scale (SDS) in patients with bipolar disorder. Psychiatry Res 2009;165:163–74.
[41] Coles T, Coon C, DeMuro C, McLeod L, Gnanasakthy A. Psychometric evaluation of the Sheehan disability scale in adult patients with attention-deficit/hyperactivity disorder. Neuropsychiatr Dis Treat 2014;10:887–95.
[42] Leon AC, Shear MK, Portera L, Klerman GL. Assessing impairment in patients with panic disorder: the Sheehan disability scale. Soc Psychiatry Psychiatr Epidemiol 1992;27:78–82.
[43] ATOP Project Team. Manual for ATOP version 4 (unpublished manual). Sydney: NSW Health, 2013.
[44] Cohen J. Statistical power analysis for the behavioral sciences. Burlington: Elsevier Science, 2013.
[45] Krippendorff K. Content analysis: an introduction to its methodology, 3rd edn. Thousand Oaks, CA: Sage, 2013.
[46] Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. Commun Methods Meas 2007;1:77–89.
[47] Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess 1994;6:284–90.
[48] Cicchetti DV, Sparrow SA. Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. Am J Ment Defic 1981;86:127–37.
[49] R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, 2016.
[50] Wickham H. Tidyverse: easily install and load 'tidyverse' packages. R package version. 2017.
[51] Gamer M, Lemon J, Fellows I, Puspendra S. irr: various coefficients of interrater reliability and agreement. R package version 0.84. Available at: http://CRAN.R-project.org/package=irr
[52] Marsden J, Eastwood B, Ali R *et al*. Development of the addiction dimensions for assessment and personalised treatment (ADAPT). Drug Alcohol Depend 2014;139:121–31.
[53] Pulford J, Deering DE, Robinson G *et al*. Development of a routine outcome monitoring instrument for use with clients in the New Zealand alcohol and other drug treatment sector: the alcohol and drug outcome measure (ADOM). NZ J Psychol 2010;39:35–45.
[54] Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther 2005;85:257–68.
[55] Mitchell SK. Interobserver agreement, reliability, and generalizability of data collected in observational studies. Psychol Bull 1979;86:376–90.
[56] Castillo-Carniglia Á, Marín JD, Soto-Brandt G *et al*. Adaptation and validation of the instrument treatment outcomes profile to the Chilean population. J Subst Abuse Treat 2015;56:39–47.
[57] Wang M, Shen J, Liu X *et al*. Reliability and validity of the treatment outcome profile among patients attending methadone maintenance treatment programs in Kunming, China. J Subst Abuse Treat 2017;77:89–94.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. Interrater reliability estimates for Australian Treatment Outcomes Profile continuous variables
Table S2. Interrater reliability estimates for ATOP dichotomous variables
Table S3. Sensitivity analysis: interrater reliability estimates for ATOP continuous variables under different restrictions of sample size based on lag between medical screen ATOP and researcher ATOP
Table S4. Sensitivity analysis: interrater reliability estimates for ATOP categorical variables under different restrictions of sample size based on lag between medical screen ATOP and researcher ATOP
Figure S1. Histogram of time between clinician and researcher Australian Treatment Outcomes Profiles (n = 117)