Australian Treatment Outcomes Profile (ATOP) Manual 2: Research and Service Evaluation

AoD Treatment Clinical Outcomes and Quality Indicators Program

c/o The Langton Centre, 591 South Dowling St, Surry Hills NSW 2010

T +61 2 9332 8777

E seslhd-coqi@health.nsw.gov.au

Further copies of this publication can be obtained from the COQI Program Manager at seslhd-coqi@health.nsw.gov.au

Suggested reference: Lintzeris, N., Deacon, RM., Mills, L., Bruno, R., Black, E., Holmes, J., Dunlop, A., Mammen, K. (2021). The Australian Treatment Outcomes Profile (ATOP) Manual 2: Research and Service Evaluation. Retrieved from URL: www.seslhd.health.nsw.gov.au/australian-treatment-outcomes-profile

ISBN: 978-0-6489170-1-4

Published: 8 November 2021

Contents

Acknowledgements

The Clinical Outcomes and Quality Indicator (COQI) Project team represents a team of project managers, clinicians, researchers and consumer workers working across South Eastern Sydney Local Health District (SESLHD) and University of Sydney, Centre for Population Health, New South Wales (NSW) Ministry of Health and Hunter New England Local Health District (HNELHD). The COQI Project team would like to thank the following agencies and people for their contributions to the development of the Australian Treatment Outcomes Profile (ATOP) and its materials.

The ATOP is based on the British TOP (Treatment Outcomes Profile), developed by the National Treatment Agency, National Health Service (NHS), United Kingdom. A large component of the original ATOP manual was adapted from the TOP support materials to better suit the Australian context. For more information on the British TOP please see: https://www.gov.uk/government/publications/drug-and-alcohol-treatment-outcomes-measuring-effectiveness/collecting-drug-and-alcohol-treatment-outcomes-information.

The COQI Project Team acknowledges the funding that has contributed to developing the ATOP and its implementation resources including the NSW Ministry of Health (initially through the Mental Health and Drug and Alcohol Branch, and more recently through the Centre for Alcohol and Other Drugs). This work would not have been possible without the contribution of the clients and staff (including consumer workers) from the treatment services involved in adapting the ATOP, in the validation and evaluation studies. Clinical and research staff and clients from the Drug and Alcohol services in SESLHD, HNELHD, South Western Sydney Local Health District (SWSLHD), and Western Sydney Local Health District (WSLHD) have contributed throughout the adaptation and validation process. They have been supported by staff and clients of Alcohol and Other Drug (AoD) services across NSW Health including Central Coast Local Health District (CCLHD), Illawarra Shoalhaven Local

Health District (ISLHD), Murrumbidgee Local Health District (MLHD), Mid-North Coast Local Health District (MNCLHD), Northern New South Wales Local Health District (NNSWLHD), Northern Sydney Local Health District (NSLHD), Sydney Local Health District (SLHD), Southern New South Wales Local Health District (SNSWLHD), and Western New South Wales Local Health District (WNSWLHD). The Agency for Clinical Innovation (ACI), Network of Alcohol and other Drugs Agencies (NADA), and New South Wales Users and AIDS Association (NUAA) have supported the work through their membership on the COQI Advisory Group and Clinical Consultation Groups. Finally, we wish to thank the participating researchers and academics from the University of Tasmania and the University of New South Wales who contributed to the development and validation of the ATOP.

Chapter 1: Overview - Using ATOP data for service improvement and research

1.1 Applications of the ATOP

As discussed in the Manual Part 1, The Australian Treatment Outcomes Profile (ATOP) is a brief, 22-item instrument that assesses various parameters of (a) substance use and (b) general health and wellbeing over the preceding 4 weeks. It is a patient reported outcome measure (PROM) and clinical risk screening tool, eliciting responses directly from clients and is designed to be incorporated into routine clinical care in Alcohol and other Drug (AoD) treatment settings.

The ATOP is usually administered either face-to-face or by telephone by a clinician or researcher and requires minimal training for administration or interpretation. It typically takes approximately 10 minutes to complete. In addition to the clinical applications of the ATOP (brief structured assessment, screening for risk, care planning, routine monitoring of treatment progress, and standardised communication between providers), the ATOP can also be used for service evaluation, continuous quality improvement, and clinical research.

The primary focus of this manual (Part 2) is on the use of ATOP data to develop a better understanding of the profile of clients accessing treatment and their clinical outcomes. It will provide ideas on how to display, analyse and report ATOP data in a variety of ways. The ultimate decision on how best to use the data will always depend on the questions being asked and the intended audience.

The following Chapters 2 and 3 provide practical advice on using ATOP data to summarise the characteristics of clients at a single point in time, and how to calculate and interpret changes in ATOP scores over time. These 2 chapters will be the focus for readers who want to get straight to cleaning and using their ATOP data.

Chapter 4 gives a deeper explanation of the processes by which cutoffs for ATOP single time point and change scores were established. Chapter 5 pulls together the ATOP cutoffs and minimum change scores into the COQI Outcome Metric for the ATOP.

Appendix A describes how ATOP data from AoD services across NSW was collected and used as a 'normative' dataset, which helped with formulation of the COQI Outcome Metric.

Appendix B gives an overview of other uses of the ATOP and its predecessor, the Treatment Outcomes Profile (TOP) in the research literature.

1.2 Why measure outcomes?

More than 10% of Australia's gross domestic product is spent on healthcare (Australian Institute of Health and Welfare 2020). The result of this spending on patients' health however, is largely unknown. While there may be measures on the number of procedures or interventions delivered, survival rates, safety indicators, and other measures, measurement of health outcomes is sub-optimal.

Health outcomes are disease-specific measures that show change in a medical condition. Each condition may have a range of related health outcomes that may vary according to the perspective of the stakeholder (e.g. the client, their family, the treatment provider, the health service administrator, the local community etc.).

Outcomes from the perspective of the patient are increasingly being referred to as PROs (Patient Reported Outcomes). A PRO can be defined as an outcome that is, "directly reported by the patient without

interpretation of the patient's response by a clinician or anyone else and pertains to the patient's health, quality of life or functional status associate with health care or treatment" (Food and Drug Administration 2019).

While measurement and benchmarking is occurring across many areas of healthcare, it often focuses on inputs (e.g. number of staff, qualifications, facilities), throughput (e.g. number of patients treated), and the completion of treatment processes (e.g. adherence to guidelines or standards for a procedure), rather than the effectiveness of those treatments and systems. Efforts to measure disease specific health variables have typically been limited to clinical trials of treatment efficacy or time limited evaluations of a treatment or program where a set number of clients are followed up at fixed time intervals. However, these approaches are time-consuming to conduct and the results of these evaluations cannot be assumed to transfer to all settings.

When health outcomes are routinely measured and feedback can be provided to staff and clients within a short timeframe, this information can be used to inform individual treatment decisions as well as to evaluate the treatment being provided by the service. This approach is consistent with NSW Health's Analytics Framework, which supports increased attention to health analytics for 'gaining insight for making informed decisions to improve health outcomes and health system performance' (NSW Ministry of Health 2016).

Chapter 1 key points:

- **Chapters 2-3 are practical 'how to' guides for selecting, cleaning and analysing your ATOP data**
- **Chapters 4-5 are a 'deep dive' into development of the COQI change metric**

Chapter 2: Using your ATOP data: describe characteristics of clients engaged with AoD Services

This chapter will cover the key considerations for selecting, cleaning, and displaying your ATOP data to provide descriptions of people in your treatment service.

The ATOP can be used to describe client populations with regard to their patterns of recent substance use, health status, quality of life, and areas of risk. When that is paired with client demographics and service descriptors, this also provides a valuable description of the profile of clients engaged in the service.

There are a number of ways to aggregate and present group ATOP data, depending on the aims of the review. We suggest two key 'cross sectional' approaches as a starting point:

(i) reporting data from clients at their entry into a treatment service, usually during a given period of time (e.g. the first ATOP collected with clients who entered a treatment program during a three, six or 12 month period);

(ii) reporting data collected from all clients in treatment during a defined period of time, regardless of when they began treatment (e.g. data from the most recent ATOP collected on all clients in a treatment program at three, six or 12 monthly intervals).

As these reports rely on clients regularly repeating ATOPs throughout treatment, the accuracy and therefore utility of these reports will be improved by ensuring regular ATOP completion among AoD teams.

Nuanced descriptions of client populations and treatment services can be achieved by combining ATOP data with the NSW Minimum Data Set for Drug and Alcohol Treatment Services (NSW MDS DATS) data collected at the start of a treatment episode. Data can be grouped according to a range of categories, such as by:

- client characteristics (e.g. age or gender)
- socio-economic variables (e.g. employment or housing status),
- client reported patterns of substance use (e.g. primary drug of concern at treatment entry, or clients reporting any use of a substance in the preceding 4 weeks),
- treatment type (e.g. counselling, withdrawal, opioid agonist treatment), and/or
- service location (e.g. clinic A versus clinic B).

2.1 Selecting your data

We recommend taking the following steps, and keep a record of these in case you need to review or repeat your actions at a later date:

1. Identify the group of clients you would like to review ATOP scores for (e.g. all clients entering OTP treatment; clients in treatment for alcohol problems).
2. Identify the time period you would like to review client profiles for (e.g. all clients who entered treatment from 1 July 2020 – 31 December 2020); all clients in treatment in the past 6 months).
3. Extract the relevant ATOPs.
4. If there are multiple ATOPs for a single person, decide which ATOP scores you will retain. This should be a single rule applied across all clients (e.g. the most recent ATOP; the first ATOP completed during the time period). Consider which rule makes the most sense for the questions you are attempting to answer.

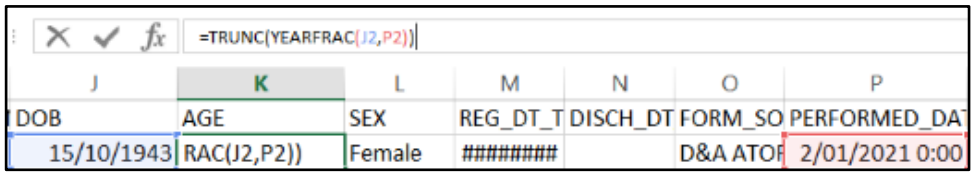
2.2 Cleaning your ATOP data

To have confidence in your analyses, you need to start by cleaning your data first. The steps listed below are intended as a guide to this process. We have also included ways to address common pitfalls in using ATOP data, and examples of which Excel functions to use.

The procedures can be conducted in Excel or other data analysis programs. An SPSS script, and a macro for use in Excel are available to apply steps 3, 4, 8-12, and 14-16 - please contact the COQI team to obtain these.

1. Save a copy of your raw ATOP data, and continue to make backup copies throughout processing!
2. Delete ATOPs that are not completed (TX_STAGE = clinically inappropriate or client refused).
3. Calculate age: number of year's difference between birth date and ATOP performed date, truncated.

In Excel: insert a column to the right of the column containing dates of birth and type "AGE" as a header into the first cell. Type the following into cell K2: "=TRUNC(YEARFRAC(J2,P2)":



Adjust cell references if your data are in different columns to this screenshot – similarly for the other Excel code examples following. Copy the formula to the bottom of column K and format as number, no decimal points. To prevent cell references being corrupted as data is further manipulated, insert another column to the right of AGE, copy the AGE column and paste as value into the blank column (right-click on blank column, select 'Paste special - value'). Then delete original AGE column. Follow this pattern for other Excel code examples given here.

4. Rename variables TOTAL_CANABIS to TOTAL_CANNABIS (removes typo), and OVERAL_QUALITY_OF_LIFE to QOL_Rating (removes typo, shortens variable name and converts to the same naming convention as PSYCH_RATING and PHYSICAL_RATING).

5. Optional: Quantity of alcohol used – If you need to use this variable, it is advised to clean data manually as ALCOHOL_TYPICAL_QTY is a string (free text) variable and ALCOHOL_UNITS can be edited by clinicians when they complete ATOPs. Use an Australian standard drinks chart (example) to assist in calculating descriptions such as "1 cask of wine", "2 long necks" into standard drinks. However, some descriptions such as "1 bottle" are too vague to permit recalculation, and ALCOHOL_TYPICAL_QTY should be left blank for these cases.

6. Optional: manually review OTHER_SUBSTANCE1 and OTHER_SUBSTANCE2 variables. Often clinicians enter substances here that should have been entered under one of the previous categories, most commonly benzodiazepines and other opioids. Where this is the case, copy the weekly and total days' used from OTHER_SUBSTANCE to the correct substance type. Where there is already non-zero use recorded, use the higher of the weekly days' used values and adjust total days accordingly.

BF	BG	BH	BI	BJ	BK	BL
)THER_OPIOIDS_TYPIC	OTHER_OPIOIE	OTHER_OPIOIDS_WEEK_4	OTHER_OPIOIDS_WEEI	OTHER_OI	OTHER_O	TOTAL_OTHER_OPIOIDS
1	tab	1	0	1	0	2

For example, the ATOP entry below records 2 days of use of other opioids.

BT	BU	BV	BW	BX	BY	BZ	CA
OTHER_SUBSTANCE1_NA	OTHER_SU	OTHER_SU	OTHER_SU	OTHER_SU	OTHER_SU	OTHER_SU	TOTAL_OTHER_SUBSTANCE_1
Panadeine Forte	2	daily	6	6	6	6	24

But recorded in OTHER_SUBSTANCE_1 is additional opioid use:

As Panadeine Forte is an opioid, and not used to treat opioid use disorder, this should have been entered under

OTHER_OPIOIDS. Using the recoding rule above, 6 days use in a week is larger than 1 day use, so replace OTHER_OPIOIDS

BF	BG	BH	BI	BJ	BK	BL
OTHER_OPIOIDS_TYPIC	OTHER_OPIOIE	OTHER_OPIOIDS_WEEK_4	OTHER_OPIOIDS_WEEI	OTHER_OI	OTHER_OI	TOTAL_OTHER_OPIOIDS
2	tabs panadeine	6	6	6	6	24

variables with:

7. Further advice on use of the OTHER_SUBSTANCE variables: unless use of other substance types (e.g. GHB, synthetic cannabis) is of clinical interest, these variables can be discarded after the optional manual review step 6.

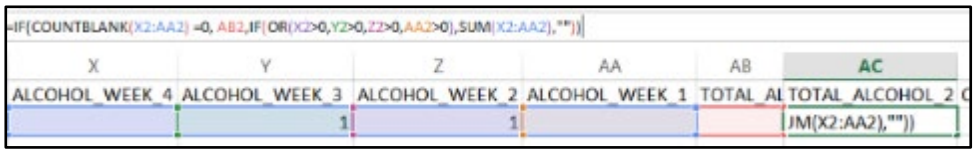
8. Correction for missing data in the weekly substance use and injecting variables. Sometimes not all weeks are filled in, creating a situation such as for weeks 3 and 2 there is 1 day each of alcohol use recorded (non-zero use), but weeks 4 and 1 are blank, and the TOTAL_ALCOHOL variable is missing, as in the example below.

W	X	Y	Z	AA	AB
ALCOHOL	ALCOHOL_WEEK_4	ALCOHOL_WEEK_3	ALCOHOL_WEEK_2	ALCOHOL_WEEK_1	TOTAL_AL
Std drinks		1	1		

In order to minimise missing data, the days of use for weeks 4 and 3 can be summed and copied to the TOTAL_ALCOHOL variable (in this example, TOTAL_ALCOHOL would equal 2), and an indicator variable can still be created indicating there was some alcohol use. The assumption here is that clinicians (where they make this error) fill in weeks where there is substance use, and skip weeks where there is none. However, this rule should not be applied where zero use is indicated for one or more weeks but the other weeks are blank – in this case it is less certain that there was truly no use of that substance in the previous 28 days.

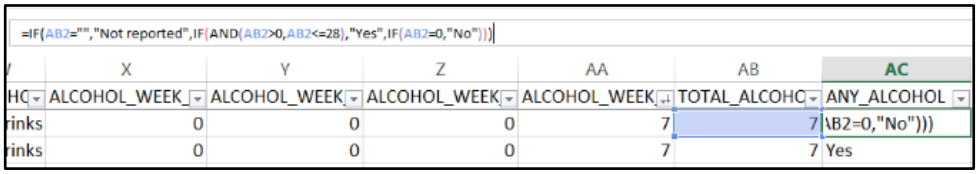
In Excel: In cell AC2, enter "=IF(COUNTBLANK (X2:AA2) = 0, AB2, IF(OR(X2>0, Y2>0, Z2>0, AA2>0), SUM(X2:AA2), ""))". Cell AC2 then calculates to 2. Copy and paste column AC as value only to a new column as before, delete TOTAL_ALCOHOL and original

TOTAL_ALCOHOL_2 columns, and rename TOTAL_ALCOHOL_2 to TOTAL_ALCOHOL. Repeat for other substance use categories.



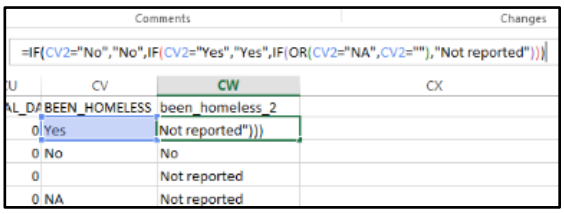
9. Create indicator variables (variables indicating any use/no use in the last 28 days) for the substance use categories, injecting, work and study. Calculate these from total days used for each substance/injected/worked/studied, and include a level for missing data. E.g. TOTAL_ALCOHOL can be recoded into another variable called ANY_ALCOHOL where 0=no use (TOTAL_ALCOHOL=0), 1=any use (TOTAL_ALCOHOL is between 1 and 28), 2=missing/not completed (TOTAL_ALCOHOL is blank).

   In Excel: "=IF(AB2="", "Not reported", IF(AND(AB2>0, AB2<=28), "Yes", IF(AB2=0,"No")))"



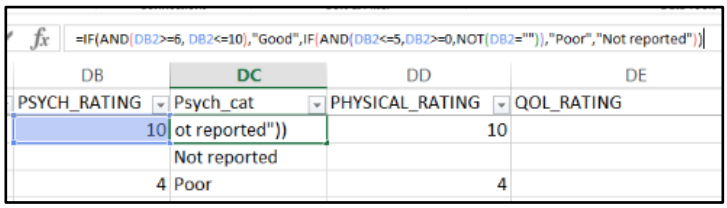
10. Recode the categorical variables BEEN_HOMELESS, AT_RISK_OF_EVICTION, CHILDREN_UNDER_5, CHILDREN_5_TO_15, BEEN_ARRESTED, BEEN_VIOLENT, and SUFFERED_VIOLENCE, respectively, so that there is a level combining missing data and where the response was "NA" (not answered). E.g. 0=No, 1=Yes, 2=Not reported.

    In Excel: =IF(CV2="No", "No",IF(CV2="Yes", "Yes", IF(OR(CV2="NA", CV2=""), "Not reported")))



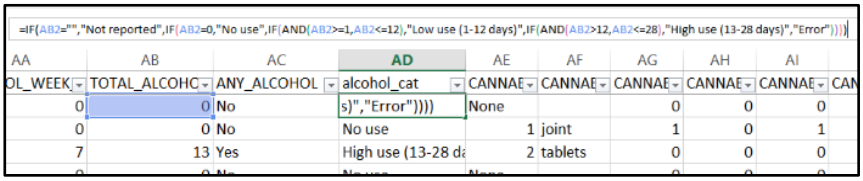
11. Optional: Consider creating "all opioids" variables – total days use of any form of opioids (summing the heroin and other opioids use variables), set to a maximum of 28 days, and an indicator variable for any heroin and/or other opioids use. This reasoning is based on Time-Line Followback data from the COQI study (Deacon, Bruno et al. 2018) which showed that the overwhelming majority of people who used both heroin and other opioids had done so on separate days, not on the same days. This step can simplify output by reducing the number of substance categories to report, but there may also be clinical interest in reporting heroin and other opioid use separately.
12. Optional: Similarly, depending on the clinical focus, "any stimulants" variables can be created, being the sum of total days of ATS and cocaine use and an indicator variable for any ATS or cocaine use. Again the assumption here is that clients of AoD treatment services will tend to use one or the other of ATS or cocaine on a day of use, not on the same day.
13. Principal drug of concern (PDOC): this is available for Assessment ATOPs only. Where this is to be used, it can be helpful to reclassify PDOC into a variable with fewer levels – e.g. by combining different opioid types into one 'opioids' level, combining different amphetamine-type stimulants into one 'ATS' level. This can also be done with OTHER_SPECIFY (where sufficient responses have been provided for this to be worth your while.
14. Recode the three variables PSYCH_RATING, PHYSICAL_RATING and QOL_RATING (the "PPQ" variables) into categorical variables based on the clinical cutoffs established by Mammen, Mills et al. (2021) and described in ATOP Manual part 1 (Chapter 2 section 2). A score of 5 or less indicates a client is likely to have a 'clinical problem' in that area. Thus, recode PPQ scores of 0-5 as 1=poor, scores of 6-10 as 2=good, and any missing scores into 3=missing.

In Excel: "=IF(AND(DB2>=6, DB2<=10),"Good", IF(AND(DB2<=5, DB2>=0, NOT(DB2="")), ""Poor","Not reported"))"



15. Optional: Recode continuous days of substance use variables into categorical variables with three levels where 1=no use, 2=low use (1-12 days) and 3=high use (13-28 days)[1].

    In Excel: "=IF(AB2="","Not reported", IF(AB2=0, "No use", IF(AND(AB2>=1, AB2<=12), "Low use (1-12 days)", IF(AND(AB2>12, AB2<=28), "High use (13-28 days)", "Error"))))"



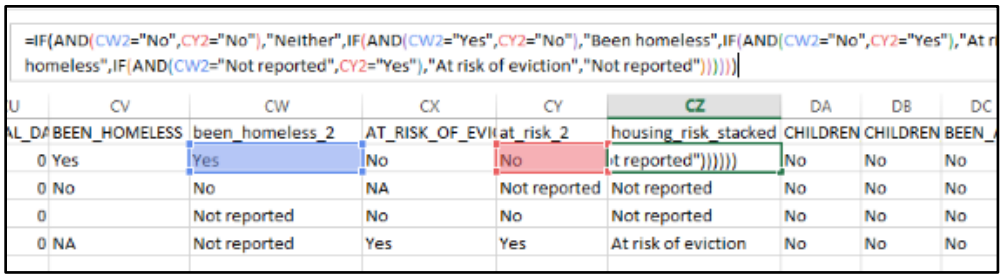
    Alternatively, eight-level variables - no use, 1-4 days, 5-8 days, 9-12 days, 13-16 days, 17-20 days, 21-24 days, and 25-28 days.

16. Optional: Compute stacked variables, combining the variable pairs work and study, homeless and at risk of eviction, living with children under 5 and living for children aged 5-15, experienced violence and inflicted violence. For example, work and study can be combined into one 5-level variable where 1=worked only, 2=studied only, 3= worked and studied, 4=neither, and 5=not answered/missing. Similarly for children: 1=living with children under 5 years only, 2=living with children 5-15 years only, 3=living with any children, 4=no children and 5=missing/not answered. Where one variable in each pair is reported as yes but the other is missing (e.g. someone worked, but did not answer the study question) we recommend this is coded as valid – in this example, that someone worked. Where

[1] These cutoffs for 'low' and 'high' substance use frequencies were discussed and determined by consensus in the AoD clinicians expert consultation group, see section 4.4 for further detail.

someone answered no to one variable and the other is missing, code the stacked variable as missing. The SPSS script and macro each follow these coding rules. Stacked variables can make reporting both simpler, as there are fewer variables, and more informative, e.g. it is easy to see the overall proportion of clients with violence around them.

In Excel: =IF(AND(CW2="No", CY2="No"), "Neither", IF(AND(CW2 ="Yes", CY2="No"),"Been homeless", IF(AND(CW2="No", CY2="Yes"), "At risk of eviction", IF(AND(CW2="Yes", CY2="Yes"), "Both", IF(AND(CW2="Yes", CY2="Not reported"), "Been homeless", IF(AND(CW2="Not reported", CY2="Yes"), "At risk of eviction", "Not reported"))))))



17. Output: Excel can be used to create summary PivotTables and charts in various formats, such as in Figures 1 and 2 below, or other available programs can also be used.

2.3 Displaying your ATOP data.

Table 1 lists the ATOP items and describes a number of ways to display the data.

Figures 1 and 2 give an example of how ATOP variables can be displayed graphically. Here, PDOC has been recoded into fewer categories by combining heroin and other opioids, and amphetamines and cocaine. Substance use frequency is presented in no use/low use/high use/not reported categories. Figure 1 shows demographics and drug use data for new clients of a hypothetical AoD team attending over a 6 month period. Figure 2 shows their ATOP social situation, risk factors and wellbeing variables.

Table 1: ATOP items and summary of ways to analyse and graph the items.

Item	**Display and analyses**
Substance use in the previous 28 days	1. Report % reporting any use and mean (SD) or median (IQR) for those reporting any use 2. Histogram grouped by: no use, 1-12 days in the past 28, 13-28 days in the past 28 *(See Figure 1 for an example).* 3. Histogram grouped by average number of days per week of use: no use, 1-4 days in the past 28, 5-8 days in the past 28, 9-12 days in the past 28, 13-16 days in the past 28, 17-20 days in the past 28, 21-24 days in the past 28, 25-28 days in the past 28
Injecting drug use and shared equipment in past 28 days.	1. Report % reporting any use (*see Figure 1*), and mean (SD) or median (IQR) for those reporting any injecting drug use. 2. Histogram grouped by no use, 1-12 days in the past 28, 13-28 days in the past 28 Histogram grouped by average number of days per week of use: no use, 1-4 days in the past 28, 5-8 days in the past 28, 9-12 days in the past 28, 13-16 days in the past 28, 17-20 days in the past 28, 21-24 days in the past 28, 25-28 days in the past 28).
Shared injecting equipment	1. % shared equipment for those who injected (*see Figure 1*)
Working or studying days (combined)	1. % reporting any work / study; and mean (SD) or median (IQR) of days for those reporting any.

Item	Display and analyses
	2. Histogram grouped by average number of days per week of work or study: none, 1-4 days in the past 28, 5-8 days in the past 28, 9-12 days in the past 28, 13-16 days in the past 28, 17-20 days in the past 28, 21-24 days in the past 28, 25-28 days in the past 28) 3. Cumulative bar graph: Worked only, studied only, both, neither, missing/not answered (*Figure 2*).
Living with children 0-5	1. Single bar graph with proportion reporting yes recorded 2. Cumulative bar graph with the following combined variables: Housing instability (combining homeless only, at risk of eviction only, both homeless and at risk of eviction, neither homeless nor at risk of eviction, missing/not answered); Violence (experienced violence, been violent, both experienced and been violent, and neither been or experienced violence, missing/not answered); Living with children (Under 5 yrs only, Under 5 years and 5-15 years, 5-15 years only, not living with children, missing/not reported) *(see Figure 2)*
Living with children 6-15	
Arrested	
Homeless	
At risk for eviction	
Experienced violence	
Been violent	
Self-rated psychological health	1. Mean, SD 2. % no clinical concern; % clinical concern, % not reported/missing (*see Figure 2*).
Self-rated physical health	
Self-rated quality of life	

Figure 1 Example presentation of demographics and ATOP substance use variables for new clients at an AoD clinic.

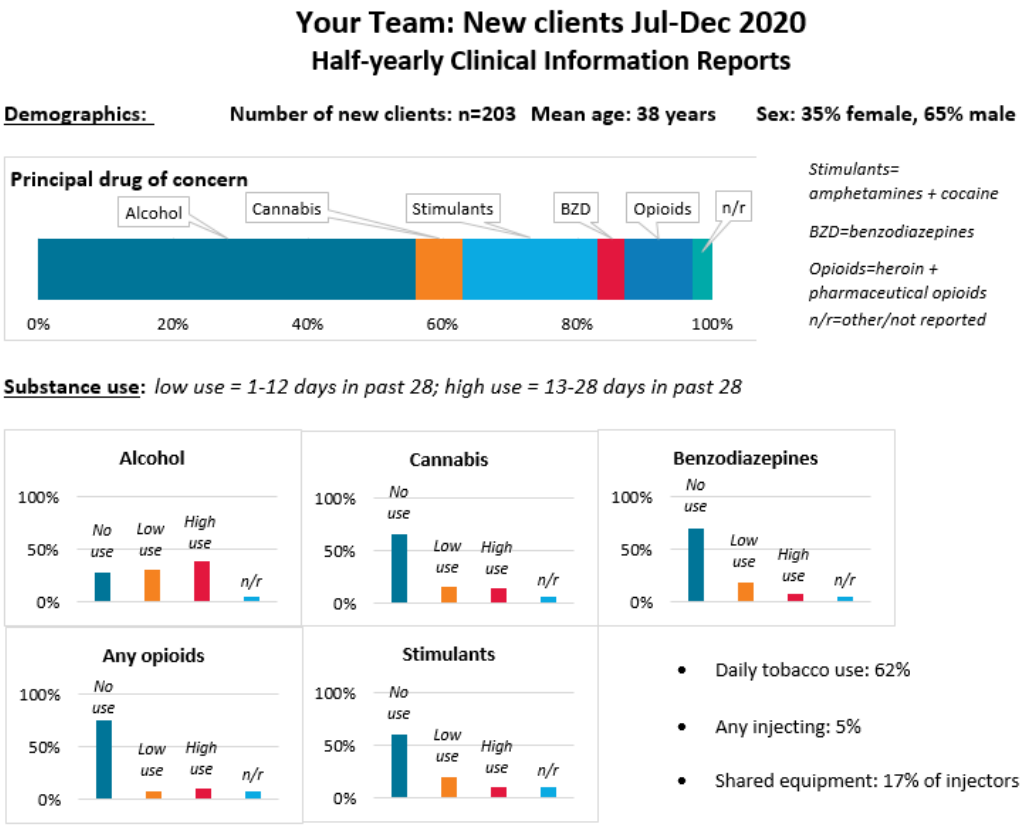
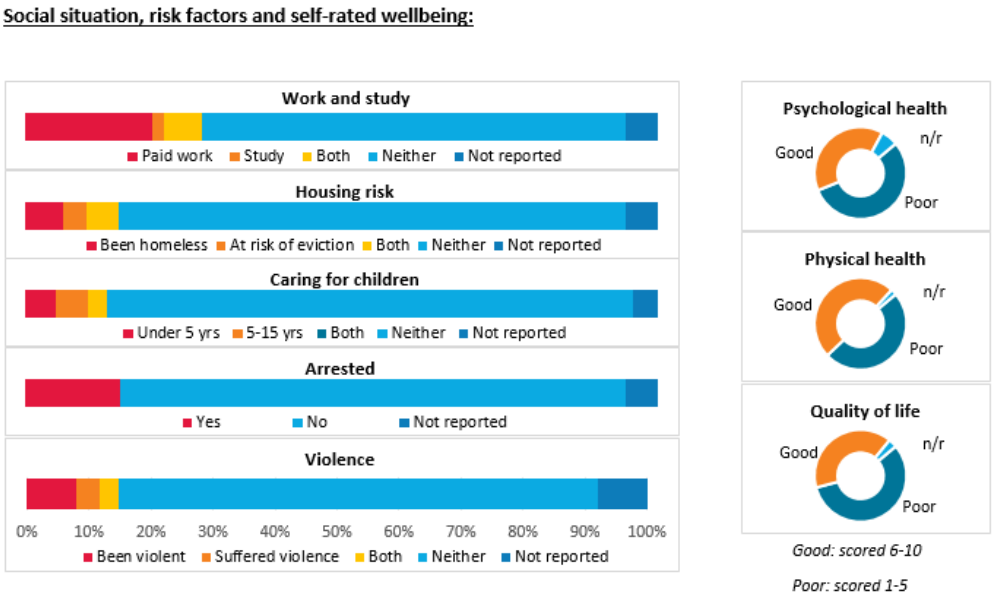


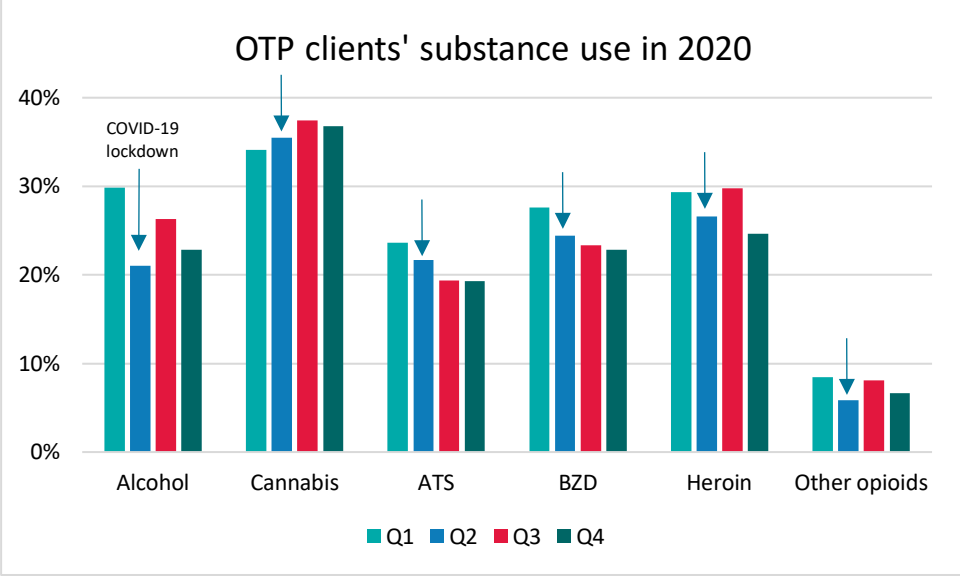
Figure 2 Example presentation of ATOP social situation, risk factors and wellbeing variables for new clients at an AoD clinic.



2.4 Comparing 'cross sectional' scores over time

Services may want to review their ATOP scores over time. This can be a way of looking at a number of things such as changes in drug use trends and in social-economic circumstances for the people attending your service. This serial display of ATOP scores is different to using ATOPs for outcome measurement (examined in Chapter 3) as we are not looking at changes in ATOP scores for individuals over time, but rather the group trends. See figure 3 for an example.

Figure 3 Quarterly cross-sectional substance use for all OTP clients attending a Sydney AoD clinic in 2020. Data from approx 400-450 unique clients was available per quarter. Arrows indicate the quarter where COVID-19 restrictions were first imposed. *ATS: amphetamine type substances; BZD: benzodiazepines.*



COVID-19 hit Australia in early 2020, with stay-at-home orders imposed from late March which in NSW extended through to late May (Storen and Corrigan 2020). Figure 3 suggests OTP clients' substance use may have changed in response – possible decreases in alcohol, amphetamine-type substances, benzodiazepines, heroin and other opioids, and a possible increase in cannabis use.

Chapter 2 key points:

- **Steps for selecting, cleaning and analysing your ATOP data are described**
- **We present example graphical representation of single time-point and repeated cross-sectional ATOP data**

Chapter 3. Using your ATOP data: describe change among clients engaged with AoD services

3.1 What question and which data?

The use of information from ATOPs collected over time enables services to examine client outcomes across a number of relevant domains. More than one ATOP needs to be completed by clients for it to be used as an outcome measure and the timing of those ATOPs depends on the evaluation question the service wants to ask. Some common questions might be:

a. What are the 6-month (or 12-month) outcomes of clients entering our service in, e.g. 2019?

   To answer this question you need to be able to:

   i. identify the clients who started treatment in that time period
   ii. identify their start of treatment ATOP data
   iii. identify their 6 month (or 12 month) ATOP data
      a. decide what time range the 6- or 12-month ATOP can fall in, e.g. follow up ATOP completed between 4 and 8 months after start of treatment ATOP; completed between 10 and 14 months after start of treatment ATOP. When using routinely collected data, measures are not always taken at precise times but may be several weeks before or after, depending on individual client attendance at the services. This is in contrast to clinical trials where

follow-ups are standardized to a protocol (e.g. every 3 months +/- 1 week).

iv. Match the start of treatment and 6 (or 12) month treatment data

v. Analyse their change scores (see sections X and Y for more details)

b. What were the outcomes of all the clients who were discharged from our service in the past 12 months?

To answer this question you need to be able to:

i. Identify the clients who were discharged from the service in that time period. This would usually be through MDS.

ii. Identify their start of treatment ATOP data (this may date back more than a year).

iii. Identify their end of treatment ATOP data. In order to reduce missing data, this may need to be the last ATOP collected, even though it may not be within 1-2 months of client discharge. Consider whether this is an appropriate option for your service.

iv. Match the start of treatment and end of treatment data.

v. Analyse their change scores (see section 3.2 for more details).

c. What were the outcomes of clients who were in our service in X (e.g. 2019) year?

This question looks at change regardless of entry or exit ("all in").

To answer this question you need to be able to:

i. Identify all clients of the service with more than one ATOP completed in 2019.

ii. For these clients, select the earliest and latest ATOPs completed within 2019.

iii. Ensure the 2 ATOPs for each client were completed at least 28 days apart, or a longer period of time if you want longer term outcomes.

iv. Match the start of treatment and end of treatment data
v. Analyse their change scores (see section 3.2 for more details)

Beware of the limitations of reporting on data that represents only a small proportion of your clients. In general, it is considered difficult to 'generalise' findings where there is data available for fewer than 70% of eligible clients, and indeed generalizability increases when reporting on >80% of eligible clients. To minimize missing data and enhance generalizability, clinical business process should be implemented whereby routine clinical reviews are conducted using the ATOP at a minimum of every three months. For many services this will provide a good source of outcomes data to answer a variety of questions.

First clean and recode ATOP data according to the data cleaning process in section 2.2.

3.2 How to analyse

3.2.1. Mean change scores

Here, we demonstrate how ATOP change scores can be represented using inferential statistics. As an example, we present data collected during the COQI validation study (Deacon, Bruno et al. 2018). Presented here are baseline and follow-up ATOP scores for 108 participants whose PDOC was alcohol. Participants were a mix of new and ongoing clients, and follow-up ATOPs were completed 3 months after baseline. Table 1 displays data for the ATOP alcohol use variables, and the three PPQ variables (psychological health, physical health and quality of life).

Table 1 Changes in alcohol use, psychological health, physical health and quality of life for 108 participants with alcohol as their principal drug of concern, enrolled in the COQI validation study. Participants had an ATOP completed at baseline and 3 months' followup.

ATOP variable	Baseline	Follow up	Change statistic
Alcohol			
Any use, n (%)	75 (69%)	61 (56%)	χ^2=6.036, p=0.014[1]
Days use, mean (SD)	10 (10)	7 (10)	t=3.257(107), p=0.002[2]
Days use, median (IQR)	10 (0-20)	2 (0-11.75)	Z=-3.053, p=0.002[3]
Psychological health	5.6 (2.1)	5.9 (1.9)	t=-1.571 (107), p=0.119[2]
Physical health	6.0 (2.1)	6.0 (2.0)	t=-0.102(107), p=0.919[2]
Quality of life	5.8 (2.3)	6.5 (1.9)	t=-2.476(107), p=0.015[2]

[1] McNemar test; [2] paired t-test; [3] Wilcoxon signed-rank test

The table shows that there was a statistically significant ($p<0.05$) decrease in alcohol use – both in the number of people using, and in days use. However, this approach says nothing about how many people decreased or increased their use, only that overall there was a reduction across the group of clients. In practice, we often want to be able to describe the proportion of clients who increased or decreased their use, rather than only reporting group changes. There was also a significant increase in quality of life ratings. However, the increase was from 5.8 to 6.5, on a 0-10 scale – is under 1 point. It is crucial to consider 'does this statistically significant increase have any clinical meaning'? When using large data sets, even quite minor changes in means may be statistically significant, but may not represent a clinically meaningful change.

3.2.2 Clinically meaningful change

There are a number of reasons why it is important to report 'clinically meaningful' rather than statistically significant changes.

- While the traditional method for analysing change in clinical outcomes for continuous variables include analyses such as paired t-tests, ANOVAs etc., these approaches can lead to statistically significant results that may not reflect a clinically meaningful change.

That is, the statistically significant change may represent a magnitude of change that is not considered clinically important or meaningful to clients and service providers. Statistical significance is also influenced by the number of people in the dataset – in a large enough dataset, even very small changes may reach statistical significance without being clinically meaningful.

- In some cases, small changes may reflect 'measurement error' in the use of a scale. For example, a client may report using cannabis on 15 in the past 28 days at one ATOP, but may report using 16 days at the next ATOP. This may reflect a genuine increase by 1 day, but may also reflect recall or reporting errors (that is frequency of cannabis use stayed the same, but the client reported a difference).
- Group changes also do not reflect that some clients have improvements, whilst others have deteriorations in their outcomes, which may not be reflected by group changes. For example, one client may increase their alcohol use by 10 days, another reduces their alcohol use by 4 days. The mean change suggests a group increase of 3 days each, but this does not reflect the true clinical picture. This section describes how to calculate the proportion of clients who have improved or deteriorated by a clinically meaningful amount.
- It also easier for services to use these methods to report the proportion of clients who demonstrated clinically meaningful change, without the need for complicated statistical analyses requiring assistance of a statistician.

Chapter 4 discusses in depth our approach for determining clinically significant change thresholds for the ATOP substance use and PPQ variables. The approach identified change thresholds to determine whether a client has significantly improved, significantly deteriorated, or stayed the same. There is one threshold for substance use (days' used variables) and a second for the PPQ items, which are detailed in Table 2. These thresholds incorporate the change required to overcome reliability/reporting errors, as well as clinical significance.

3.2.3 How can a binary 'outcomes metric' assist?

The COQI metric extends the idea of clinically meaningful change and transforms it into a binary metric. The next step, after determining the amount of change needed for a 'successful treatment outcome' in AoD treatment, is to consider if a client always needs to 'improve' their ATOP scores in order to be considered a good clinical outcome. There are a number of scenarios where 'staying the same' is a successful outcome in its own right. For example, a client already engaged in treatment may not be using any substances of concern in the past 28 days, and report high ratings for physical and psychological health (e.g. 7/10 and 8/10 respectively); or where a client transfer from residential rehab treatment (with no substance use in preceding month) to a community setting for ongoing relapse prevention counselling. A successful outcome may be to maintain their abstinence and high health ratings (indeed it is not possible to reduce your substance use below 0 days, or to improve your health rating score further). In contrast, no change in substance use over time would not be generally considered a good outcome for a client who entered treatment using heroin 28/28 days. In both examples, the client made no change in their substance use over time – but the clinical meaning associated with 'no change' is quite different. Hence, assigning meaning to change scores requires consideration of the client's 'starting point'.

To assist this process of assigning clinical meaning to change scores, we identified 'cut off' points on the ATOP items that would indicate whether a client whose score 'stays the same' should be defined as a successful or unsuccessful outcome: for substance use the cutoffs represent low and high substance use frequency as per Section 4.4, and for the psychological and physical health, and quality of life items the cutoffs indicate whether the client is likely to be experiencing a problem in those areas as per Section 4.4 The details of the steps undertaken to determine the algorithm are outlined in Chapter 4.

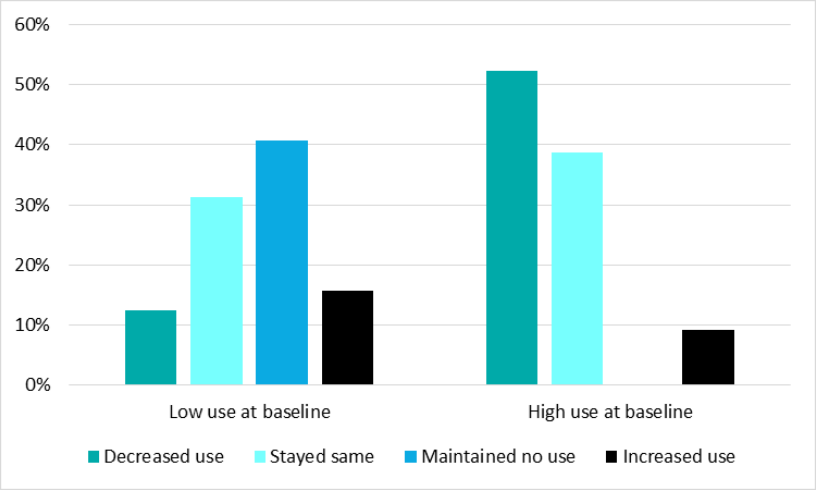
Table 2 outlines the algorithm for the COQI metric for substance use variables and PPQ variables.

Table 2: COQI algorithm for clinically meaningful change in substance use, and psychological health, physical health and quality of life (the PPQ variables). Treatment outcome is deemed successful or unsuccessful depending on both the starting point, and the threshold of change required to be clinically meaningful.

Frequency of substance use at measurement A	**Relative change in frequency of substance use in previous 28 days at measurement B**	**Change category**	**Treatment outcome**
Low (≤12 days in previous 28)	Increased by ≥4 days use compared to measurement A	Significant increase	**Unsuccessful**
	Reduced by ≥4 days use compared to measurement A	Significant decrease	**Successful**
	Increase or decrease of <4 days use compared to measurement A	No significant change	**Successful**
	No change from zero use	Maintained no use	**Successful**
High (>12 days in previous 28)	≥30% increase in days use compared to measurement A	Significant increase	**Unsuccessful**
	≥30% decrease in days use compared to measurement A	Significant decrease	**Successful**
	<30% increase or decrease in days use compared to measurement A	No significant change	**Unsuccessful**
Rating of PPQ variable at measurement A	**Relative change in rating of PPQ variable at measurement B**	**Change category**	**Treatment outcome**
Clinical concern (score of ≤5 on 0-10 scale)	Score increase of ≥2 from score at measurement A	Significant increase	**Successful**
	Score decrease of ≥2 from score at measurement A	Significant decrease	**Unsuccessful**
	Increase or decrease of <2 from score at measurement A	No significant change	**Unsuccessful**
No clinical concern (score of > 5 on 0-10 scale)	Score increase of ≥2 from score at measurement A	Significant increase	**Successful**
	Score decrease of ≥2 from score at measurement A	Significant decrease	**Unsuccessful**
	Increase or decrease of <2 from score at measurement A	No significant change	**Successful**

So let's look at how to analyse and display data according to this metric. Taking the same dataset as used in section 3.2.1 (data from participants with PDOC alcohol from the COQI validation study), lets first look at alcohol days' use outcomes, split by the threshold for low and high substance use as per Table 2, and presented in Figure 1.

Figure 1 Proportion of clients with decreased use, same use, maintained no use and increased use of alcohol from baseline to 3 month's follow up (alcohol principal drug of concern), split according to whether use at baseline was low or high frequency.



Now we can see, for example, that among clients with low use at baseline, 40% have maintained no use of alcohol at all. Among clients with high use at baseline, over 50% decreased their use by a clinically meaningful amount.

Figure 2a and b: Data as per figure 1 but a) outcomes categories colour-coded to represent successful and non-successful outcomes; b) successful and non-successful outcome categories grouped together by low and high use.

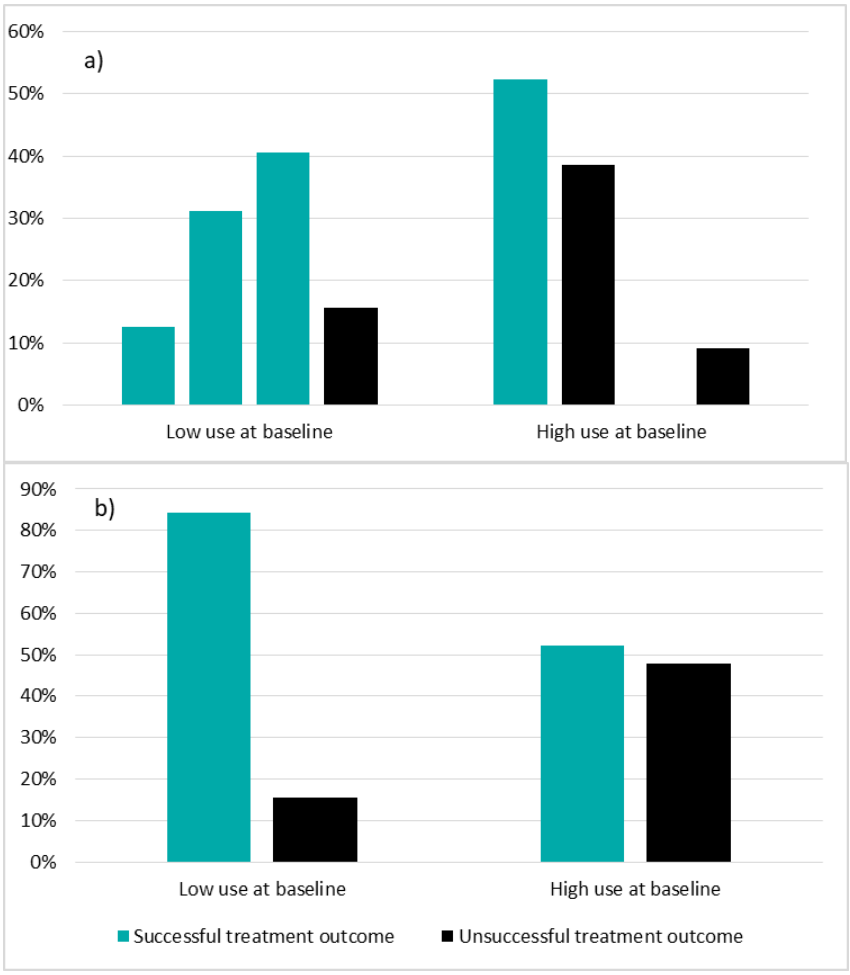
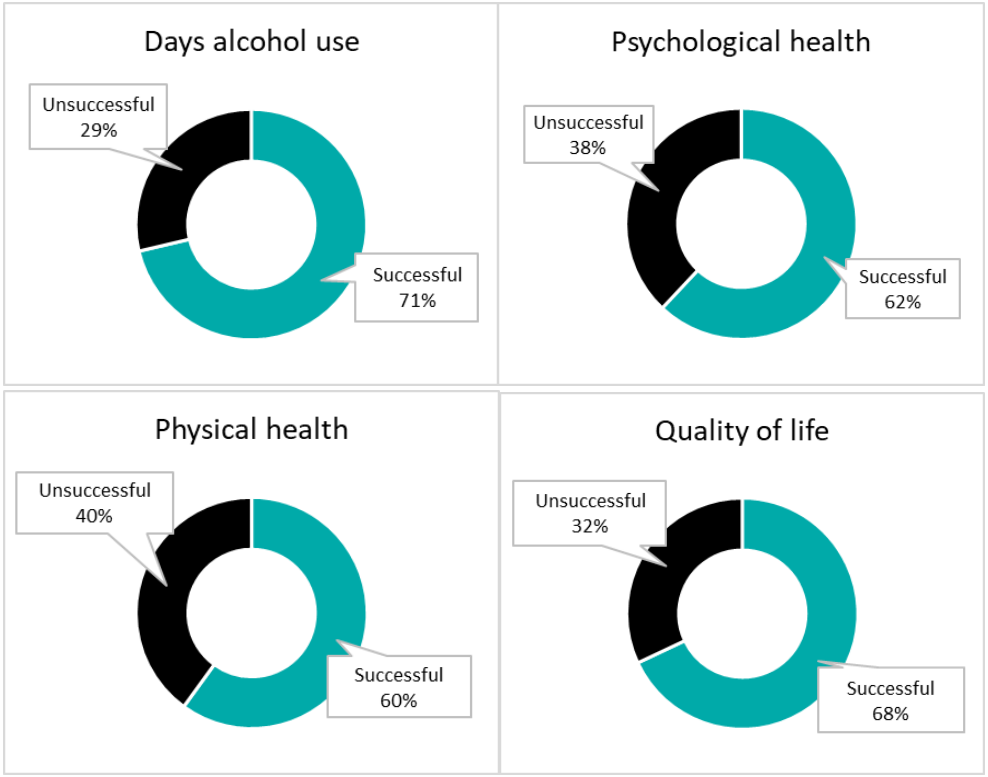


Figure 2a is the same as figure 1, with outcomes now colour-coded to show which are considered successful and non-successful treatment outcomes according to table 3. Figure 2b groups successful and non-successful groups together. We can now see that the majority of participants with low use at baseline had a good outcome.

Finally, combining low and high baseline use groups, and applying the same steps to the PPQ variables, outcomes can presented as per Figure 3. At least 60% of participants had successful outcomes on one or more of these ATOP variables.

Figure 3 Successful and unsuccessful outcomes for days alcohol use, psychological health, physical health and quality of life



Chapter 3 key points:

- **The algorithm for the COQI change metric for substance use variables and psychological health, physical health, and quality of life are given in Table 2.**
- **We present example graphical representation of ATOP change data**

Chapter 4. In depth: Clinically Meaningful Change

4.1 Introduction

As noted in Chapter 3, the traditional method for analyzing change in clinical outcomes – standard inferential statistics such as t-tests and ANOVAs – can lead to statistically significant results that don't represent a clinically meaningful change. That is, the statistically significant change is less than a magnitude of change that is considered meaningful to clients and service providers. They also don't give a clear picture of which clients or how many are showing improvements or deterioration in their outcomes.

The concept of a clinically meaningful change is proposed as an alternative to statistically significant change for treatment outcomes research (Jacobson and Truax 1991, Crosby, Kolotkin et al. 2003, Marsden, Eastwood et al. 2011). Clinically meaningful change has been operationalized in a number of different ways and is often discussed alongside 'reliable change' (RC), an idea first proposed by Jacobson, Follette et al. (1984) and then refined by Jacobson and Truax (1991). Jacobson and Truax (1991; p14) propose RC as a measure of how likely it is that the observed change in score on a question is due to actual change for a client rather than merely "the fluctuations of an imprecise measuring instrument", and this method has received much support (e.g. Speer 1992, Cisler, Kowalchuk et al. 2005, Marsden, Eastwood et al. 2011). Jacobson and Truax extend the RC concept into Clinically Significant change, whereby an individual's change is denoted as clinically significant if it moves

them from a dysfunctional population to a functional population for the construct of interest, sometimes termed 'recovery'.

We are not seeking here a measure to define AoD clients as 'recovered' or as falling within a general or functional population. 'Recovery' has meaning beyond simply days of substance use. For example, a person can have no substance use at all and still not deem themselves 'recovered'. Furthermore, client perceptions of 'recovery' or attainment of treatment goals can vary over time as client circumstances and goals change – a common feature in AoD treatment (e.g. a client may initially identify reduced substance use as their goal, but once this is achieved, their goal may shift to include other things such as finding employment). Rather, we are seeking a threshold of change for the ATOP PPQ and substance use variables, either an increase or decrease, to warrant a clinician noting this as meaningful change and aid interpretation of change in aggregated data.

Given this, we used a two-pronged approach to determining 'clinically meaningful change' thresholds for ATOP items. For the PPQ variables, we calculated the Reliable Change Index (RCI) to determine the minimum amount of change that could be considered statistically reliable, using data from the COQI ATOP validation study (Section 4.2; Deacon, Bruno et al. 2018). We also looked at additional data driven methods for determining clinically meaningful change for the PPQs as a sensitivity analysis.

It is challenging to directly apply the RCI method to substance use frequency. Whereas the PPQ variables are normally distributed, in our sample the days of substance use variables are not. Thus, for the days of substance use variables, we calculated the RCI in an adapted way (Section 4.3), using a large, representative sample of ATOPs collected at treatment entry into NSW AoD treatment services.

In Section 4.4 we explored the data-driven findings from Sections 4.2 and 4.3 chapter with a key stakeholder group to come to a

consensus on what constitutes clinically meaningful change for ATOP data from NSW AoD treatment services.

Going further, in Chapter 5 we established the COQI metric which extends the idea of clinically meaningful change and transforms it into a binary metric of AoD treatment success.

4.2 Clinically meaningful change for ATOP psychological health, physical health and quality of life variables

This section reports how we arrived at a threshold for clinically significant change for the ATOP items psychological health, physical health and quality of life (PPQ variables). By a clinically significant change threshold we mean the minimum amount of change – be it improvement or deterioration – in the outcome of interest that would need to be observed for that change to be considered clinically meaningful.

In order to arrive at these thresholds we drew on an extensive literature describing the best approaches for determining what amount of a change in score on a questionnaire can safely be regarded as a true representation of actual change in the construct of interest (e.g. substance use, psychological health). This literature is informed by the idea that statistical significance and clinical significance are two different things and that the standard errors that provide the basis for determining statistical significance of change are usually too small to be useful to clinicians on a case-by-case basis. The challenge for the researcher wishing to decide on a threshold for clinically significant change is to choose the method, from among the many that exist, that yields a threshold that is:

- Large enough that a clinician can be certain that the amount of change observed reflects real change and not measurement error (such as errors in recall)
- Small enough that small but meaningful actual change is not overlooked
- Best suited to the constraints of the study and the instrument
- Is straightforward enough to be used by busy clinicians as part of routine clinical care.

4.2.1 Method

Data

Data came from 278 participants in the ATOP validation project conducted by the COQI team (Deacon, Bruno et al. 2018).

In this study participants were administered a series of questionnaires both by clinicians and researchers at baseline, and again at one-month and three-month follow-up sessions. Included in this series were the ATOP and some well-tested and psychometrically validated questionnaires that were used to validate the ATOP. Here, we use PPQ variable data from the baseline and one-month ATOPs, and one set of the comparison measures.

Comparison measures

Clinician Global Impression of Severity (CGI-S) and Clinician Global Impression of Change (CGI-C) (Guy 1976):
The CGI-S measured clinicians global impression of how well the patient is functioning at the time of measurement. As well as an overall score, clinicians rated clients' substance use, mental health, physical health, and socioeconomic status. Responses were made on a 7-item response scale: 1-'Normal, not at all ill', 2-'Borderline Ill', 3-'Mildly Ill', 4-'Moderately Ill', 5-'Markedly Ill', 6-'Severely Ill', 7-'Among the Most Extremely Ill Patients'. Unlike the CGI-S, which *only* refers to current functioning, the CGI-C scale measured, for each

participant, the clinician's overall impression of how well the patient has been functioning overall compared to their functioning when they started the study. For CGI-C the clinician rates the overall change in functioning, as well as change in specific domains of substance use, physical health, mental health, and quality of life. Responses were made on a 7-item semantically symmetrical response scale: 1-'Very Much Improved since commencing the study', 2-'Much Improved', 3-'Minimally Improved', 4-'No change from commencing the study', 5-'Minimally worse', 6-'Much worse', 7-'Very much worse since commencing the study'.

Data Analysis

Clinically Significant Change: We used a sensitivity analysis to help us arrive at the thresholds for clinically significant change for the PPQ variables. A sensitivity analysis consists of using multiple methods of analysing the same data and documenting *all* the results. If all the different methods yield the same answer then it increases the confidence in and validity of that answer (McElreath 2018).

For this data-driven approach we followed the methodology outlined in Crosby, Kolotkin et al. (2003), which describes the many approaches that have been taken to establishing clinically significant change in the past. These approaches are summarised in Table 1. We used both anchor-based methods and distribution-based methods. Anchor-based methods for establishing clinically significant change all require operational definitions of what constitutes a 'condition present' state (using the original clinical language of the tests) and what constitutes a normal, recovered, or 'illness absent' state. Once criteria have been established for the two states the mean values for clients who meet criteria for each state can be calculated and the difference between those means used as the threshold for significant change. The first of the anchor-based methods we used took a cross-sectional approach, using scores on

the CGI-S at a single time point (baseline). For all four CGI-S domains (mental functioning, physical functioning, overall functioning and socioeconomic functioning) clients who were rated by clinicians as being 'Normal, not at all ill' (CGI-S = 1) or 'Borderline Ill' (CGI-S = 2) at baseline were considered to be normal or recovered and clients who were rated as being 'Markedly Ill' (CGI-S = 5), 'Severely Ill' (CGI-S = 6), or 'Among the Most Extremely Ill Patients' (CGI-S = 7) were considered to be 'disease present'. Mean ATOP psychological health scores were calculated for clients who met either 'condition present' or 'condition free' status for CGI-S-Mental Functioning, ATOP physical health for CGI-S-Physical Functioning, ATOP quality of life for both CGI-S-Overall and CGI-S-Socioeconomic functioning. The difference between these means rounded to the nearest integer was taken as the minimum difference in ATOP PPQ variable scores necessary to demonstrate clinically significant change between measurements.

The second anchor-based method took a longitudinal approach, using CGI-C scores – clinician ratings of *change* – as the anchor used to group clients' scores. Change in ATOP Psychological Health was compared to CGI-C Mental Health, ATOP Physical Health to CGI-C Physical health, and ATOP Quality of Life to CGI-C Quality of Life. For all three ATOP PPQ domains, the threshold for clinically significant improvement was defined as the mean change from baseline to one-month follow-up for all participants designated as having improved in the equivalent CGI-C domain (CGI-C scores = 1-2). The threshold for clinically significant deterioration was defined as the mean change from baseline to one-month follow-up for all participants designated as having deteriorated in the equivalent CGI-C domain (CGI-C scores = 6-7). CGI-C scores from 3-5 were defined as having experienced no change.

There are drawbacks to anchor-based approaches, including potentially non-linear relationships between the anchor and the outcome measure and large variability between global ratings made

by different clinicians using the anchor measurement. Therefore, in addition to these anchor-based methods, several distribution-based approaches to establishing thresholds for clinically significant change in the PPQ variables were also used. Distribution-based approaches are based on the statistical properties of the obtained sample rather than on reference to an outside criterion measure, and therefore are free from any of the weaknesses of that criterion measure. We used two broad classes of distribution-based measures, those based on sample variation, and those based on measurement precision.

Sample Variation: We used three methods based on sample variation to obtain thresholds for clinically significant change:

Effect Size: threshold for change is simply the standard deviation of the outcome measure at baseline.

Standardised Response Mean (SRM): threshold for change is the standard deviation for change scores (baseline – follow-up)

Responsiveness Statistic: threshold for change is the standard deviation for change for a group of clients who are deemed stable at baseline.

For all three measures clinically significant change is expressed in standard deviations from the mean. The provisional clinical threshold for each of these sample variation-based approaches was set at 1 SD.

Measurement Precision: We used two methods based on measurement precision to obtain clinically-significant change thresholds:

Standard Error of Measurement: threshold for change is the standard error of measurement, derived from sample standard deviation and the sample reliability coefficient.

Reliable Change Index: threshold for change is based on the standard error of measurement difference

Each individual's change score for all three PPQ variables was divided by each of the change thresholds referred to above, yielding frequency tables for the proportion of participants who showed clinically significant deterioration, clinically significant improvement, or no change.

Table 1: Methods for determining clinically significant change in ATOP psychological health, physical health, and quality of life variables.

Method	Individual Change Measured in relation to	Calculation	Advantages	Disadvantages
Anchor-based Methods				
Cross-sectional	Difference in mean score at baseline for people rated by clinicians on the CGI-S as having normal or borderline normal functioning and mean score for those rated as being extremely ill	$\frac{x_1 - x_0}{\mu_{CGI-S\ not\ ill} - \mu_{CGI-S\ ill}}$	• Threshold tied to clinical judgements of whether the condition is present vs whether it is absent	• May be subject to unreliability due to weaknesses of the anchor (e.g. unreliability) or to a non-linear relationship between the anchor and the outcome variable
Longitudinal	Average change score for those who were rated by clinicians as having improved or deteriorated	$\frac{x_1 - x_0}{\mu_{CGI-C\ improved}}$ or $\frac{x_1 - x_0}{\mu_{CGI-C\ deteriorated}}$	• Threshold tied to clinical judgements, with potentially different thresholds for improvement and deterioration.	• Same as above

Table 1 cont.

Method	Individual Change Measured in relation to	Calculation	Advantages	Disadvantages
Distribution-based Methods				
Sample Variation based				
Effect Size	Baseline Standard Deviation	$\frac{x_1 - x_0}{\sqrt{\frac{\Sigma(x_0 - \overline{x}_0)^2}{n-1}}}$	• Standardised units • Benchmarks for interpretation • Independent of sample size	• Sensitive to variability of outcome measure • Does not consider variability of change • May vary among samples
Standardised Response mean	Standard deviation of change	$\frac{x_1 - x_0}{\sqrt{\frac{\Sigma(d_i - \overline{d})^2}{n-1}}}$	• Standardised units • Independent of sample size • Based on variability of change	• Sensitive to relative success of treatment in individual samples
Responsiveness Statistic	Standard deviation of change in a stable group	$\frac{x_1 - x_0}{\sqrt{\frac{\Sigma(d_{i\,stable} - \overline{d}_{stable})^2}{n-1}}}$	• Standardised units • More conservative than effect size • Independent of sample size • Takes into account spurious change due to measurement error	• Data on stable subjects often not available

Table 1 cont.

Method	Individual Change Measured in relation to	Calculation	• Advantages	• Disadvantages
Measurement Precision based				
Standard Error of Measurement	Standard error of measurement	$\frac{x_1 - x_0}{\sqrt{\frac{\Sigma(x_0 - \overline{x}_0)^2}{n-1}}(\sqrt{1-r})}$	• Stable across populations • Takes account of precision of instrument • Cutoffs based on confidence intervals	• Assumes measurement error to be constant across the possible range of scores
Reliable Change Index	Standard error of the measurement difference	$\frac{x_1 - x_0}{\sqrt{2(SEM)^2}}$	• Stable across populations • Takes account of precision of instrument • Cutoffs based on confidence interval	• Assumes measurement error to be constant across the possible range of scores

$x_0 =$ individual baseline score
$x_1 =$ individual 4-week follow-up score
$\mu_{CGI-S\ not\ ill}$ = Mean PPQ score for those rated on the CGI-S at baseline by as being 'Not at all ill' or 'Borderline Ill'
$\mu_{CGI-S\ ill}$ = Mean PPQ score for those rated on the CGI-S at baseline as being 'Severely ill' or 'Among the most Extremely Ill Patients'
$\mu_{CGI-C\ improved}$ = Mean PPQ change score for those rated on CGI-C at follow-up as being 'Very much improved' or 'Much Improved'
$\mu_{CGI-C\ deteriorated}$ = Mean PPQ change score for those rated on CGI-C at follow-up as being 'Very much improved' or 'Much Improved'
$\overline{x}_0 =$ mean baseline score
$d_i =$ difference score (baseline – follow-up) for participant i
$\overline{d} =$ mean difference score
$d_{i\ stable} =$ difference score (baseline – follow-up) for participant i who belongs to stable group
$\overline{d}_{stable} =$ mean difference score for those in stable group
$r =$ test-retest reliability of the measure in question
$SEM =$ Standard Error of Measurement

4.2.2 Results

Table 2 shows the results of the sensitivity analysis of data-driven methods for deriving clinically-significant change thresholds for the ATOP PPQ variables.

Rounded clinically-significant change thresholds using the anchor-based cross-sectional method were 3-2-3 for psychological health, physical health, and quality of life respectively. The longitudinal anchor-based method had thresholds of 1 (for clinical improvement) and 2 (clinical deterioration) for psychological health, 0 (improvement) and 0 (deterioration) for physical health, and 1 (improvement) and 1 (deterioration) for quality of life. All anchor-based methods rely on having gathered data from both an 'ill' population and a healthy population. We attempted to simulate these two populations within our cohort of treatment seekers, by measuring mean PPQ scores for those rated as normal or borderline ill on the CGI-S (anchor-based cross-sectional method) or by measuring mean PPQ change scores for those rated as having improved or deteriorated on the CGI-C (anchor-based longitudinal). Unfortunately however, the numbers of participants who were either rated as ill on the CGI-S or as having improved on the CGI-C was very small. Furthermore, it is doubtful that the people selected by clinicians as 'not ill', all of whom were actively seeking treatment for D&A problems, could be considered a properly recovered or normal population. Until we have ATOP data from people who have used substances then stopped problematic use and experienced important sustained improvements in their health/social circumstances or from general populations, anchor-based methods are probably not suitable ways of estimating clinically significant change, making data-driven methods more suitable.

Four of the five distribution-based methods agreed on a threshold of 2 points for all three PPQ variables. The Standard Error of Measurement was the only data-driven method that yielded a

different threshold – 1 point for all three PPQ scales – however Crosby, Kolotkin et al. (2003) point out that this method is more prone to false positives than other methods and recommend the Reliable Change Index instead.

Table 2: Results of seven different methods for determining clinically significant change in ATOP psychological health, physical health, and quality of life variables

		Threshold for Sign Change	Proportion Changed, n (%)		
Method	**PPQ Variable**	**Threshold (rounded up); n**	**Improved**	**No Change**	**Deteriorated**
Anchor-based methods					
Cross-sectional	*Psychological Health*	2.76 (3); n_{ill}=17, n_{well}=184	26 (11)	190 (81)	19 (8)
	Physical Health	1.79 (2); n_{ill}=16, n_{well}=177	43 (18)	153 (65)	39 (17)
	Quality of Life	2.81 (3); n_{ill}=32, n_{well}=114	35 (15)	186 (79)	14 (6)
Longitudinal	*Psychological Health*				
	CGI-C Improved	0.61 (1); *n*=51	98 (42)	96 (41)	41 (17)
	CGI-C Deteriorated	1.80 (2); *n*=6			
	Physical Health				
	CGI-C Improved	0.43 (0); *n*=41	82 (35)	75 (32)	78 (33)
	CGI-C Deteriorated	0.00 (0); *n*=4			
	Quality of Life				
	CGI-C Improved	1.15 (1); *n*=66	59 (25)	99 (42)	77 (33)
	CGI-C Deteriorated	0.50 (1); *n*=7			

Table 2: cont

		Threshold for Sign Change	Proportion Changed, n (%)		
Method	**PPQ Variable**	**Threshold (rounded up); n**	**Improved**	**No Change**	**Deteriorated**
Distribution-based methods					
Sample Variation based					
Effect Size	*Psychological Health*	2.14 (2)	26 (11)	190 (81)	19 (8)
	Physical Health	2.15 (2)	22 (9)	195 (83)	18 (8)
	Quality of Life	2.31 (2)	35 (15)	186 (79)	14 (6)
Standardised Response mean	*Psychological Health*	1.91 (2)	55 (23)	139 (59)	41 (17)
	Physical Health	1.82 (2)	50 (21)	146 (62)	39 (17)
	Quality of Life	2.00 (2)	59 (25)	140 (60)	36 (15)
Responsiveness Statistic	*Psychological Health*	1.85 (2)	55 (23)	139 (59)	41 (17)
	Physical Health	1.77 (2)	43 (18)	153 (65)	39 (17)
	Quality of Life	1.92 (2)	59 (25)	140 (60)	36 (15)

Table 2: cont

		Threshold for Sign Change	Proportion Changed, n (%)		
Method	**PPQ Variable**	**Threshold (rounded up); n**	**Improved**	**No Change**	**Deteriorated**
Measurement Precision based					
Standard Error of Measurement	*Psychological Health*	1.10 (1)	55 (23)	139 (59)	41 (17)
	Physical Health	1.26 (1)	43 (18)	153 (65)	39 (17)
	Quality of Life	1.36 (1)	59 (25)	140 (60)	36 (15)
Reliable Change Index	*Psychological Health*	1.55 (2)	55 (23)	139 (59)	41 (17)
	Physical Health	1.79 (2)	43 (18)	153 (65)	39 (17)
	Quality of Life	1.92 (2)	59 (25)	140 (60)	36 (15)

Key

$x_0 =$ individual baseline score

$x_1 =$ individual 4-week follow-up score

$\mu_{CGI-S\,not\,ill} =$ Mean PPQ score for those rated on the CGIS at baseline by as being 'Not at all ill' or 'Borderline Ill'

$\mu_{CGI-S\,ill} =$ Mean PPQ score for those rated on the CGIS at baseline as being 'Severely ill' or 'Among the most Extremely Ill Patients'

$\mu_{CGI-C\,improved} =$ Mean PPQ change score for those rated on CGIC at follow-up as being 'Very much improved' or 'Much Improved'

$\mu_{CGI-C\,deteriorated} =$ Mean PPQ change score for those rated on CGIC at follow-up as being 'Very much improved' or 'Much Improved'

$\overline{x_0} =$ mean baseline score

$d_i =$ difference score (baseline – follow-up) for participant *i*

$\overline{d} =$ mean difference score

$d_{i\,stable} =$ difference score (baseline – follow-up) for participant *i* who belongs to stable group

$\overline{d}_{stable} =$ mean difference score for those in stable group

$r =$ test-retest reliability of the measure in question

$SEM =$ Standard Error of Measurement

4.2.3 Discussion

Four of the five distribution-based methods reported in the sensitivity analysis of the ATOP PPQ variables suggest that a threshold for change in the PPQ variables of 2 points represents a clinically significant change. This threshold for change was also endorsed by a consensus of clinical experts in D&A treatment (Section 4.4), hence it is with some confidence that we recommend that a change of two points in either direction in either ATOP psychological health, physical health, or quality of life, represents a change worthy of clinical attention.

It is important to note that the ATOP outcome algorithm is designed for interpreting group data, rather than for analysing an individual client's outcomes in a clinical context. The clinical function of the ATOP is as (1) a tool to allow clinicians to track a client's progress over time, (2) as a way of structuring a clinical interview, (3) to potentially alert them to change in their client's health and wellbeing that is worthy of their (and the client's) attention. The ATOP outcome algorithm was *not* designed to be a substitute for the experience and judgement of the individual clinician. What action the individual clinicians takes when reviewing changes in ATOP scores is described in ATOP Manual 1 (Chapter 3 section 1iv).

4.3 Data-driven methods for calculating clinically significant change for ATOP substance use variables

Issues with the distributions of ATOP substance use variables precluded direct application of the RCI method for determining significant change in these variables. To address this, we divided our frequency data into two samples – high use and low use. We describe the approach and rationale below.

Skew in substance use frequency data is not uncommon (e.g. Maisto, Kaczynski et al. 1996, Roberts, Neal et al. 2000). Drug and alcohol researchers have used an array of methods to address this problem. One approach has been to mathematically transform the outcome variable. Marsden, Eastwood et al. (2011; p296) for example, applied a root arcsine transformation to their substance use response data to "improve distributional characteristics". Transformations are useful in that they convert the data to a normal distribution that is consistent with the assumptions of the change metrics, but yield thresholds that are very difficult to interpret or implement. Furthermore, our frequency of substance use data – reported here and in most clinical samples the COQI team work with – are distributed bimodally (a 'bathtub' or 'U' shaped curve) with the majority of values lying at either extreme of the scale, and very few in the middle of the scale. That is, most clients use either daily or near daily, or alternatively use very infrequently or not at all. In these circumstances transformations such as root arcsine are of little use.

There are other data-driven methods for determining meaningful change that are reliant on having normative data, that is, data on the outcome in question taken from a 'normal' population as

described in Section 4.2. When it comes to substance use however these methods are problematic. For example days of heroin use in the general population samples would yield frequencies of use where the overwhelming majority use 0 days per week. Normative samples should also have a normal distribution of scores for the variable of interest - a distribution where 99% of scores are 0 is not useful.

4.3.1 Method

The RCI, conceptualized by Jacobson and Truax (1991), is equal to the difference in scores (x_1 and x_2) at two time-points divided by the standard error of the difference (S_{diff}) of the measure in question:

$$RCI = \frac{x_1 - x_0}{S_{diff}}.$$

Here x_1 and x_0 are a subject's follow-up and initial scores on the measure respectively. S_{diff} can be calculated from the standard deviation (SD) of a "normal" population's scores on the measure, and the test-retest reliability (r) of the measure, such that

$$RCI = \frac{x_1 - x_0}{SD\sqrt{2(1-r)}}.$$

An RCI larger than 1.96 is considered to be unlikely to occur at the 95% level of confidence ($p<0.05$). The minimum change in a score for there to be reliable change is thus:

$$x_1 - x_0 = 1.96 * SD\sqrt{2(1-r)}$$

Given there is no concept of a 'normal' level of use for most substances, generally in addiction research the standard deviation of the sample at time 0 is used (e.g. Kelly 2021, Marsden 2010) instead of a reference sample.

4.3.2 Data sources

- to obtain test-retest reliabilities *r*, we sampled 94 clients at SESLHD D&A services, who completed two ATOPs within a period of 3 days with the same researcher, and
- to obtain standard deviations, ATOP data from a normative clinical sample of 6,100 entrants to AoD treatment was obtained (see Appendix A). For each substance category, we selected only those ATOPs from clients whose PDOC was the substance of interest.

Days of use data in the normative sample were skewed towards the minimum and maximum (0 and 28 days), indicating two separate underlying populations – lower and higher use.

Visual observation of the distributions and Gaussian model fitting indicated there was no distinct separation between the two populations. Different categorisations of days of use were trialled to separate the low and high use populations –10 12, and 14 days – and the clinical consultation group (Section 4.3) was presented with these options to determine which would be most clinically practical.

There was consensus that 12 days was best as it represents a whole 'average' number (3) of days per week of use (out of 28 days). Therefore, we calculated standard deviations of the normative datasets on split samples – 0-12 days, and 13-28 days of use, for each substance type.

4.3.3 Results

In Table 3, we present *r* values, SDs and corresponding 95% RCIs for each substance use category, and for 0-12 and 13-28 days of use.

Table 3 Test-retest reliabilities for ATOP substance use variables from the test-retest dataset, means and standard deviations (SD) for clients with primary drug of concern, and 0-12 or 13-28 days of use for each substance from the 7-LHD normative dataset, and calculated Reliable Change Indices (RCIs).

Substance use variables	Test-retest reliability	0-12 days of use		13-28 days of use	
		SD	95% RCI	SD	95% RCI
Alcohol days use	0.94	4.3	3.0	4.9	3.4
Cannabis days use	0.92	4.0	3.2	4.2	3.3
Amphetamines days use	0.96	3.8	2.1	5.3	2.9
Benzodiazepines days use	0.89	3.7	3.4	2.9	2.6
Heroin days use	0.98	2.7	1.1	4.6	1.8
Other opioids days use	0.94*	2.9	2.0	3.9	2.6
Any opioids days use	0.94*	2.6	1.8	4.1	2.8
Injecting days	0.97	2.8	1.3	4.4	2.0

** average of alcohol, cannabis, amphetamines, benzodiazepines, heroin and injecting days, as unable to be measured using the test-retest dataset*

In order to be consistent for clinicians to apply change rules across low and high use and different substance types, 4 days – the largest 95% RCI in table 1 rounded up to the nearest whole number - was chosen as the minimum change required to indicate a statistically reliable change. This was presented to the clinical consultation group (see section 4.4. below).

4.3.4 Discussion

Applying the RCI method to our split distribution, we have calculated that a change of 4 or more days in 28-day substance use is required

to be a clinically meaningful change. We note that previous uses of RCI calculation on substance use days have acknowledged their samples not being normally distributed. Marsden, Eastwood et al. (2011) approached the issue by considering (1) days of substance use only among those with any use (equal to or greater than 1 day) at time 0, and (2) each substance use response variable was computed as a proportion and root arcsine-transformed. Cisler, Kowalchuk et al. (2005), alcohol use was measured using percent days abstinent in the last 90, giving a range 0-1. At 6-month's follow up, their 1726 participants with alcohol problems had a mean of 79% days abstinent with standard deviation of 0.29, indicating a skewed distribution. They acknowledge methodological issues in determining clinical significance including the distributional properties of the measures, yet went with the Jacobson and Truax approach for determining functional status and individuals' reliable change. They calculated that a change in substance use of 14 days in the previous 90-day period constituted a reliable change. Roberts, Neal et al. (2000) used a 0-5 rating scale for drinking frequency, from 0 (less than once per month) to 5 (nearly every day). The distributions of all five dependent measures were heavily skewed and the authors acknowledged the problems of calculating clinical significance in outcomes with non-normal distributions.

The issue of calculating reliable change for substance use frequency measures is complex, and there are multiple approaches. Our approach, which sets a lower limit of 4 days in the previous 28-day period for statistically reliable change on the ATOP substance use frequency variables, is a new way of approaching RCI for non-normal distributions.

4.4 Consultation with clinical stakeholders on meaningful change

A day-long workshop was held in September 2018 to gather feedback from experienced clinicians working in the public AOD sector in NSW concerning what they would consider clinically significant changes in frequency of substance use and on the PPQ items. At this workshop clinicians were presented with a range of options for deriving clinically significant change thresholds, as well as categorisations for 'low' and 'high' substance use frequency, and cutoff flags for the PPQ variables (reported in Mammen, Mills et al. 2021).

Broadly speaking there are two approaches that researchers can use to determine thresholds for clinically-significant change: Data-driven and expert-consensus-driven. Each of these methods have their strengths and weaknesses. Expert-consensus can be unreliable due to the biases introduced by clinical folklore and individual opinion, but data-driven methods can sometimes be inflexible and can run into difficulties when the outcome variables are complex. Where possible it is best to use both methods, for, if both methods agree, it increases the likelihood that the threshold arrived at is both clinically useful and empirically valid.

4.4.1 Method

Clinicians were seated at tables of 4-6 and each table was asked to discuss the relative merits of different criteria for clinically significant change and clinical cutoffs and then to indicate a preference. The options that were a clear majority among the groups were adopted as the criteria for clinically meaningful change.

For substance use:

Categorisation of 'low' and 'high' frequency use

Frequency distributions are bimodal, indicating two different populations are entering treatment. What should the delineation between 'low' and 'high' use be?

- Midway: 0-13 and 14-28 days
- 3 days per week: 0-12 and 13-28 days
- 4 days per week: 0-16 and 17-28 days

Clinically significant change

The reliable change index shows 4 days of change is considered as the minimum reliable change – but is this clinically meaningful? Four other options were presented: 8 days per month (i.e. approximately 2 days per week); 12 days per month (i.e. approximately 3 days per week); 30% change in days used (minimum 4 days), and 50% change in days used (minimum 4 days). The last two options were selected as typical minimum changes used in the medical literature and are consistent with advice around interpretation of effect sizes (e.g. Cohen 1988).

For PPQ variables:

Clinically significant change

The reliable change index shows 2 points of change is considered as the minimum reliable change – but is this clinically meaningful?

4.4.2 Results

3.2 *Clinically Significant Change: Substance Use*

For the categorisation of low and high frequency use, the clinical expert group decided that 12 days was appropriate as it was equivalent to a whole number of days (3) per week. Low frequency use thus refers to 0-12 days, and high frequency is 13-28 days.

Table 4 shows each D&A clinical experts group's response to the question "How much change is needed in the frequency of use of principle drug of concern for us to determine that there has been a clinically significant change?"

Table 4. Expert-informed thresholds for clinically significant change by group, based on clients' PDOC and their frequency of use at assessment

Group	Alcohol		Opioid	
	Low frequency	High frequency	Low frequency	High frequency
Group 1	30%	30%	50%	50%
Group 2	30%	30%	NA[a]	NA
Group 3	2 days/week	2 days/week	2 days/week	2 days/week
Group 4	30%	30%	30%	30%
Group 5	30%	30%	50% (at least)	50% (at least)

Note: Each group was made up of 4-6 people, mostly clinicians but some researchers and Health-related data managers working in the D&A treatment sector. Groups were asked to consider four threshold options: 8 days per month (ie approximately 2 per week); 12 days per month (ie approximately 3 per week); 30% change (minimum RCI, equivalent to 4+ days of change), and 50% change (minimum RCI). a: Group 2 did not nominate a change threshold for opiates.

The most common threshold chosen by the clinical experts groups for clients with alcohol as PDOC was a 30% change in either direction in number of days' used in the previous 28-day period. The clinical expert groups' most common choice of change threshold for clients with opioids as PDOC was 50%. However, after a common discussion among all groups in the room they decided that 30% would also be the change threshold for clients whose PDOC is opioids, so as to simplify the guidelines for minimum clinically-significant change to make them easier to apply for busy working clinicians. While percentage-change thresholds are more flexible than thresholds based on absolute number of days' change, they do encounter mathematical difficulties when initial days used are low. Consider, for example what a 30% increase in days used would be for someone whose initial days used was 0 or 1 day in the past 28.

To sidestep these issues the 30% threshold was combined with the reliable change index such that whichever is the higher, is the minimum clinically significant change.

3.3 *Clinically Significant Change: PPQ variables*

Clinicians were shown the change thresholds for the PPQ variables as derived from the different data-driven methods in Section 4.2. Clinicians were told that these were minimum thresholds, and were asked whether they would consider these thresholds large enough, or whether larger thresholds would be required to be clinically significant. The group endorsed 2 points as the threshold for clinically significant change for all PPQ variables.

Chapter 4 key points:

- **A threshold for change in the PPQ variables of 2 points represents a clinically significant change**
- **A change of 4 or more days in 28-day substance use is required to be a clinically meaningful change**
- **Clinical stakeholders' consultation decisions:**
  - **Categorization of low and high frequency substance use to be ≤12 days and 12+ days**
  - **30% minimum change or 4+ days (whichever is higher) is the threshold for clinical significant change for substance use**
  - **2+ points changes is the threshold for clinically significant change for PPQ variables**

Chapter 5: Defining the COQI change metric.

Outcomes metrics are predetermined measure of success and they are used in many areas of health care, and more broadly in social services, education, business and finance. When used for measuring treatment outcomes, the metric defines the conditions that are needed to say that a treatment was successful, providing a standardised approach for a service sector to make comparisons across treatment types, service models, and over time. It can also be applied in clinical research, so that each study is not required to independently define treatment success.

The COQI Program aims to facilitate services to use their routinely collected clinical information to measure and improve the quality and outcome of treatment. Services that have Clinical Information Systems (CIS) have access to large amounts of clinical data that can be used to evaluate and improve service delivery, but knowing how to use the information can be challenging for treatment services. One of the ways the COQI Program has attempted to do this is by working with key stakeholders to predefine what a treatment success on the ATOP would look like, allowing us to develop an outcomes metric.

Why do we need an 'outcomes metric'?

The development of an outcomes metric in AoD treatment is challenging as there are a range of approaches and stakeholder perspectives on what constitutes a 'treatment success'. Success may mean different things for different treatment types and at different stages in a person's treatment journey. It is also unlikely that all aspects of treatment outcome are quantitatively measureable. However, the benefits of a predefined metric for assisting treatment services to use their own data for improvement, outweigh the challenges of defining success.

So what are some of these benefits? An outcome metric allows us to: identify whether an expected treatment outcome has been achieved; it can be programmed into services' clinical information systems, supporting services to make meaning of the large volume of clinical information available to them, and; facilitate more timely and efficient access to outcomes results.

Services can then use the information in a range of ways including in quality improvement and research. Having direct access to this information at a service level may facilitate a sense of ownership of any findings and increase the likelihood that this will be translated into clinical practice improvement.

In terms of data quality, that sense of ownership can also improve the completeness and accuracy of outcome information and improve motivation for routine collection.

A standardised outcomes metric also opens up opportunities for benchmarking (over time within the one service, across services within an organisation, and externally with like services), best conducted using the benchmarking principles of respectful and collaborative inquiry. A binary outcome (ie 2 categories - % good outcome and % poor outcome) has good communication power within the health system including through the use of analytics platform such as dashboards, and it becomes particularly powerful when it accompanied by a narrative that describes the treatment provided, the clients, and their experiences.

As already noted, it is also **very** important to remember that the outcome metric is not meant as a client feedback tool

Success does not always mean change

The COQI metric extends the idea of clinically meaningful change and transforms it into a binary metric. The next step, after determining the amount of change in scores needed to represent clinically significant change, is to consider whether a client *always* needs to improve their ATOP

scores for their AoD treatment to be considered 'successful' . In fact there are a number of scenarios in AoD treatment where 'staying the same' is a successful outcome. For example, when clients who are stable transfer to a new treatment setting, a successful outcome includes keeping their entry level of substance use and psychological and physical health and quality of life rates the same. Likewise for clients attending a counselling service for relapse prevention: remaining on zero days of substance use from when counselling began to the end of treatment is most definitely a successful outcome! The metric will also need to be able to detect when a client is deteriorating.

We identified cut off points on the items that would indicate whether a client whose score stay the same should be defined as a successful or unsuccessful outcome: for substance use the cutoffs represent low and high substance use frequency, and for the psychological and physical health, and quality of life items the cutoffs indicate whether the client is likely to be experiencing a problem in those areas.

The details of the steps undertaken to determine the algorithm were discussed in Chapter 4. The algorithm for the COQI metric is in Table 1 (reproduced from Chapter 3 Section 3.2.3) which defines successful and not successful treatment outcomes.

Table 1: COQI algorithm for clinically meaningful change in substance use, and psychological health, physical health and quality of life (the PPQ variables). Treatment outcome is deemed successful or unsuccessful depending on both the starting point, and the threshold of change required to be clinically meaningful.

Frequency of substance use at measurement A	**Relative change in frequency of substance use in previous 28 days at measurement B**	**Change category**	**Treatment outcome**
Low (≤12 days in previous 28)	Increased by ≥4 days use compared to measurement A	Significant increase	**Unsuccessful**
	Reduced by ≥4 days use compared to measurement A	Significant decrease	**Successful**
	Increase or decrease of <4 days use compared to measurement A	No significant change	**Successful**
	No change from zero use	Maintained no use	**Successful**
High (>12 days in previous 28)	≥30% increase in days use compared to measurement A	Significant increase	**Unsuccessful**
	≥30% decrease in days use compared to measurement A	Significant decrease	**Successful**
	<30% increase or decrease in days use compared to measurement A	No significant change	**Unsuccessful**
Rating of PPQ variable at measurement A	**Relative change in rating of PPQ variable at measurement B**	**Change category**	**Treatment outcome**
Poor (score of ≤5 on 0-10 scale)	Score increase of ≥2 from score at measurement A	Significant increase	**Successful**
	Score decrease of ≥2 from score at measurement A	Significant decrease	**Unsuccessful**
	Increase or decrease of <2 from score at measurement A	No significant change	**Unsuccessful**
Normal-well (score of > 5 on 0-10 scale)	Score increase of ≥2 from score at measurement A	Significant increase	**Successful**
	Score decrease of ≥2 from score at measurement A	Significant decrease	**Unsuccessful**
	Increase or decrease of <2 from score at measurement A	No significant change	**Successful**

Appendix A: ATOP Normative Data: Clients entering treatment.

The objective of this study was to derive normative means and standard deviations for continuous items from the ATOP (days in last 28 of substance use, injecting drug use, work and training/education) and normative proportions for dichotomous items from the ATOP (daily tobacco use, sharing of injecting equipment, homelessness or risk, arrests, caring for children, experience of violence).

Method

Data from *N*=6,100 entrants to public drug and alcohol treatment services were extracted from the Community Health and Outpatient Care (CHOC) electronic clinical information system for the period January – December 2017. Seven NSW Health Local Health Districts (LHDs) participated, representing a spread of data across metropolitan and regional areas. We have not identified the LHDs for confidentiality purposes.

Data items included LHD, gender, principal drug of concern, age and all ATOP data items.

Analysis

For continuous items of the ATOP, mean and standard deviations were calculated for days of use for principal drug of concern (PDOC) and overall. The following continuous items were also recoded into dichotomous variables: any substance use (by type); any injecting; any paid work; and any education. Similarly, quality of life, physical and psychological wellbeing items were dichotomised such that scores of 0-5 were collapsed to indicate possible challenges in this area suitable for clinical follow up ('poor', and

scores of 6-10 were collapsed to indicate this life area was 'good'; (Lintzeris, Mammen et al. 2020).

For dichotomous items, proportions reporting each item were calculated for each PDOC and overall.

Findings

Client descriptors

Clients were predominantly male [where details were available] with a mean age of 39. Treatment entrants were primarily seeking support related to their alcohol use (Table 1).

Table 1: Demographic characteristics and principal drug of concern at treatment entry, 2017

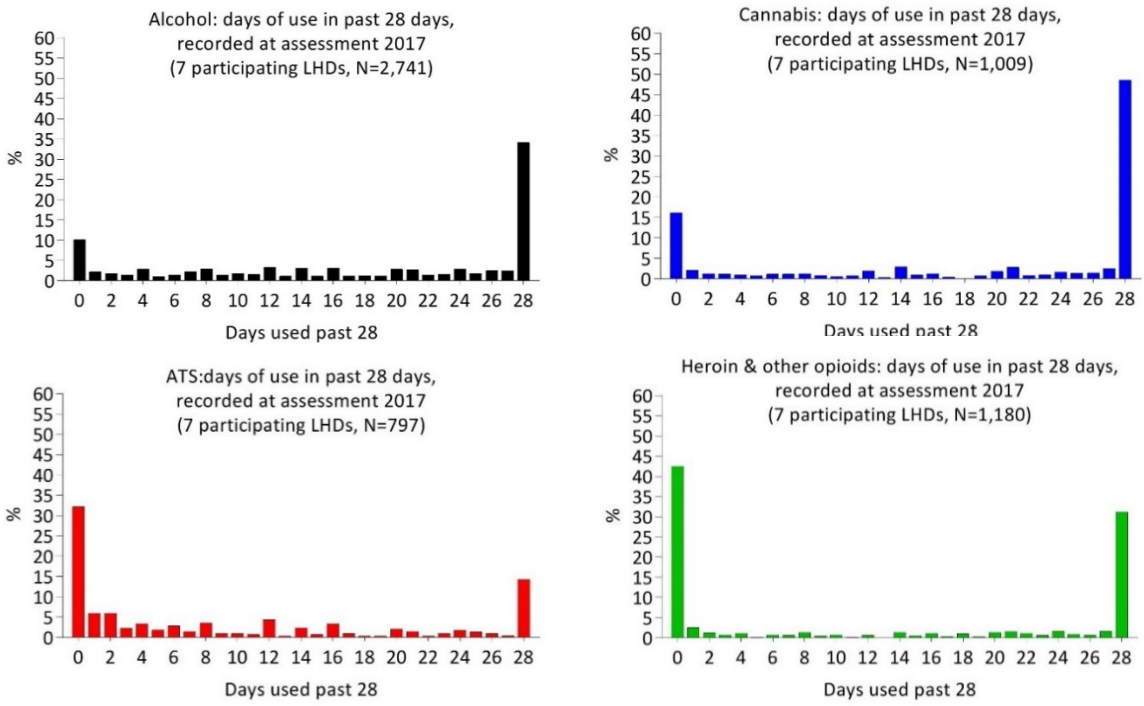
Demographics	
	N=6100
Age: M (years)	39
Age SD, range	*12, 9-86*
% Female	32[1]
Principal drug of concern	%
Alcohol	46
Amphetamine type substances	14
Benzodiazepines	1
Cannabis	17
Cocaine	1
Heroin & other opioids	21

[1]Sex was not provided by a number of LHDs due to issues with extraction

Patterns of drug use in the 4 weeks preceding assessment

Inspection of the data reveals that the number of days of use of each PDOC typically has a bimodal distribution, such that the most common number of days of use is 0 days (abstinence) or 28 days (daily use) in the 4 weeks preceding assessment. Therefore these items are displayed below as frequency figures (Figure 1).

Figure 1: Patterns of recent drug use among people reporting the drug as their principal drug of concern



Dichotomous ATOP items and dichotomous items derived from the ATOP are shown in Table 2. Overall these findings reflect a wide range of activities and life stresses experienced by clients preceding treatment entry.

Table 2: Normative proportions endorsing situational items from the ATOP

ATOP item	Overall
	N=6100
% daily tobacco use	62
% any days paid work	32
% any days education	5
% any injecting	15
among those who had recently injected: % shared injecting equipment	*15*
% acute housing problem	10
% at risk of eviction	7
% caring for children 0-4y	10
% caring for children 5-15y	15
% arrest	11
% experienced violence to self	8
% violent to others	7

Self-reported wellbeing

Means and standard deviations for the psychological health, physical health, and quality of life (the "PPQ items") are shown in Table 3 below by PDOC. Scores range from 0 (poorest) to 10 (highest).

Table 3: Psychological health, physical health and quality of life self-ratings by PDOC at assessment

		Alcohol	Cannabis	ATS[1]	Opioids[2]	Overall
Psychological health	n	2446	901	745	1113	5317
	M	5.1	5.6	5.3	5.6	5.3
	SD	2.2	2.2	2.2	2.2	2.2
Physical health	n	2440	903	743	1111	5308
	M	5.8	6.4	6.2	5.9	6.0
	SD	2.1	2.0	2.0	2.1	2.1
Quality of life	n	2434	892	741	1109	5286
	M	5.4	5.9	5.6	5.8	5.6
	SD	2.3	2.2	2.3	2.3	2.3

[1]*Amphetamine type stimulants*

[2]*Includes heroin and other opioids*

In general, data completeness was very high (80% or greater) and consistent at initial assessment across all LHDs (Tables 9a-9e). Please note that gender is routinely recorded, however was not provided by a number of LHDs due to issues with extraction. The physical and psychological wellbeing and quality of life (PPQ) items typically had the lowest rates of completion.

Appendix B: Overview of the use of ATOP, and related TOP tools, in research

Adaptation

The ATOP has been adapted for use in Greece (Karakoula, Kokkolis et al. 2021) . The team selected the ATOP rather than the TOP to adapt due to the greater similarity of the ATOP's substance use categories to use patterns in Greece, use of 0-10 rather than 0-20 scales, and less detailed crime questions. Additional changes for the Hellenic population were inclusion of an item assessing sports and/or volunteer work as well as employment and study and addition of route of administration for each substance.

Use of the full ATOP

Examples of studies that have used the instrument in its entirety include Berry, Jacomb et al. (2019) who use the ATOP at baseline and follow-up to assess a stepped wedge cluster randomised trial of a cognitive remediation intervention in alcohol and other drug residential treatment services. Monds, Ridley et al. (2017) also included most ATOP items, including the three PPQ variables and substance use items to examine cognition and adaptive functioning in older people attending AOD services.

Use of parts of the ATOP

Other studies have included a partial ATOP, such utilising only the PPQ variables, or the substance use items alone.

PPQ items:

Barker, Best et al. (2016) used the PPQ measures to assess differences between groups of people seeking help for substance use who were classified using a modified tiered model of substance use severity and life complexity.

Bathish, Best et al. (2017) used the quality of life measure only in a cross-sectional survey of people in recovery from addiction to assess the role of social network and social identity factors on their transition to recovery. The quality of life variable was also used in an audit of screening data for clients of specialist AoD clinics in Victoria, Australia, which examined severity of alcohol use disorder by country of birth (Savic, Barker et al. 2014).

Substance use items:

Reilly, Wand et al. (2020) have conducted a national survey of methamphetamine use among Aboriginal and Torres Strait Islander people and non-Indigenous people. Allsop, Rooney et al. (2017) used the ATOP to measure participants' substance use in the past 28 days in a randomised controlled trial investigating the effect of exercise on cannabis withdrawal symptoms. ATOP substance use variables were also used in an RCT of depot buprenorphine in people with opioid dependence (Larance, Byrne et al. 2020). Dore, Sinclair et al. (2015) used the substance use variables as one set of measures to examine progress of 40 alcohol dependent clients in the NSW Involuntary Drug and Alcohol Treatment program. Furthermore, Haslam and colleagues (Best, Haslam et al. 2016, Dingle, Haslam et al. 2019) used the quality of life and substance use items to describe characteristics and measure change in people enrolled in a prospective cohort study of clients entering therapeutic communities throughout Australia.

Use of the UK TOP

The TOP instrument was developed by a team working in the UK drug and alcohol sector (Marsden, Farrell et al. 2008, Marsden, Eastwood et al. 2011) and it is now integrated into the National Drug Treatment

Monitoring System (NDTMS). The NDTMS records anonymous information about people receiving structured care-planned drug treatment in the UK with the aim of monitoring and describing the number of people attending treatment and to assist in reviews of treatment effectiveness. This has facilitated further use of routinely collected TOP data from the NDTMS in research studies.

Examples of research studies using the TOP include a study of variability in the effectiveness of treatment for substance abuse disorder for clients with heroin-related problems (Marsden, Eastwood et al. 2012) and an examination of changes in clients receiving treatment for heroin or crack cocaine use (Marsden, Eastwood et al. 2009). Treatment outcomes have also been examined among people attending inpatient withdrawal and residential rehabilitation treatment for alcohol use disorder (Eastwood, Peacock et al. 2018) and treatment for cannabis use (Kovac, Abbasi et al. 2015). It was also used in a study of women in treatment for opioid addiction (Cornford, Close et al. 2015) focusing on contraceptive use, pregnancy outcomes and their association with a range of risk factors including substance use and physical health, psychological health and overall wellbeing. In addition, Dalton, Crowley et al. (2017) used the TOP to compare exit from treatment TOP scores to post-discharge "check in" TOPs, and the TOP was also included in a feasibility randomised controlled trial of behavioural activation vs cognitive behavioural therapy-based guided self-help (Delgadillo, Gore et al. 2015).

Further development of the TOP instrument includes the development of the Addiction Dimensions for Assessment and Personalised Treatment (Marsden, Eastwood et al. 2014). We also note an audit of the use of the TOP in three treatment services which found large disparities between the cost of clients' reported drug use, and their reported income, suggesting clients strongly underreported acquisitive crime and that these items may not be valid (Luty, Varughese et al. 2009).

References

Allsop, D. J., K. Rooney, J. C. Arnold, A. K. Bhardwaj, R. Bruno, D. J. Bartlett, M. Montebello, T. Arkell, E. Richards, J. Gugusheff, S. Rooke, W. Kerley, B. Murnion, P. Haber, I. S. McGregor and N. Lintzeris (2017). "Randomised controlled trial (RCT) of daily aerobic exercise for inpatient cannabis withdrawal: A study protocol." Mental Health and Physical Activity **13**: 57-67.

Australian Institute of Health and Welfare (2020). Health expenditure Australia 2018-19. Canberra.

Barker, S. F., D. Best, V. Manning, M. Savic, D. I. Lubman and B. Rush (2016). "A tiered model of substance use severity and life complexity: Potential for application to needs-based planning." Substance Abuse **37**(4): 526-533.

Bathish, R., D. Best, M. Savic, M. Beckwith, J. Mackenzie and D. I. Lubman (2017). ""Is it me or should my friends take the credit?" The role of social networks and social identity in recovery from addiction." Journal of Applied Social Psychology **47**(1): 35-46.

Berry, J., I. Jacomb, J. Lunn, A. Sedwell, A. Shakeshaft, P. J. Kelly, P. Sarrami, M. James, S. Russell, T. Nardo, D. Barker and J. Holmes (2019). "A stepped wedge cluster randomised trial of a cognitive remediation intervention in alcohol and other drug (AOD) residential treatment services." BMC Psychiatry **19**(1): 70.

Best, D. W., C. Haslam, P. Staiger, G. Dingle, M. Savic, R. Bathish, J. Mackenzie, M. Beckwith and D. I. Lubman (2016). "Social networks and recovery (SONAR): characteristics of a longitudinal outcome study in five therapeutic communities in Australia." Therapeutic Communities **37**(3): 131-139.

Cisler, R. A., R. K. Kowalchuk, S. M. Saunders, A. Zweben and H. Q. Trinh (2005). "Applying clinical significance methodology to alcoholism treatment trials: determining recovery outcome status with individual- and population-based measures." Alcoholism: Clinical and Experimental Research **29**(11): 1991-2000.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences Lawrence Erlbaum Associates.

Cornford, C. S., H. J. Close, R. Bray, D. Beere and J. M. Mason (2015). "Contraceptive Use and Pregnancy Outcomes among Opioid Drug-Using Women: A Retrospective Cohort Study." PLoS One **10**(3).

Crosby, R. D., R. L. Kolotkin and G. R. Williams (2003). "Defining clinically meaningful change in health-related quality of life." Journal of Clinical Epidemiology **56**(5): 395-407.

Dalton, M., K. F. Crowley, J. W. J. Crouch and S. E. Kelly (2017). ""Check in, check up": an evaluation of the impact of post-treatment follow-ups on substance users' recovery." Journal of Substance Use **22**(3): 260-264.

Deacon, R. M., R. Bruno, K. Mammen, J. Holmes, A. J. Dunlop, L. Mills, E. Murray, M. Jefferies, C. Rivas, M. C. Schulz, M. J. Luksza and N. Lintzeris (2018). "Concurrent validity, test-retest reliability and inter-rater reliability of the Australian Treatment Outcomes Profile among alcohol, cannabis and opioid treatment clients." Drug and Alcohol Review **37**.

Delgadillo, J., S. Gore, S. Ali, D. Ekers, S. Gilbody, G. Gilchrist, D. McMillan and E. Hughes (2015). "Feasibility Randomized Controlled Trial of Cognitive and

Behavioral Interventions for Depression Symptoms in Patients Accessing Drug and Alcohol Treatment." Journal of Substance Abuse Treatment **55**: 6-14.

Dingle, G. A., C. Haslam, D. Best, G. Chan, P. K. Staiger, M. Savic, M. Beckwith, J. Mackenzie, R. Bathish and D. I. Lubman (2019). "Social identity differentiation predicts commitment to sobriety and wellbeing in residents of therapeutic communities." Social Science & Medicine **237**: 112459.

Dore, G., B. Sinclair and R. Murray (2015). "Treatment Resistant and Resistant to Treatment? Evaluation of 40 Alcohol Dependent Patients Admitted for Involuntary Treatment." Alcohol and Alcoholism **51**(3): 291-295.

Eastwood, B., A. Peacock, T. Millar, A. Jones, J. Knight, P. Horgan, T. Lowden, P. Willey and J. Marsden (2018). "Effectiveness of inpatient withdrawal and residential rehabilitation interventions for alcohol use disorder: A national observational, cohort study in England." Journal of Substance Abuse Treatment **88**: 1-8.

Food and Drug Administration, U.S. Department of Health and Human Services (2019). Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims

Guy, W. (1976). ECDEU Assessment Manual for Psychopharmacology —Revised. Rockville, MD.

Jacobson, N. S., W. C. Follette and D. Revenstorf (1984). "Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance." Behavior Therapy **15**(4): 336-352.

Jacobson, N. S. and P. Truax (1991). "Clinical Significance: A Statistical Approach to Defining Meaningful Change in Psychotherapy Research." Journal of Consulting & Clinical Psychology **59**(1): 12-19.

Karakoula, P., K. Kokkolis, C. Tsatsaroni, A. Kitsiou, E. Koukidis, P. Danousi, S. Nikolaidi, G. Kolios, G. Vlachou, A. Fragkouli and N. Lintzeris (2021). "Cultural Adaptation and Psychometric Validation of the Australian Treatment Outcomes Profile for the Greek Population." International Journal of Mental Health and Addiction.

Kovac, G., Y. Abbasi and D. Owens (2015). "A retrospective outcomes evaluation of cannabis use at an addictions center." Drugs and Alcohol Today **15**(1): 2-8.

Larance, B., M. A.-O. Byrne, N. Lintzeris, S. A.-O. Nielsen, J. Grebely, L. A.-O. Degenhardt, J. Shahbazi, M. A.-O. Shanahan, K. Lancaster, G. Dore, R. Ali and M. Farrell (2020). "Open-label, multicentre, single-arm trial of monthly injections of depot buprenorphine in people with opioid dependence: protocol for the CoLAB study." BMJ Open **10**(7): e034389.

Lintzeris, N., K. Mammen, J. Holmes, R. Deacon, L. Mills, E. Black, L. Gardner and A. Dunlop (2020). Australian Treatment Outcomes Profile (ATOP) Manual 1: Using the ATOP with Individual clients. Sydney, Australia, AoD Treatment Clinical Outcomes and Quality Indicators Program.

Luty, J., S. Varughese and J. Easow (2009). "Criminally invalid: the treatment outcome profile form for substance misuse." Psychiatric Bulletin **33**(11): 404-406.

Maisto, S. A., N. A. Kaczynski and R. T. Ammerman (1996). "Analysis of change in adolescents involved in alcohol and drug treatment." Experimental and Clinical Psychopharmacology **4**(3): 322-329.

Mammen, K., L. Mills, R. M. Deacon, R. Bruno, A. Dunlop, J. Holmes, J. Luksza, A. Shakeshaft, M. Farrell and N. Lintzeris (2021). "Determining Clinical Cutoff Scores for the Australian Treatment Outcomes Profile Psychological Health,

Physical Health, and Quality of Life Questions." Drug and Alcohol Review(accepted).

Marsden, J., B. Eastwood, R. Ali, P. Burkinshaw, G. Chohan, A. Copello, D. Burn, M. Kelleher, L. Mitcheson, S. Taylor, N. Wilson, C. Whiteley and E. Day (2014). "Development of the Addiction Dimensions for Assessment and Personalised Treatment (ADAPT)." Drug and Alcohol Dependence **139**: 121-131.

Marsden, J., B. Eastwood, C. Bradbury, A. Dale-Perera, M. Farrell, P. Hammond, J. Knight, K. Randhawa and C. Wright (2009). "Effectiveness of community treatments for heroin and crack cocaine addiction in England: a prospective, in-treatment cohort study." **374**(9697): 1262-1270.

Marsden, J., B. Eastwood, H. Jones, C. Bradbury, M. Hickman, J. Knight, K. Randhawa, M. White and G. the National Drug Treatment Monitoring System Outcomes Study (2012). "Risk adjustment of heroin treatment outcomes for comparative performance assessment in England." Addiction **107**(12): 2161-2172.

Marsden, J., B. Eastwood, C. Wright, C. Bradbury, J. Knight and P. Hammond (2011). "How best to measure change in evaluations of treatment for substance use disorder." Addiction **106**(2): 294-302.

Marsden, J., M. Farrell, C. Bradbury, A. Dale-Perera, B. Eastwood, M. Roxburgh and S. Taylor (2008). "Development of the treatment outcomes profile." Addiction **103**(9): 1450-1460.

McElreath, R. (2018). Statistical rethinking: A Bayesian course with examples in R and Stan, Chapman and Hall/CRC.

Monds, L. A., N. J. Ridley, C. Rivas, A. Withall, B. Draper and N. Lintzeris (2017). "Cognition and adaptive functioning in older people attending drug and alcohol services." International Psychogeriatrics **29**(5): 815-823.

NSW Ministry of Health (2016). NSW Health Analytics Framework. Sydney.

Reilly, R., H. Wand, R. McKetin, B. Quinn, N. Ezard, A. Dunlop, K. Conigrave, C. Treloar, Y. Roe, D. Gray, J. Stephens, J. Ward, N. C. R. on the behalf of the and S. Partner (2020). "Survey methods and characteristics of a sample of Aboriginal and Torres Strait Islander and non-Indigenous people who have recently used methamphetamine: the NIMAC survey." Drug and Alcohol Review **39**(6): 646-655.

Roberts, L. J., D. J. Neal, D. R. Kivlahan, J. S. Baer and G. A. Marlatt (2000). "Individual drinking changes following a brief intervention among college students: Clinical significance in an indicated preventive context." Journal of Consulting and Clinical Psychology **68**(3): 500-505.

Savic, M., S. F. Barker, D. Best and D. I. Lubman (2014). "Alcohol problems among migrants in substance use treatment: the role of drinking patterns in countries of birth." Australian Journal of Primary Health **20**(3): 220-221.

Speer, D. C. (1992). "Clinically Significant Change: Jacobson and Truax (1991) Revisited." Journal of Consulting & Clinical Psychology **60**(3): 402-408.

Storen, R. and N. Corrigan. (2020). "COVID-19: a chronology of state and territory government announcements (up until 30 June 2020)." Retrieved 30/07/2021, from https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/rp/rp2021/Chronologies/COVID-19StateTerritoryGovernmentAnnouncements.