STATISTICAL ANALYSIS PLAN

FOR

EXPO – PHASE 1

Date: 08/06/2022

Version: 4

North Wales Organisation for Randomised Trials in Health (NWORTH)

Bangor University, College of Health & Behavioural Sciences,

Institute of Medical & Social Care Research, Y Wern, Normal Site, Holyhead Road, Bangor, Gwynedd LL57 2PZ

Telephone: 01248 388095 Email: nworth@bangor.ac.uk Website: http://nworth-ctu.bangor.ac.uk/

Extended-release pharmacotherapy for opioid use disorder (EXPO): Protocol for an open-label randomised controlled trial of injectable maintenance buprenorphine with personalised psychosocial intervention. The EXPO study.

IRAS Project ID: 255522

REC: London – Brighton and Sussex

EudraCT: 2018-004460-63

Based on protocol version: V5.1 (06/10/2021)

Chief Investigator:	Mike Kelleher
Co-Investigators/Co-Applicants	
	John Marsden
	Luke Mitcheson
	Caroline Murphy
	Joanna Kelly
	Eilish Gilvarry
	Ed Day
	Jonathan Dewhurst
	Fiona Cowden
Trial Manager	Jatinder Bisla
Senior Trial Statistician	Zoë Hoare
Trial Statistician	Rachel Evans

Approval signatures

Rachel Evans Author's Name	Date and Signature	08/06/2022
Zoë Hoare Second Statistical Advisor	Date and Signature	ZHoare 08/06/2022
Mike Kelleher Chief Investigator	Date and Signature	8/8/22

Professor Tim Millar

DMC Chair

12/6/22

Date and Signature

Professor Richard Holland

TSC Chair

10/6/2022

Date and Signature

DOCUMENT HISTORY

Updated version no.	Effective date	Authorship	Section changed	Summary/justification of changes
1	*n.a	Rachel Evans	n.a	New document
2	15.12.2020	Rachel Evans	Changes made following comments from TSC Chair	
			Section 4.6	PO missing data clarification
			Section 6.8	Subgroup analysis added
			Section 6.9	Sensitivity analysis for PO missing data added and months in treatment being current episode clarified
			Section 5.2	Mediator analysis models clarified
			Section 5, 5.2 & 6	Minor typos amended
			Several	References for PO reliability added, and subsequent numbering of references updated
3	18.05.2022	Rachel Evans	Changes made following discovery of protocol violations (ineligible participants), missing measures (CUD and exit interviews) and missing data review highlighting administration error of final visit time.	
			Summary page	Protocol version update
			Several	CUD – Cocaine use disorder added
			Section 3.1	Edits to which SAPs cover what study phase analysis
			Section 3.1	Handling of eligibility violations updated
			Section 4.3	Further clarification on adjustments for multiple comparisons added
			Section 4.5	Ineligible participants being randomised added as an example of a protocol violation
			Section 4.6	Clarification added to highlight that missing UDS not resulting in overall missing data for the primary outcome
			Section 5.1	Table updated in line with latest protocol version
			Section 5.2	Text added to summarise missing data due to administration error (i.e. missing days off end of some participant TLFB due to final visit

			being conducted early) and use of medical notes and rule for where this has occurred
		Section 5.2 other measures	PHQ-15, PHQ4 and OSTQOL added for exit interviews
		Section 5.3	Update to text to clarify safety data being reported in final analysis report
		Section 6	Analysis timelines updated
		Section 6.8	Detail on sub group analysis added in Updates to sensitivity analysis; Addition of UDS missing data sensitivity analysis,
		Section 6.9	Clarification of per protocol and mITT and addition of 14-day TLFB truncated analysis for administration error.
		Acronym table and throughout document	Addition of sensitivity analysis for Grace period XR-Bup amended to BUP-XR Met amended to MET Bup amended to BUP-SL Esp emended to ESP
		Section 3.1	Background text updated to align with latest protocol
		Appendix 2	Adjustments to scoring of CEQ-F, WSAS and SURE to align with correct scoring.
		Appendix 3	Addition of MOCA scoring
4	Rachel Evans	6.1 Analysis timeframe	Update to planned analysis timelines
		5.2 Definitions and calculations of outcome measures	Clarification that objective 3, measure 3 will not be conducted as part of the analysis within this SAP
		6.9 Sensitivity analyses or model testing	Table 3 dates of UDS collection updated Update of grace period sensitivity from day 0 to 161 to day 0 to 168

	4.6 Missing Data	Addition of sentence to clarify item imputation for continuous outcome measures. MOCA removed as potential predictor of missingness as measure was unobtainable during study due to COVID. Other minor clarifications.

*Version 1 signed off by Author, Principal Statistician, CI and TSC chair. Updates occurred before DMEC chair final sign off.

TABLE OF CONTENTS

1. ACRONYMS AND DEFINITION OF TERMS

Acronym	Meaning
ADAPT	Addiction Dimensions for Assessment and Personalised Treatment
AD-SUS	Alcohol and Drug Service Use Schedule
AE	Adverse Event
ALC-QFM	Alcohol – Quantity, Frequency and Maximum Consumption (ALC-QFM)
AR	Adverse Reaction
ARC	Addiction recovery clinic
BUP-SL	BUP-sublingual
BUP-SL-NX	BUP-Sublingual-Naloxone (Suboxone®)
BUP-XR	BUP- Extended-Release (US Sublocade®; prev. RBP-60000)
CEQ-F – C	Craving Experience Questionnaire (frequency version) (Cocaine)
CEQ-F – H	Craving Experience Questionnaire (frequency version) (Heroin)
CGI-I	Clinical Global Impression (improvement)
CGI-S	Clinical Global Impression (severity)
CI	Chief Investigator
CONSORT	Consolidated Standards of Reporting Trials
CRA	Clinical Research Associate
CTIMP	Clinical Trial of Investigational Medicinal Product
CTU	Clinical Trials Unit
CUD	Cocaine Use Disorder
DERS-SF	Difficulties in Emotion Regulation Scale – Short Form

DSM5	Diagnostic and Statistical Manual of Mental Disorders 5th Edition
DMC	Data Monitoring Committee
eCRF	Electronic case report form
EDC	Electronic data capture
ESP	ESPranor®
EQ-5D-5L	EurolQol Health Status (5 level)
EXPO	Extended-release Pharmacotherapy for Opioid Use Disorder
GCP	Good clinical practice
GLM	Generalised Linear Model
HEAP	Health Economics Analysis Plan
IMP	Investigational Medicinal Product
ITT	Intention to Treat
KCF	Keyworker Contact Form
LFT	Liver Function Tests (AST, ALT, albumin and bilirubin)
MET	METhadone (oral)
mITT	modified Intention to Treat
MoCA	Montreal Cognitive Assessment
NHS	National Health Service
NWORTH	North Wales Organisation for Randomised Trials in Health
OUD	Opioid Use Disorder
PRO-I	Patient Reported Outcome (improvement)
PRO-S	Patient Reported Outcome (severity)
PSI	Personalised Psychosocial Intervention
QIDS-SR	Quick Inventory of Depressive Symptomatology
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SAR	Serious Adverse Reaction
SCID-5-RV	Standard Clinical Interview for Dependence – Research Version (DSM-5)
SD	Standard Deviation
SURE	Substance Use Recovery Evaluator
SUSAR	Suspected Unexpected Serious Adverse Reaction
TLFB	Timeline Follow Back
TMG	Trial Management Group
TOP	Treatment Outcomes Profile
TSC	Trial Steering Committee
UDS	Urine drug Screen
USA	United States of America
VAS-N (H/C)	Visual Analogue Craving Rating (Need) (heroine/cocaine)
VAS-W (H/C)	Visual Analogue Craving Rating (Want) (heroine/cocaine)
VIF	Variance Inflation factor
WSAS	Work and Social Adjustment Scale

2. STATISTICAL ANALYSIS PLAN AUTHORSHIP

This SAP has been authored by Rachel Evans, Trial Statistician for the EXPO study, (Senior statistician NWORTH CTU) with input from CI, Mike Kelleher, Co-lead Investigator, John Marsden, Joanna Kelly and Caroline Murphy. Jatinder Bisla, EXPO Trial Manager and Zoë Hoare, Senior Trial Statistician for the trial (Principal Statistician NWORTH CTU).

Zoë Hoare (ZH Senior Trial Statistician) will remain blind throughout the conduct of the trial until the main analysis has been completed. Rachel Evans (RE) will be unblind for the purposes of reporting to the DMC. The draft plan has been circulated to the DMC members and TSC members for input and review prior to sign off.

Statistician Blinding

Upon review of the protocol, it was noted that group numbers have been listed since protocol version 1.1. The Senior Statistician had reviewed and approved these protocols and the use of these group numbers could potentially unblind them to the treatment groups. These codes appear to be unblind with group 1 and 2 relating to phase 1 and groups 3 and 4 relating to phase 2.

It has not been confirmed either way which group these relate to explicitly however to ensure that the Senior Statistician remains blind to the data for analysis, the Trial Statistician will re-label the groups prior to analysis. The data using this relabelling will be used by the Senior Statistician to ensure blinding is maintained. All work will follow a full audit trail and documented in line with NWORTHs SOPs and MHRA guidance.

3. INTRODUCTION

3.1 BACKGROUND AND DESIGN

Background

Daily doses of sublingual (tablet) buprenorphine hydrochloride (BUP-SL) and oral (liquid) METhadone hydrochloride (MET) are the first-line, standard-of-care (SOC) maintenance pharmacotherapies for OUD. MET is an opioid agonist with actions predominantly at the endogenous μ-opioid receptor. BUP is an opioid partial agonist/antagonist with actions predominantly at the endogenous μ-opioid and k (kappa) opioid receptors. In the UK, two other licensed BUP medications are also available for OUD – buprenorphine-naloxone (buprenorphine hydrochloride-naloxone dihydrate; Suboxone®; sublingual tablet; BUP-NX)

and buprenorphine-lyophilisate; ESPranor ®; sublingual wafer; ESP). BUP-NX contains the opioid antagonist naloxone (1:4 ratio with BUP) as a deterrent to injection of non-medical opioids. Indivior developed a subcutaneously injected, extended-release formulation of BUP (RBP6000). RBP6000 releases BUP for a minimum of 28 days, thereby facilitating monthly maintenance dosing. RBP6000 is now licensed as Sublocade® in the USA (BUP-XR herein).

Design

The EXPO study will determine the effectiveness of BUP-XR in a head-to-head superiority comparison with BUP-SL/MET (phase 1) and with an adjunctive psychosocial intervention (PSI) (phase 2) in the pragmatic context of specialist community setting addiction treatment programmes in England and Scotland provided by the NHS.

The Trial is a Phase III (CTIMP; IMP open-label, oral IMP comparators [no placebo]), multi-centre, pragmatic, two-phase, randomised controlled trial. With a target total sample of 604 patients with Opioid Use Disorder (OUD; moderate-severe; DSM5), participants are allocated to one of four trial arms:

- BUP-XR
- BUP-SL/MET (including BUP-SL-NX and ESP)
- BUP-XR + PSI
- BUP-SL/MET (including BUP-SL-NX and ESP) + PSI

As BUP is the active component of BUP-SL-NX and ESP they will be categorised under BUP-SL within the remainder of this SAP and for analysis between trial arms.

Treatments

BUP extended-release depot injection (USA Sublocade®; BUP-XR) is administered by subcutaneous injection into abdominal tissue. BUP-XR is available as a 100mg/0.5ml and 300mg/1.5ml pre-filled syringe. In EXPO, the drug is administered as two initial doses of 300mg given a minimum of 21 days apart, followed by 4 x 100 mg monthly maintenance doses.

Within the regimen and if the benefits are judged to outweigh the risks: (1) The 100 mg maintenance dose may be increased and maintained to 300mg if there is unsatisfactory clinical response (i.e. dose 3-6); (2) A 300 mg maintenance dose (i.e. dose 4 and 5) may be

reduced to 100 mg according to clinical assessment; (3) BUP-SL (including BUP-SL-NX and ESP) will be given on a rescue basis for patients who do not achieve a satisfactory clinical response to BUP-XR.

Active comparator product(s);

(1) BUP-SL (sublingual [tablet] BUP-SL), or substitute 4:1 ratio BUP-SL/naloxone 'Suboxone' [BUP-SL-NX]) usual dose 12-24mg/day; or substitute oral lyophilisate ESPranor (ESP) usual dose 9-18mg/day) are longstanding standard-of-care maintenance pharmacotherapy for OUD in the NHS;

(2) MET (oral methadone hydrochloride; usual dose 60-120mg/d), longstanding standard-of-care maintenance pharmacotherapy for OUD in the NHS.

Trial Phases

The study consists of two phases. Phase 1 of the study refers to the analysis of BUP-XR and BUP-SL/MET and phase 2 refers to the analysis of BUP-XR + PSI and BUP-SL/MET + PSI. There will be separate statistical analysis plans for the two study phases. Recruitment in phase 1 of the study is on a 1:1 allocation ratio to BUP-XR and BUP-SL/MET across 4 clinical recruitment sites in the West Midlands, Greater Manchester, Newcastle and Tayside. During phase 1, the South London site will recruit to groups on a 4:4:1:1 allocation ratio for approximately 6 months whilst resources are in place to allow more participants to be allocated to psychosocial therapy (e.g. therapists in post). During the subsequent approximate 18 months of recruitment, there will be a 1:1 allocation ratio to BUP-XR + PSI and MET/BUP-SL + PSI (only at the South London site). The timescales for these ratios may change depending on recruitment levels and resource availability. Figure 1 below represents the study's randomisation structure across sites. The current statistical analysis plan (SAP) only concerns phase 1 of the study.

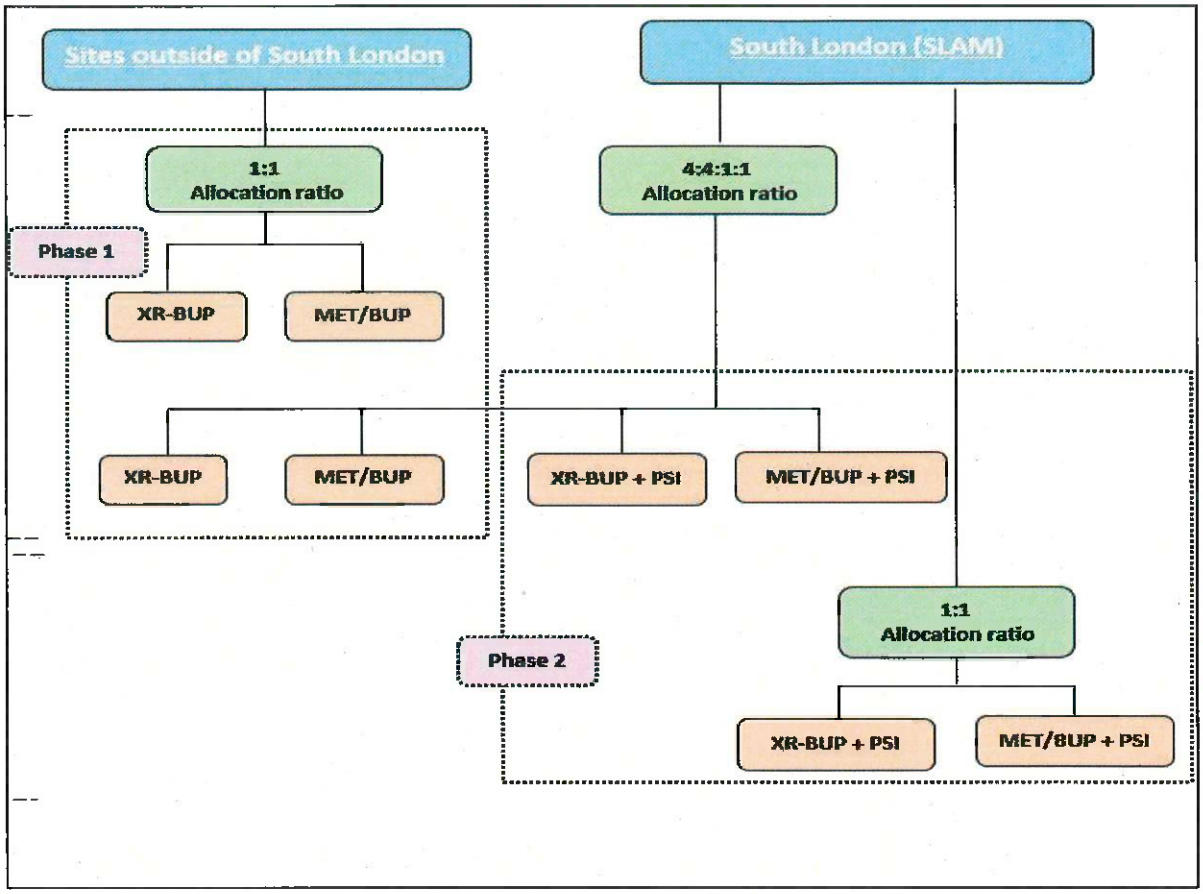


Figure 1: Randomisation summary for Phase 1 and Phase 2 of EXPO study

Analysis set

Primary analysis will be conducted on an intention to treat (ITT) basis, defined as all participants that are randomised are analysed as allocated, regardless of their treatment.

Certain circumstances may occur that result in randomised participants not being analysed. Such circumstances can arise if patients are randomised in error, e.g.

- If the same participant is mistakenly randomised twice
- If an ineligible patient is randomised and must be withdrawn from the study Eligibility violations

If any post randomisation eligibility violations are identified, the study team along with the sponsor and independent committees will establish whether the participant can remain in the study. For cases where the criteria violated does not affect clinical outcome or patient safety these participants will be kept in the ITT analysis set. Sensitivity analysis will be conducted

removing eligibility violations by conducting a modified intention to treat (mITT) analysis, see section 6.9 for more details.

All exclusions due to randomisation errors or eligibility violations will be reported clearly.

3.2 TRIAL OBJECTIVES

Primary Objective:

During six months of active trial treatment, the primary objective of this superiority study is to determine the difference in effectiveness (in terms of reduced use of heroin and illicit opioids) of BUP-XR and either MET or BUP-SL. Study participation will be offered to patients already enrolled in BUP-SL and MET maintenance treatment.

With a 1-week measurement grace period from randomisation, the primary endpoint is the count of days abstinent from heroin and illicit opioids during six months of treatment combined with urine drug screen samples which are negative for non-medical opioids.

Secondary Objectives:

To explore study group differences on the following measures:

(1) *Clinical superiority of BUP-XR plus PSI versus BUP-SL/MET plus PSI;

(2) Safety of BUP-XR;

(3) Treatment retention;

(4) OUD and CUD remission status;

(5) Clinician rated impression of patient response to treatment;

(6) Patient reported recovery outcomes;

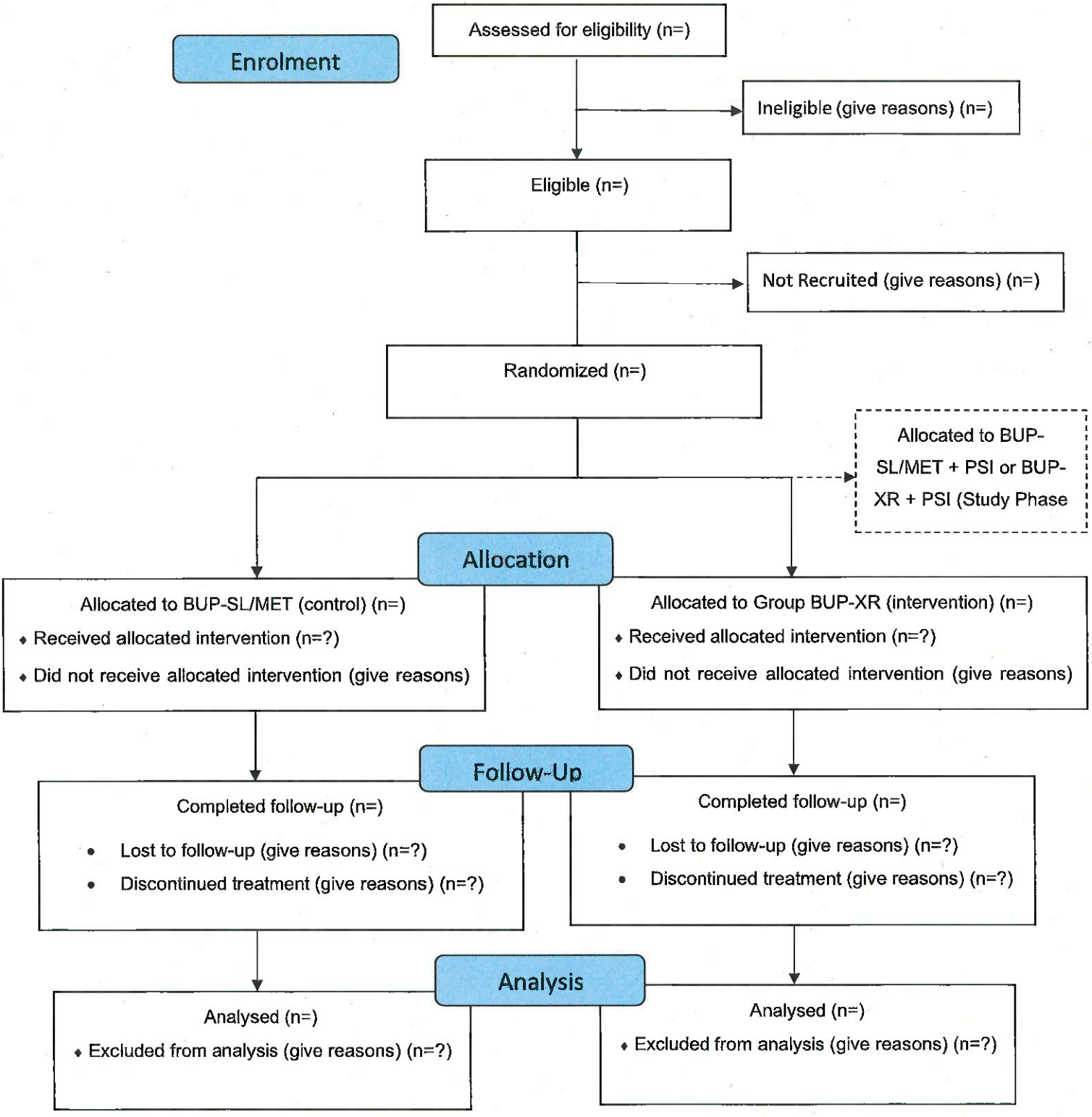
(7) Mediation of craving, depression, work and social adjustment, and cognitive function on outcome.

* Secondary Objective (1) not covered by this SAP as is part of phase 2 analysis.

Additionally, longer-term outcomes will be collected through data-registry indicators of criminal convictions and health service utilisation. The analysis of these outcomes are not included in this SAP.

3.3 CONSORT DIAGRAM

CONSORT 2010 Flow Diagram



4. STATISTICAL PRINCIPLES

4.1 SAMPLE SIZE JUSTIFICATION

The required number of participants has been estimated using that required for a Poisson regression model taking a baseline rate of 0.6 (60% abstinent days per month) and assuming a 23% target difference in the count of days of abstinence from heroin over the study period of 161 days. This target difference is informed by results of the ARC trial (ISRCTN69313751): 56.78% for the treatment-as-usual control (i.e. assuming 17 abstinent days over a period of 28 days).

To obtain 90% power, for a Poisson regression, with alpha at 5% and with 15% inflation for attrition, EXPO has a target total of 304 participants for the statistical comparison of the primary outcome recorded for BUP-XR and MET/BUP-SL.

A sensitivity check on the power calculation on the assumption of a greater response has been completed. With a sample of 304 subjects and a 23% target difference achieved, a higher baseline rate of 0.75 observed would achieve 97.5% statistical power.

Additionally, a sensitivity check has been conducted assuming a reduced baseline rate. With a sample of 304 subjects and a 23% target difference achieved, a baseline rate of 0.5 observed would achieve 88% statistical power.

A power analysis will be completed as part of the main analysis for the study.

4.2 RANDOMISATION

Online randomisation system will be provided by King's CTU. Randomisation procedure will be on a 1:1 allocation ratio by stratified block with random varying block sizes, stratified for Centre (West Midlands, Manchester, Newcastle and Tayside) and Current (last 28 days) drug injecting status (yes/no). At South London for approximately the first 6 months (subject to change depending on resource and recruitment rates) of recruitment, patients will be randomised to one of four groups on a 4:4:1:1 allocation ratio, with Current (last 28 days) drug injecting status (yes/no). Figure 1 in section 3.1 portrays this.

Justification for these stratification factors is as follows: (a) site is required for the random-effects model; (b) it is expected that there will be relatively poorer outcome among participants who are drug injectors. The randomisation procedure will use stratified random blocks of varying size to ensure even allocation.

4.3 LEVELS OF CONFIDENCE AND P-VALUES

All statistical tests for the primary endpoints will be two-sided and will be performed using a 5% significance level. Confidence intervals for estimated effects will be presented as 95% and two-sided.

Secondary outcomes will not be adjusted for multiple comparisons, as standard. The primary outcome is powered for at the significance level stated. Analysis of all secondary outcomes is considered exploratory, and any statements of results related to these outcomes will be reported appropriately. 95% confidence intervals will be reported along with the p-value of the effect.

4.4 ADHERENCE

BUP-SL (including BUP-SL-NX and ESP) (active sublingual comparator): The usual dose is 12-24mg/day (ESP usual dose 8-18mg/day). Clinical practice in NHS clinics is supervised (directly observed dosing) in a community pharmacy followed by provision of patient self-administered 'take home' doses according to clinical response (i.e. informed by negative urine drug screens [UDS]; adherence to prescription; retention in treatment).

MET (active oral comparator): The usual dose is 60-120mg/day. Clinical practice in NHS clinics is supervised (directly observed dosing) in a community pharmacy followed by provision of patient self-administered 'take home' doses according to clinical response (i.e. informed by negative UDS; adherence to prescription; retention in treatment).

Adherence to the intervention will be reported descriptively using the BUP-XR Injection Log and for the comparator the BUP-SL/MET Medication Record form and UDS. This will include injection frequencies, patterns of dosage (i.e. injection size in mg), oral medication frequencies and patterns of dosages.

4.5 PROTOCOL VIOLATIONS

Definition of protocol violations and deviations. Violation is an intended failure to adhere to the protocol such as wrong treatment being prescribed or administered or incorrect data being collected and documented. A protocol deviation is an unintended failure to adhere to the protocol and examples include errors in applying inclusion/exclusion criteria or missed follow-up visits due to error. Per protocol analysis is not planned for the study but deviations and violations will be monitored and reported throughout the study and detailed descriptively in the analysis report. If protocol deviation occurrences rise above 10% then sensitivity analysis as detailed in section 6.5 will be conducted. Some examples of protocol deviations which might occur in this study include;

- Patients not receiving randomised allocation
- Patients not receiving treatment within 7 days
- Ineligible participants being randomised

4.6 MISSING DATA

Primary and secondary count-based measures For the count-based measures two situations of missing are expected. Firstly, there will be those who have missing observations periodically through the study observation period but have remained in the study for these patients multiple imputation as described above will be adopted. The memory recall for the TLFB is 3 months therefore if a patient hasn't been to visits for a couple of months recall of up to 3 months is deemed reliable [4 - 6]. Where patients have gaps in their recall on the primary outcome (at most this would be 3 months) we would use the available data to impute for the patient. Sensitivity analysis will be conducted on this imputation using a best-case worst-case analysis, see section 6.9 for more details.

Secondly, there will be some cases of participants who completely drop out of the study during the observation period i.e. withdrawals or loss to follow ups. For these patients on the primary outcome and secondary count-based measures an offset (censoring) will be used to represent different exposures rather than multiple imputation.

Furthermore, some of the primary and secondary end points (TLFB) are based on self-report and UDS tests. Where data is available for self-report and UDS, if the results are inconsistent the UDS data will override the self-report, as indicated in section 5.2. Essentially where UDS is not available this does not result directly in missing data for the outcome as the self-report

will provide this data. However, the reliability of the self-report may be impacted therefore sensitivity analysis on missing UDS data will be conducted, see section 6.9 for more detail.

Continuous outcome measures

Missing observations are expected within the dataset. For missing items within an outcome measure, the published rules for completing missing data for the relevant measure will be applied. Where there are no missing data rules for the measure, if the number of missing items on an outcome is 20% or less, then the missing value for the item will be substituted by the individual's mean score for the remaining items on the scale (Bono, Ried, Kimberlin and Vogel 2007). If there are more than 20% missing items in the scale the outcome measure will not be calculated for the participant at that time point and multiple imputation methods will be used.

Predictive mean matching multiple imputation methods will be adopted for an outcome measure total score if missing data arises above 5%. For multiple imputations, the number of imputations completed will be dependent upon the percentage of missing data [2]. The missing outcome measures will be imputed using group allocation and stratification variables along with any baseline characteristics that are deemed to be predictors of missingness. Baseline characteristics will be assessed for being predictors of missingness by running statistical tests on completers versus non completers and evaluating if any differences are present between the two.

Baseline factors to be assessed as predictors of missingness include;

- Centre
- Current injecting status
- Length of current treatment episode
- Preference of MET or BUP-SL
- Use of cocaine in the past month
- Sex at birth
- Age
- *Ethnicity
- Baseline outcome scores on the following measures:
    - QIDS-SR
    - WSAS
    - PRO-S

- ALC-QFM
- ADAPT (all subscales)

*Ethnicity is being collected with five categories - Black, White, Mixed, Asian and other. Depending on the prevalence of responses in each of these categories, the variable will be treated either with all 5 categories as collected or if appropriate will be dichotomised into 'White' and 'Non-White' if the range across categories are low.

4.7 ASSUMPTION CHECKING

All assumptions relating to the models will be checked and evaluated whether appropriate to use with the data. If any of the assumptions are substantially violated then appropriate non-parametric tests will be conducted. Appendix 4 contains details of the assumptions associated with each model and the methods to be used to assess these assumptions.

Outliers or unusual values will be assessed by running Grubbs [7] test for outliers and visually inspecting a boxplot. No outliers will be discarded if they are within plausible range. Primary analysis will be conducted keeping the outliers in the dataset and if necessary, sensitivity analysis will be conducted by removing the outliers and evaluating any effects on the results and conclusions of analysis. Any outliers removed will be fully reported.

The distribution of the data will be checked. For the count data a decision will need to be made as to which analysis method to use for the outcome based on the distribution of the data. Appendix 1 contains a flow diagram to be followed to decide which analysis model is to be run based on the data distribution.

For continuous data a decision will be made as to whether a transformation should be applied to the data and if so which transformation should be used. If transformation is required the distribution of the transformed data will be checked. Analysis will be reported on the original scale, transforming variables back. If a transformation is inappropriate then non parametric analysis methods will be considered.

For the logistic regressions, if detection rates (number of responses observed) are very low, where appropriate, either a Firth's Bias-Reduced Logistic Regression will be conducted or descriptive statistics reported.

5. DATA

KCTU MACRO database;

All baseline and follow-up data will be entered on the online InferMed MACRO electronic data capture (EDC) system (infermed.com). This system is regulatory compliant (GCP, and the EC Clinical Trial Directive). An electronic case report form (eCRF) using the MACRO EDC will be programmed by the KCTU and hosted on a dedicated secure server. The eCRF system will have full audit trail, data discrepancy functionality, database lock functionality, and supports real time data cleaning and reporting. The Trial Manager will request usernames and passwords to any new researchers. Only those authorised by the Trial Manager will be able to use the system.

After written recording, each research worker will aim to transcribe data onto the eCRF within one working week of a participant assessment. After completion of all follow-ups and prompt entry of data, the Clinical Research Associate (CRA) will review the data and resolve queries with sites. At the end of the trial, each centre will be supplied with their eCRF data for storage. This will be filed locally for any future regulatory or internal audit.

Database lock for phase I will be undertaken within approximately 3 months of last patient, last visit and will be controlled by TMG of participants enrolled in groups 1 and 2. Database lock for phase II will be within approximately 3 months of last patient, last visit and will be controlled by TMG of participants enrolled in groups 3 and 4. This will be once the data has been monitored, all queries resolved and PI sign off has been obtained for participants in the BUP-XR and BUP-SL/MET (including BUP-SL-NX and ESP) arms.

KCTU Randomisation System;

The King's Clinical Trials Unit Randomisation system will be used in EXPO. This will be a secure password-protected, web-accessed system.

Once a patient has consented and it is confirmed that all potential medication is in place within the next 7 days, all baseline data collected and eligibility confirmed, the staff member will log

into the randomisation system (www.ctu.co.uk) and randomise the participant, detailing the relevant stratification factor information;

*(a) Site (South London, the West Midlands, Greater Manchester, Newcastle and Tayside)

(b) Current (last 28 days) drug injecting status (yes/no)

*Note there are four systems built for the study three for South London and one for the other sites.

1 system for other sites at 1:1 ratio XR BUP-SL and MET/BUP-SL

1 system South London at ratio 4:4:1:1 BUP-XR, MET/BUP-SL, BUP-XR + PSI, MET/BUP-SL + PSI

1 system South London at ratio 1:1:1:1 BUP-XR MET/BUP-SL, BUP-XR + PSI, MET/BUP-SL + PSI

1 system South London at ratio 1:1: BUP-XR + PSI, MET/BUP-SL + PSI

The randomisation procedure will use stratified random blocks of varying size to ensure even allocation. See section 4.2 for more details on randomisation.

5.1 TIME POINTS OF OUTCOMES MEASURES

Table 2 indicates the data and study measures to be collected during the trial and at which time points across the study they are collected.

Table 1: List of outcome measures to be collected and time point of collection

Measure	Baseline	Randomisation	Ongoing	Week 1	Week 2	Week 4	Week 6	Week 8	Week 10	Week 12	Week 14	Week 16	Week 18	Week 20	Week 22	Week 24	Withdrawal	Week 24 onwards	Ext. Phase Interview +++
Consent §	X																	X	X
Screening	X																		
UDS					X	X	X	X	X	X	X	X	X	X	X	X	X		
LFT *	X					X				X						X		X	
BUP-XR**				X		X		X		X		X		X				X	
BUP-SL/MET **			X																
ALC-QFM (TOP)***	X															X			
SCID-5-RV	X									X						X			X
ADAPT	X					X				X						X			
CGI-S	X																		
CGI-I						X				X						X			
CEQ-F – H***	X					X		X		X		X		X		X	X		
CEQ-F – C***	X					X		X		X		X		X		X	X		
VAS-N (H/C)***	X					X		X		X		X		X		X	X		
VAS-W (H/C)***	X					X		X		X		X		X		X	X		
DERS-SF	X					X				X						X			X
WSAS	X					X				X						X			
EQ-5D-5L	X									X						X			
AD-SUS	X									X						X			
KCF***						X				X						X			
QIDS-SR	X					X				X						X			
SURE						X				X						X			
PRO-S	X																		
PRO-I						X				X						X			
MoCA***	X									X									
OSTQOL																			X
PHQ-4																			X
PHQ-15																			X
Exit Interview+																X			
BUP-SL/MET + PSI Interview++																X			
TLFB	X		X																X
Con Meds****			X			X				X						X		X	
Adverse Events Log			X															X	
Research Payments *****	X 20			X 10	X 10	X 10	X 5	X 10	X 10	X 10	X 5	X 10	X 10	X 10	X 5	X 10			X 20

§ screening to include pregnancy test, demographics and medical history. Participants receiving XR- BUP-SL reconsented once at week 24 to continue receiving monthly injections;
* LFT for all participants at screening; during treatment for BUP-XR participants only under protocol;
** refer to section 6 of the protocol for guidelines;
*** used for mediation model;
**** ConMeds to be recorded continuously and reviewed at week 4, 12 and 24;
***** research payments are time offset to attend the clinic to complete research measures and cover travel (either a £20, £10, or £5 value, as detailed per visit applicable; loaded onto a prepaid card).
+ Exit Interview conducted at South London, Newcastle, West Midlands and Tayside sites with BUP-XR participants only. Participants will be reimbursed an additional £20 for their time in taking part in the interview (loaded onto a prepaid card).
++ BUP-SL/MET +PSI Interview conducted at South London with BUP-SL/MET +PSI participants only. Participants will be reimbursed an additional £20 for their time in taking part in the interview (loaded onto a prepaid card).
+++ Extension Interview conducted at the South London and Newcastle sites with BUP-XR participants who consented to continue treatment after week 24 between 12-24 months of them being randomised in the trial.

5.2 DEFINITIONS AND CALCULATIONS OF OUTCOME MEASURES

The following outcomes measures have been identified in relation to the objectives as described in section 3.2

Primary objective;

Objective 1 - effectiveness of BUP-XR and MET/BUP-SL

Measure 1; Number of days abstinent from 'substance' during days 8-168 (weeks 2 to 24, 161 days); combining self-report information from time-line follow-back interview adapted from the Treatment Outcomes Profile and incorporating data from 12 scheduled urine drug screens (UDS). The UDS results will provide biological verification for up to 36 out of the 161 days required.

- Days abstinent from illicit/non-medical opioids (Primary outcome)
- Days abstinent from cocaine (secondary outcome)
- Days abstinent from benzodiazepines (secondary outcome)

The primary outcome will be determined by combining the self-report results and UDS results to give an overall count for each participant. Where data is available for self-report and UDS, if the results are inconsistent the UDS data will override the self-report. Where this occurs, the result will overrule the self-report for the day of the UDS and the two days prior. Examples are indicated below.

Example 1

A negative self-report for 4 days, day 4 a positive UDS, figure 2 below.

	Day 1	day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10
Self-report	no drug use	no drug use	no drug use	no drug use	drug use	drug use	drug use	drug use	drug use	drug use
UDS (Day 4)				positive						

***Figure 2**: Example 1 of Self-report and UDS results*

The UDS would trump the 'no drug use' self-report for days 2 – 4, resulting the outcome in figure 3.

	Day 1	day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10
Outcome:	no drug use	drug use	drug use	drug use	drug use	drug use	drug use	drug use	drug use	drug use

***Figure 3:** Self-report and UDS example 1 outcome*

Example 2

A positive self-report for 2 days, day 4 a negative UDS as shown in figure 4.

	Day 1	day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10
Self-report	no drug use	no drug use	drug use	drug use	drug use	drug use	drug use	drug use	drug use	drug use
UDS (Day 4)				negative						

***Figure 4:** Example 2 of Self-report and UDS results*

The negative UDS would trump the self-report for days 2 – 4 (day 2 no change) resulting in the outcome below in figure 5.

	Day 1	day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10
Outcome:	no drug use	no drug use	no drug use	no drug use	drug use	drug use	drug use	drug use	drug use	drug use

***Figure 5:** Self-report and UDS example 2 outcome*

Example 3

UDS and self-report align figure 6, therefore no change to the outcome in figure 7

	Day 1	day 2	Day 3	Day 4		Day 13	Day 14	Day 15	Day 16	Day 17
Self-report	drug use	drug use	drug use	drug use		no drug use	no drug use	no drug use	no drug use	drug use
UDS (Day 4 and day 16)				positive					negative	

***Figure 6:** Example 3 of Self-report and UDS results*

	Day 1	day 2	Day 3	Day 4		Day 13	Day 14	Day 15	Day 16	Day 17
Outcome:	drug use	drug use	drug use	drug use		no drug use	no drug use	no drug use	no drug use	drug use

***Figure 7:** Self-report and UDS example 3 outcome*

The syntax for calculating this variable will be validated and verified fully prior to database lock.

During a data review it was identified that some participants had been called in for their last visit early, resulting in the full primary outcome of 161 days not being obtained. The participants in question are not lost to follow up or withdrawals and are just missing due to an administration error and in most cases only by a day or two. It was noted that patient records could provide data for some of these missing days therefore, with input from independent committees, it was decided that for the TLFB data some of the final follow up days are being input based on these medical notes.

Where the TLFB data is not obtainable from the patient notes, the missing days will be imputed using a devised rule. The missing data rule used for these missing days will look at the last 14 days of data that are available for the participant, the percentage of drug use will be calculated. For all missing days a random number will be generated (either a 0 or 1, no use or use) using a probability based on their percentage used in the last 14 days.

Measure 2; Longest time in days ***continuously*** abstinent from 'substance' during days 8-168 (weeks 2 to 24, 161 days); combining self-report information from time-line follow-back and 12 urine drug screen (UDS) results.

- Longest count of days continuously abstinent from illicit/non-medical opioids
- Longest count of days continuously abstinent from cocaine
- Longest count of days continuously abstinent from benzodiazepines

Secondary objectives;

Objective 1 – Clinical superiority of BUP-XR plus PSI versus BUP-SL/MET plus PSI.

Will be analysed as part of phase 2 of the study and detailed in a separate SAP

Objective 2 - Safety profile

- Tabulation of AEs, ARs, SAEs, SARs and SUSARs in each group

Objective 3 - Treatment retention

- **Measure 1;** Total retention time - Days enrolled in maintenance medication for opioid use disorder (OUD) from first day following grace period (day 8) to day 168 (across 161 days weeks 2 to 24)
- **Measure 2;** Survival time - Days from randomisation to first discontinuation (i.e. time (days) enrolled in study treatment).
- ***Measure 3;** Time (days) enrolled in study treatment (retention) from week 24 to date last participant randomised plus 9 months.

*Measure 3 Will be analysed as part of the extended phase of the study and detailed in a separate SAP

Objective 4 OUD and CUD remission status

- Assessed by Standard Clinical interview for Dependence - research version (SCID-RV)
    - Remission status (Yes/No) at 12 weeks
    - Remission status (Yes/No) at 24 weeks

Objective 5 Clinician rated impression of patient response to treatment (at 24 weeks)

- **Measure 1;** Clinical Global Impression scale (severity and improvement - CQI-S/CGI-I)
- **Measure 2;** Addiction Dimensions for Assessment and Personalised Treatment (ADAPT)

Objective 6 Patient reported recovery outcomes (at 24 weeks)

Measure 1; Craving response using Craving experience questionnaire (CEQ-F)

Measure 2; Visual analogue craving rating (Need) (VAS-N)

Measure 3; Visual analogue craving rating (Want) (VAS-W)

Measure 4; Social functioning using Work and Social Adjustment Scale (WSAS)

Measure 5; Depression using Quick inventory of Depressive Symptomatology (QIDS-SR)

Measure 6; Improvement and Recovery using Substance Use Recovery Elevator (SURE)

Measure 7; Patient Reported Outcome Status (severity and improvement - PRO-S/PRO-I)

Measure 8; Emotion regulation using Difficulties in Emotion Regulation Scale short form (DERS-SF)

Measure 9; Alcohol quantity, frequency, maximum consumption (ALC-QFM (TOP) - quantity and frequency only - maximum will be reported descriptively).

Objective 7 Mediation analysis

Mediation of craving, work and social adjustment, and cognitive function on outcome.

Mediator measures;

- Visual analogue scale (VAS) for the perceived need and want for non-medical opioids and cocaine (VAS-N and VAS-W)
- Craving Experience questionnaire, frequency version (Heroin) (CEQ-F-H)
- Quick Inventory of Depressive Symptomatology – Self-Report (QIDS-SR)
- Montreal Cognitive Assessment (MoCA)
- Work and Social Adjustment Scale (WSAS)

Moderator measures;

- Sex at birth

- Age
- Current injecting status
- Centre
- Measure/Outcome baseline

Separate models will be run for each of the mediator measures using the moderator measures listed above, therefore, in total five models will be evaluated (i.e. baseline and each mediator)

Other Measures

Appendix 2 contains further details of each outcome measure and Appendix 3 contains scoring details for the relevant measures.

Many of the measures are being collected across several time points during the study, as indicated in table 2 but are only being primarily analysed here at 24 weeks. Exploratory analysis of the secondary outcomes collected at all time points other than the primary endpoint of 6 months will be detailed in a separate SAP. Additionally, the EQ5D-5L, AD-SU, PHQ-15, PHQ-4, OSTQOL and keyworker contact form are being collected but not analysed here. These measures will be used for health economic analysis and qualitative evaluation at a later date. Details on this analysis is to be included in a separate health economics analysis plan (HEAP) that will align with this SAP and signed off prior to data lock.

5.3 SAFETY DATA

All adverse events, reactions and serious and unexpected events/reactions will be reported immediately following the procedure noted in the protocol (section 11.2). Safety data in the form of all AEs will be presented to the DMC on a scheduled basis. This report will cover the number of events and number of subjects affected split by treatment arm and categorised by seriousness (not serious, death, is life-threatening, requires hospitalisation or prolongation of hospitalisation, persistent or significant disability or incapacity, consists of a congenital anomaly or birth deficit, important medical event or pregnancy). Frequencies of system organ class will also be presented split by treatment arm. No statistical testing of these data will be undertaken in the DMC report. These data will also be included in the final analysis report. Precipitated withdrawal symptoms will be included in the discussion section of the primary paper as a narrative in relation to the Adverse Events table. This will not be a formal analysis but to check if there is a correlation between the first 24-48 hours prior to treatment and first

month after treatment and numbers of precipitated withdrawal symptoms reported between the treatment arms.

6. STATISTICAL ANALYSES

6.1 ANALYSIS TIME FRAME

Table 2: Planned target timeline for analysis dependent on recruitment period

Task	*Updated following COVID recruitment pause
Last patient recruited	~October 2021
Final patient visit	~May 2022
Database Lock	~June/July 2022
Final Analysis	~August 2022

*The analysis timelines are dependent on recruitment rates, COVID-19 impacts and discussions with the funder. Delivery of the analysis report will require at least one month following database lock to complete.

6.2 BASELINE ANALYSIS

Characteristics collected at baseline will be presented descriptively both overall and by treatment arm. No statistical testing of baseline differences will be completed, only used to assess balance. Categorical variables presented using counts and percentages, continuous variables with mean, SD and range.

6.3 INTERIM ANALYSIS

There are no planned interim analyses and as such the study is not powered to achieve these. There are no planned stopping criteria based on efficacy and stopping based on safety will be guided by the DMC.

6.4 CONSORT ANALYSIS

The patient flow information, as shown in section 3.4 above as advised by CONSORT reporting standards, will be completed with values relating to participants numbers. Additionally, data will be presented on screening, eligibility, recruitment, treatment discontinuation and withdrawn/lost to follow up presented for the entire study and split by recruitment site.

Where possible reasons for ineligibility or non-recruitment will be reported. From these data related eligibility, recruitment and retention rates will be calculated and presented in the final analysis report.

6.5 DESCRIPTIVE STATISTICS

Descriptive statistics of the data will be presented in the final analysis report. This will include randomisation figures, demographics, other data characteristics of the outcome measures and data completeness levels. All will be presented overall and split by allocation group.

Figures will be presented for the Randomisation data by each stratification variable;

- Allocation group
- Centre
- Current injecting status

The following demographics and data characteristics will be presented;

- Sex at birth
- Age
- Ethnicity
- Preference of MET or BUP-SL
- Use of cocaine in the past month
- Length of treatment episode (in days)
- Age started opioid substitution therapy
- OUD medication length this episode
- Rescue medications required
- ALC-QFM (TOP) – maximum consumption

Descriptive statistics of the outcome measures will be produced for the primary and secondary variables as listed in Appendix 2.

For all statistics, continuous variables will be reported with mean values, standard deviations and ranges. Categorical variables will be presented with counts and percentages. If data are not normally distributed then medians and interquartile ranges will be reported.

Graphical representations of the objective one measures (days abstinent and longest days) will be displayed to evaluate and observe where across the study patients 'relapsed/started using again' and where withdrawals in the study were common.

In addition, we will plot a cumulative distribution function for each group on the primary outcome, other graphical representations may be used for treatment effect visualisation.

6.6 ANALYSIS OF PRIMARY OUTCOME

The primary outcome will be analysed using a Poisson regression model (i.e. generalized linear model with a log link) to assess the differences between the two treatment groups (BUP-XR or MET/BUP-SL). The randomisation stratification variables (centre and current injecting status) will be included in the model as covariates. If a Poisson distribution is not observed then the regression model to be used will be dependent on the distribution of the data and decided following the decision tree in appendix 1, figure 1.

6.7 ANALYSIS OF SECONDARY OUTCOMES

All secondary outcomes are to be analysed as the primary outcome including the randomisation stratification variables (centre and current injecting status) in the models. The model used will be based on the outcome measure (dependent variable) format and the distribution of the data.

- Count variables will be analysed using a Poisson regression model. As with the primary outcome measure the exact model to be used will be based on the distribution of the data following the decision tree in appendix 1, figure 1.
- Continuous measures will be analysed with a GLM to assess the differences between the means of the treatment groups. In addition to the randomisation stratification factors, the baseline measure scores of the associated outcome will be included in the models as covariates to account for the baseline scores.
- Binary variables will be analysed using a (binary) logistic regression model
- Ordinal variables will be analysed using an ordinal logistic regression model

Appendix 2 details the outcome measures to be analysed and their associated models.

A causal mediation analysis (including craving, MoCa, QUIDS-SR and WSAS as mediators in separate models) will be implemented using a counterfactual framework to include a treatment/mediator interaction and covariates, based on univariable path screening, and with bias corrected bootstrapped estimates.

6.8 SUBGROUP ANALYSES

Subgroup analyses will be conducted for the primary objective/outcome on the following:

- Using cocaine (Yes/No)
- Length of time in treatment (less than one month/one month or longer)
- Benzodiazepine use past month to admission (Yes/No)
- CGI-S (Mild/Severe)

6.9 SENSITIVITY ANALYSES OR MODEL TESTING

Sensitivity analysis will be conducted on the primary analysis model including the variables that were identified as predictors of missingness as covariates in the model.

Additionally, sensitivity analysis may be required around any assumptions made for the primary analysis (e.g. outliers, data distributions) but will only be conducted if necessary. This will include a best-case worst-case analysis on the self-report missing data on the primary outcome. Best case would assume that patients did not use during any missing time periods and worst case would assume that they did use substances to evaluate any impact the imputed data may have had on the results.

Furthermore, the confidence of the reliability of the self-report will be checked by conducting a best-case worst-case sensitivity analysis on the UDS missing data for the TLFB outcomes.
For those who have a missing UDS, we will assume all missing as negative for one dataset and then for another dataset assume all missing as positive. Utilising this best –worst case scenario will provide a range of values between which the real outcome is likely to lie. The width of this range is likely to be influenced by many factors (including the COVID pandemic and not being able to collect the UDS).
Primary and secondary analysis will be conducted on an ITT basis (including all those randomised to be analysed. Sensitivity analysis may be conducted, if necessary, on protocol deviations or eligibility violations by removing them from the analysis set. This includes protocol deviations (but is not limited to) such as;

- Patients not receiving randomized allocation (per protocol analysis)
- Patients not receiving treatment within 7 days (per protocol analysis)
- Ineligible patients being randomised (mITT analysis)

Sensitivity analysis will be conducted on the patient's months in treatment (for current episode) at baseline by running the model including the variable as a factor.

Additionally, sensitivity analysis of the patient's preference of BUP-SL or MET will be conducted again by running the model including that variable as a factor. This will be conducted given that there is an adequate sample of data per group at each level of the variable (BUP-SL or MET) a sample of 50% – 60% in the MET group will indicate that we will run this sensitivity analysis.

Sensitivity analysis will be conducted truncating at 14 days at the end of the TLFB data (i.e. day 8 – day 154) to assess the impacts of using the medical notes due to the administration error highlighted in section 5.2.

Furthermore, there will be a sensitivity analysis around the week 'Grace period' of the primary outcome where we will include all TLB data from day 0 to day 168.

COVID-19

A sensitivity analysis around the impact of the COVID-19 pandemic has been considered. The biggest impact the pandemic has had on the trial, to date, is the pause in recruitment which occurred March 2020 to July 2020, however, the trial is currently back recruiting in all sites. Some of the psychosocial intervention was delivered remotely (by telephone and video call) and some data collection was collected remotely during this time. The paper will include a summary of milestone dates such as the date COVID-19 was declared a pandemic by the World Health Organisation, date the Government imposed lockdown and dates the trial paused and restarted recruitment.

In addition, during the pandemic, changes to the distribution of medication for those in the oral arms of the study occurred. National clinical guidance was issued on 22 May 2020

indicating that rather than daily medication collection participants were to collect weekly and not be observed whilst taking them. This may have had an impact on some participants as they had access to more medication and more freedom with these (i.e. there was be choice as to whether to take medication as directed). This could result in participants in these groups being more likely to disengage from treatment and relapse. However, subsequent guidance has recommended that supervision and dispensing interval should return to be determined on a case-by-case basis and informed by risk assessment [8]. This advice may change again in the event of a substantial increase in Covid-19 infection incidence.

Although these changes may have had an impact, investigators have no reason to believe that this will affect any signal (or no signal) detected.

However, baseline characteristics will be summarised by enrolment period (pre Covid-19, during Covid-19, post Covid-19) to assess if there are any relevant differences in the recruited population relative to the pandemic periods [9]. If deemed necessary sensitivity analysis adding 'pandemic enrolment period' into the model will be run and if found to be significant subgroup analysis may be conducted.

Finally, as indicated in section 4.6 an evaluation of missing data will be conducted prior to analysis. Missing data will be assessed by 'pandemic enrolment period' to check whether there are any differences of missingness caused by the pandemic. Missing data will be handled in line with section 4.6 of this SAP with additional guidance from recent literature on dealing with missing data in a pandemic [10].

One element of study processes which have been largely impacted by COVID is the data collection of the UDS, which will have an impact on the primary outcome. Essentially this is seen as missing data but it doesn't impact the level of missing data just the reliability of the self-report and occurs in both arms of the trial. The trial originally planned to collect UDS tests every two weeks however the following occurred;

Table 3: Summary of UDS data collection during trial following COVID-19 impacts

Site	Date	UDS data collection	
		TAU arm	Intervention arms
All	Up to 2nd April 2020	every 2 weeks	every 2 weeks

All	2Nd April 2020 – 12th June 2020	Not collected	Not collected
South London	04.05.2020	Not Collected	Monthly
	12.06.20	Monthly	Monthly
Newcastle	07.05.2020	Not Collected	Monthly
	29.06.20	Monthly	Monthly
Manchester	13.07.20	Monthly	Monthly
West Midlands	03.07.20	Monthly	Monthly
Tayside opened for recruitment on	27.07.20	Monthly	Monthly

As at the point of writing the SAP the trial is continuing to collect UDS monthly (on the larger visit weeks 4,8, 12, 16,20, 24). As there is little study data that hasn't been impacted by this and it has affected both arms simultaneously sensitivity analysis around this is not required. If any further changes during the ongoing pandemic occur that the investigators feel will impact any signal detected this will be conducted as exploratory post-hoc analysis.

6.10 EXPLORATORY ANALYSES

As indicated in section 5.2, further analysis is to be conducted at a later date but will not form the results of the current analysis. A separate statistical plan detailing all other analysis will be written and published in addition to this SAP.

7. SOFTWARE

All quantitative analysis will be completed using STATA version 15 or later.

8. REFERENCES

1. Peacock A et al. Global statistics on alcohol, tobacco and illicit drug use: 2017 status report. Addiction. 2018 May 10. doi: 10.1111/add.14234.

2. White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: issues and guidance for practice. Statistics in medicine, 30(4), 377-399.

4. Sobell L. C., Sobell M. C., Leo G. I., Cancilla A. Behavioral assessment and treatment planning with alcohol and drug abusers: a review with emphasis on clinical application. *Clin Psych Rev* 1988; **8**: 19–54.

5. Fals-Stewart W., O'Farrell T. J., Freitas T. T., McFarlin S. K., Rutigliano P. The timeline followback reports of psychoactive substance use by drug-abusing patients: psychometric properties. *J Consult Clin Psychol* 2000; **68**: 134–44.

6. Sobell, Linda & Sobell, Mark. (1992). Timeline follow-back: A technique for assessing self-reported alcohol consumption. Measuring Alcohol Consumption: Psychosocial and Biochemical Methods. 41-72.

7. Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. Technometrics, 11(1), 1-21.

8. See:https://www.gov.uk/government/publications/covid-19-guidance-for-commissioners-and-providers-of-services-for-people-who-use-drugs-or-alcohol/covid-19-guidance-for-commissioners-and-providers-of-services-for-people-who-use-drugs-or-alcohol

9. Meyer et al (2020): Statistical Issues and Recommendations for Clinical Trials Conducted During the COVID-19 Pandemic, Statistics in Biopharmaceutical Research, DOI: 10.1080/19466315.2020.1779122

10. Cro, S., Morris, T.P., Kahan, B.C. et al. A four-step strategy for handling missing outcome data in randomised trials affected by a pandemic. BMC Med Res Methodol 20, 208 (2020). https://doi.org/10.1186/s12874-020-01089-6

APPENDICES

APPENDIX 1 – POISSON REGRESSION ANALYSIS DECISION TREE

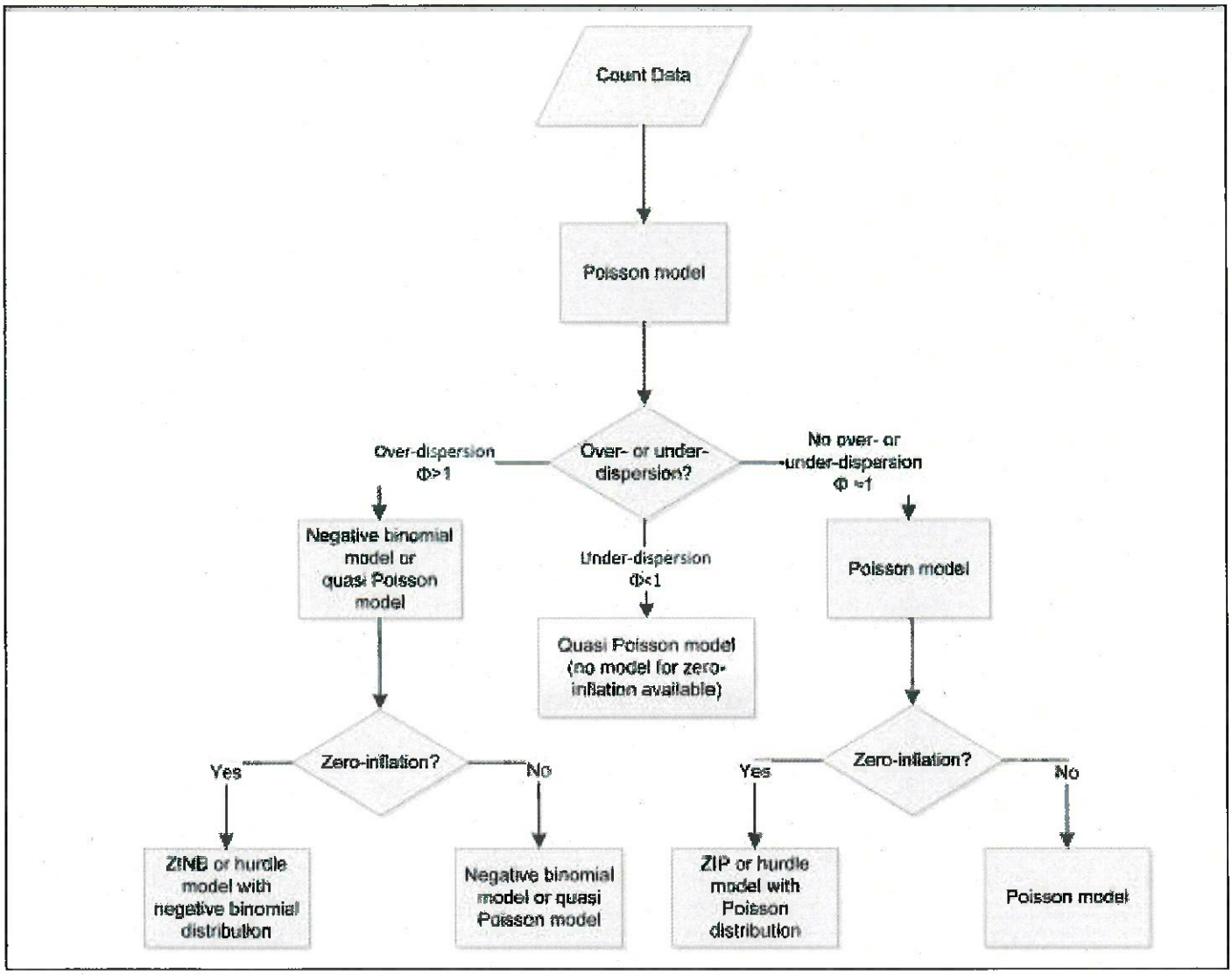


Figure obtained from;

https://pdfs.semanticscholar.org/8ab1/829f559b869eeb48e38f15c5d94dc957a0f5.pdf

APPENDIX 2 – OBJECTIVES AND ASSOCIATED OUTCOME MEASURES AND CHARACTERISTICS

Objective	Outcome Measure	Time Period	End point	Variable Type	Analysis Model	Scoring
One - effectiveness of BUP-XR and MET/BUP-SL	**Primary outcome**					
	Number of days abstinent from heroin	Days 8 – 168	24 weeks	Count	Poisson Regression	Count of days abstinent
	Secondary outcomes					
	Number of days abstinent from cocaine	Days 8 – 168	24 weeks	Count	Poisson Regression	Count of days abstinent
	Number of days abstinent from Benzodiazepines	Days 8 – 168	24 weeks	Count	Poisson Regression	Count of days abstinent
	Longest time in days continuously abstinent from heroin	Days 8 – 168	24 weeks	Count	Poisson Regression	Max Count of days consecutively abstinent
	Longest time in days continuously abstinent from cocaine	Days 8 – 168	24 weeks	Count	Poisson Regression	Max Count of days consecutively abstinent
	Longest time in days continuously abstinent from Benzodiazepines	Days 8 – 168	24 weeks	Count	Poisson Regression	Max Count of days consecutively abstinent
Two – Safety profile	Number of safety events	Study enrolment to 24 weeks	24 weeks	Count	Descriptive	n/a
Three – Treatment retention	Total retention time	Days 8 – 168	24 weeks	Count	Poisson Regression	Count of Days enrolled in maintenance medication for OUD
	Survival time to study withdrawal	Randomisation date to first discontinuation	24 weeks	Time to event	Cox regression	Time to first discontinuation
Four – OUD and CUD remission status	SCID-RV interview	Baseline to week 24	12 weeks & 24 weeks	binary	(Binary) logistic regression	Remission status (Yes/No) indicated by SCID-RV (see appendix 3 for SCID scoring)

Objective	Outcome Measure	Time Period	End point	Variable Type	Analysis Model	Scoring
Five – Clinician rated impression of patient response to treatment	Clinical Global Impression scale (severity and improvement - CGI-S/CGI-I)	Baseline to week 24	24 weeks	Ordinal	Ordinal logistic Regression	See Appendix 3
	Addiction Dimensions for Assessment and Personalised Treatment (ADAPT)	Baseline to week 24	24 weeks	Continuous	GLM	See Appendix 3
Six – Patient reported recovery outcomes	Craving response using Craving experience questionnaire (CEQ-F)	Baseline to week 24	24 weeks	Continuous	GLM	See Appendix 3
	Visual analogue craving rating (Need) (VAS-N)	Baseline to week 24	24 weeks	Continuous	GLM	See Appendix 3
	Visual analogue craving rating (Want) (VAS-W)	Baseline to week 24	24 weeks	Continuous	GLM	See Appendix 3
	Social functioning using Work and Social Adjustment Scale (WSAS)	Baseline to week 24	24 weeks	Continuous	GLM	See Appendix 3
	Depression using Quick inventory of Depressive Symptomatology (QIDS-SR)	Baseline to week 24	24 weeks	Continuous	GLM	See Appendix 3
	Improvement and Recovery using Substance Use Recovery Elevator (SURE)	Week 4 to week 24	24 weeks	Continuous	GLM	See Appendix 3
	Patient Reported Outcome Status (severity and improvement - PRO-S/PRO-I)	Baseline to week 24	24 weeks	Ordinal	Ordinal logistic Regression	See Appendix 3
	Emotion regulation using the Difficulties in Emotion Regulation Scale- short form (DERS-SF)	Baseline to week 24	24 weeks	Continuous	GLM	See Appendix 3
	Alcohol quantity, frequency, maximum consumption (ALC-QF (TOP))	Baseline to week 24	24 weeks	Continuous	GLM	See Appendix 3

APPENDIX 3 - OUTCOME MEASURES SUMMARY AND SCORING TABLE

Definition	Scoring	Item coding	Subscales	Direction	Missing value rules	Thresholds
Standard Clinical interview for Dependence – Research version **SCID-RV** *No reference available; below scoring derived from MACRO database*						
Structured Clinical Interview for DSM-5 (SCID-5) is a semi structured interview guide for making the major DSM-5 diagnoses.	Two single item scales (one for opioid and one for cocaine). Remission status is to be calculated from the SCID. Remission is satisfied (i.e. 'yes') if no symptoms are present that is a score of 1 or lower is obtained. If a score of 2 or more is given then remission status is 'no'.	Items are coded on a 5 point scale based on number of symptoms present; no symptoms, 1 symptom, Mild (2 – 3), Moderate (4-5) and Severe (6 or more)	None found	Lower better	None found	None found
Clinical Global Impression scale (severity and improvement - **CGI-S/CGI-I**) Busner, J. and Targum, S.D., The Clinical Global Impressions Scale: Applying a Research Tool in Clinical Practice (2007). *Psychiatry.*						
Brief stand-alone assessment of the clinician's view of the patient's global functioning. The scale comprises two companion one-item measures evaluating the severity of psychopathology and change from the initiation of treatment.	Two single item scales (CGI-S and GCI-I). Both items individually scored on a 1 to 7 point scale. Total score is not to be calculated from the two scales. However, the two scale scores will range from 1 to 7.	The CGI severity ranges from 1 'extremely mild' to 7 'extremely severe'	1. Severity (1 item)	Lower better	None found	1 – Extremely mild 2– Very mild 3 – Mild 4 – Moderately severe 5 – Severe 6 – Very severe 7 – Extremely severe
		The CGI improvement ranges from 1 'very much improved' to 7 'very much worse'	2. Improvement (1 item)	Lower better		1– very much improved 2 – much improved 3 – minimally improved 4 – no change 5 – minimally worse 6 – much worse 7 – very much worse
Addiction Dimensions for Assessment and Personalised Treatment **(ADAPT)** Marsden, J., et al., Deelopment of the Addiction Dimensions for Assessment and Personalised Treatment (ADAPT). *Drug Alcohol Depend.* (2014), http://dx.doi.org/10.1016/j.drugalcdep.2014.03.018						

Used for substance use disorders (SUD) clinicians to obtain a brief multi dimension patient profile of addiction related severity, health and social problem complexity and recovery strengths	A 14 item measure consisting of three subscales. Each subscale total is calculated by summing the individual items in the subscale. An overall measure total score is not obtainable, just 3 subscale scores. 1. total range 0 – 5 2. total range from 0 – 15 3. total range 0 – 11	Items 1, 3, 4, 5, 9, 10 rated; 0 'None' and 1 'Present' Items 2, 5, 6, 7, 8, 11, 12, 13, 14 rated 0 'None', 1 'Low', 2 'Moderate' and 3 'High'.	1. Addiction severity (items 1 – 3). 2. Coexisting problem complexity (items 4 – 10) 3. Recovery Strengths (items 11 – 14)	1. Higher scores indicate worse 2. Higher scores indicate better 3. Higher scores indicate better	None found	1. Low 0 – 1, moderate 2 – 3 and high 5 – 4 2. Low 0 – 2, moderate 3 – 5 and high 6 – 15 3. Low 0 – 5, moderate 6 – 8 and high 9 – 11
Craving response using Craving experience questionnaire (**CEQ-F**) *No reference available; below scoring derived from MACRO database*						
Craving Experience Questionnaire	11 item scale. Total scale calculated by summing the 11 items, therefore final score ranges from 0 to 110.	Each item rated on a scale from 0 – 10.	None identified	Higher scores indicate more severe craving	None found	None found
Visual analogue craving rating (Need/Want) (**VAS-N/VAS-W**) *No reference available; below scoring derived from MACRO database*						
Self-report craving rating	Two individual items one measures rating need and the other rating want. Total score is not to be calculated from the two scales. However, the two scale scores will range from 0 to 100.	Both items are scored on a 0 – 100 continuous scale	None identified	Higher scores indicate more severe craving	None found	None found
Social functioning using Work and Social Adjustment Scale (**WSAS**) Mundt, J.C., et al., The Work and Social Adjustment Scale: a simple measure of impairment in functioning (2002). *British Journal of Psychiatry, 180,* 461 – 464.						

Self-report scale of functional impairment attributable to an identified problem.	Five item scale. Total score calculated by summing up 5 items in the scale. Total score ranges from 0 - 40	Items rated on a 0 to 8 scale with 0 indicating no impairment and 8 indicating very severe impairment	None identified	Higher scores indicate better state.	None found	None found
Quick inventory of Depressive Symptomatology (**QIDS-SR**) Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. K., ... Keller, M. B. (2003). The 16-Item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychoMETric evaluation in patients with chronic major depression. *Biological Psychiatry, 54,* 573-583. http://www.ids-qids.org/Scoring_Instructions.pdf http://www.ids-qids.org/index2.html#table2						
self-report measure of depression	16 item scale. Subscales calculated by taking highest score of any one item of the subscale. Total score for each subscale ranging from 0 to 3 Total score calculated by summing the 9 subscales therefore total scores range from 0 to 27.	Items rated on a 0 to 3 scale; 0 indicating no symptom and 3 indicating worse.	1. Sleep disturbance (items 1 - 4) 2. Sad mood (item 5) 3. Decrease/increase in appetite or weight (items 6 - 9) 4. Concentration (item 10) 5. Self-criticism (item 11) 6. Suicidal ideation (item 12) 7. Interest (item 13) 8. Energy/fatigue (item 14) 9. Psychomotor agitation/ retardation (items 15 – 16)	Higher scores worse for all subscales	None found	For total score 1-5 = No depression 6-10 = Mild depression 11-15 = Moderate depression 16-20 = Severe depression 21-27 = Very severe depression
Improvement and Recovery using Substance Use Recovery Elevator (**SURE**) Neale, J., Vitoratou, S., Finch, E., Lennon, P., Mitcheson, L., Panebianco, D., Rose, D., Strang, J., Wykes, T., Marsden, J. (2016) 'Development and validation of 'SURE': A patient reported outcome measure (PROM) for recovery from drug and alcohol dependence', *Drug and Alcohol Dependence*. DOI: 10.1016/j.drugalcdep.2016.06.006						
Self-report measure to assess person al recovery from drug and/or alcohol dependence	21 item scale. Items 1 – 3 scored;	Section A (items 1 – 3) rated on a 5 point scale 1 – 5; ranging from 1 'Never' to 5 'everyday'	1. Drinking and drug use (items 1 – 6) 2. Self-care (items 7 – 11)	Individual items indicates lower better and higher worse but scores are then weighted higher if lower so	None Found	None Found

	'Never' OR 'On 1 or 2 days' = 3, 'On 3 or 4 days' = 2, 'On 5 or 6 days' OR 'Every day' = 1 Items 4 – 21 scored; 'All of the time' OR 'Most of the time' = 3, 'A fair amount of the time' = 2, 'A little of the time' OR 'None of the time' = 1 Score range for each subscale 1. Drinking and drug use; 6 – 18 2. Self care; 5 – 15 3. Relationships; 4 – 12 4. Material resources; 3 – 9 5. Outlook on life; items 3 – 9 Total score ranges from 21 – 63	<u>Section B</u> (items 7 – 21) rated on a 5-point scale 1 – 5; ranging from 1 'none of the time' to 5 'all of the time'. <u>Section C</u> items not used for measure scoring.	3. Relationships (items 12– 15) 4. Material resources (items 16 – 18) 5. Outlook on life (items 19 – 21)	**total score format is higher scores indicate better.**		
Patient Reported Outcome Status (severity and improvement - **PRO-S/PRO-I**) *No reference available; below scoring derived from MACRO database*						
Patient reported outcome of severity and improvement of addition problem	Two single item scales (PRO-S and PRO-I). Both items individually scored on a 1 to 7 point scale. Total score is not to be calculated from the two scales. However, the two scale scores will range from 1 to 7.	The PRO severity ranges from 1 'extremely mild' to 7 'extremely severe'	1. Severity (1 item)	Lower better	None found	1 – Extremely mild 2–Very mild 3 – mild 4 – moderate 5 – severe 6 – very severe 7 – extremely severe
		The PRO improvement ranges from 1 'very much improved' to 7 'very much worse'	2. Improvement (1 item)	Lower better		1–very much improved 2 – much improved 3 – minimally improved 4 – no change 5 – minimally worse 6 – much worse 7 – very much worse
Emotion regulation using the Difficulties in Emotion Regulation Scale- short form (Emotion regulation using the Difficulties in Emotion Regulation Scale- short form (**DERS-SF**))						

Kaufman, E. A., Xia, M., Fosco, G., Yaptangco, M., Skidmore, C. R., & Crowell, S. E. (2015). The difficulties in emotion regulation scale short form (DERS-SF): Validation and replication in adolescent and adult samples. Journal of Psychopathology and Behavioral Assessment, doi:10.1007/s10862-015-9529-3						
self-report measure for assessing emotion regulation problems	18 item scale Items in each scale summed to give individual scale scores (ranging from 1 to 5) and total score calculated by summing scales, total score ranging from 18 to 90	Each item scored on a 5 point scale ranging from 1 'Almost Never' to 5 'Almost always' Awareness scale should be reverse coded so that items range from 1 'almost always' to 5 'almost never'.	1. Strategies (items 10, 15 and 18) 2. Non-acceptance (items 7, 12 and 16) 3. Impulse (items 9, 14 and 17 4. Goals (items 8, 11 and 13) 5. Awareness (items 1, 4 and 6) 6. Clarity (items 2, 3 and 5)	Lower scores indicate better awareness scale (items 1, 4 and 6) lower scores indicate worse	None found	None found
Alcohol quantity, frequency, maximum consumption (ALC-QF (TOP)) *No reference available; below scoring derived from MACRO database*						
Alcohol – typical Quantity and Frequency. Using an adapted version of the Treatment Outcomes Profile (TOP). The TOP is the standard national instrument for monitoring the outcomes of public substance use disorder treatment services in England.	3 items in measure; number of days, average units and maximum units on any given day. Maximum units will be reported descriptively. Quantity and Frequency will be treated as two separate items.	Items rated based on number of drinking days in past four weeks; 0 to 28 Units given on scale from 0 to 100 Maximum units ranges from 0 to 500.	None found	Higher numbers indicate more drinking days and units.	None found	None found
The Montreal Cognitive Assessment (MoCA) Original ref: Nasreddine Z.S., Phillips N.A., Bédirian V., Charbonneau S., Whitehead V., Collin I., Cummings JL, Chertkow H. The Montreal Cognitive Assessment (MoCA): A Brief Screening Tool for Mild Cognitive Impairment, Journal of the American Geriatrics Society. 53 (2005) 695-699.						
Evaluates whether someone's thinking ability is impaired	10 Item scale, with an additional item (No. 11) to indicate education level. Items 1-10 are summed together to give a total score ranging from 0 to 30.	10 items are rated on various scales (0 to 1, 2, 3, 5 and 6)	None found	Higher numbers indicate less cognitive impairment	If Item 11 (education level) is missing then items 1-10 should still be calculated.	A score of 26 or above is considered normal.

	An extra point is awarded if item 11 is 'Yes' (i.e. subject has 12 years or fewer of formal education), However, the max score is 30 so If a participant scores 30 out of 30 this additional point isn't added.					

APPENDIX 4 – ASSUMPTIONS OF MODELS

Assumption	Checking
Poisson Regression	
distribution of counts (conditional on the model) follow a Poisson distribution	calculate the expected counts and plot these with the observed counts to see if they are similar
The mean and variance of the model are identical If you satisfy this assumption you have equidispersion. often this is not the case and your data are either under- or over dispersed.	Graphical display Pearson dispersion statistic See appendix 1 for more detail.
Generalised Linear mixed model	
Linearity the relationship between the independent and dependent variables to be linear	scatter plots of the model residuals vs predictor
U8Residuals/errors are independent Little or no autocorrelation in the data. (residuals should be independent from each other)	Scatter plot Durbin-Watson test
Residuals/Errors are normally distributed	Q-Q-Plot
Residuals/Errors have constant variance There should be no homoscedasticity of error terms	Scatter plot of standardized residuals versus predicted values
No or little multi-collinearity (independent variables should not be highly correlated with each other)	inspection of correlation coefficients and Tolerance/VIF values
ORDINAL LOGISTIC REGRESSION	
There should be no multi-collinearity	inspection of correlation coefficients and Tolerance/VIF values
You should have proportional odds	full likelihood ratio test comparing the fit of the proportional odds model to a model with varying location parameters
BINARY LOGISIC REGRESSION	
Linear relationship between the continuous independent variables and the logit transformation of the dependent variable.	Box-Tidwell (1962) procedure
Your data must not show multi-collinearity	inspection of correlation coefficients and Tolerance/VIF values
There should be no significant outliers, high leverage points or highly influential points	Box plots
Cox regression model	
proportional hazards the *ratio* of the hazards for any two individuals is constant over time	Kaplan Meier curves
linear covariate relationships The Cox model assumes that each variable makes a linear contribution to the model	residual plots