Statistical Analysis Plan

Final Version 4.0

14 March 2024

Study Title: *A randomised controlled trial of a brief cognitive task intervention to support NHS staff experiencing intrusive memories of traumatic events from working in the COVID-19 pandemic*

Final SAP Date: 14 March 2024

Short Study Title: *A Brief Cognitive Task Intervention for NHS Staff affected by COVID-19 trauma (GAINS-2 Study)*

Based on Study Protocol: *P1V GAINS IN02 Part 2 Study Protocol (v7.0 12Feb2024)*

Trial registration: *NCT05616676 (ClinicalTrials.gov)*

Sponsor: *P1vital Products Ltd, Manor House, Howbery Park, Wallingford, Oxfordshire, OX10 8BA*

SIGNATURES

The following people have reviewed the Statistical Analysis Plan and are in agreement with the contents.

Name	Role	Signature	Date
Dr. Amy Beckenstrom	Chief Investigator Senior Digital Health Scientist. P1vital Products Limited (Sponsor)	Amy Beckenstrom	14/03/2024
Prof. Emily Holmes	Scientific Lead Professor, Department of Women's and Child's Health, Uppsala University	Emily Holmes	14/03/2024
Jonathan Kingslake	CEO, P1vital Products Limited (Sponsor)	J Kingslake	15/03/2024
Dr. Rebecca Dias	Director of Operations, P1vital Products Limited (Sponsor)	Rebecca Dias	14/03/2024
Prof. Michael Bonsall	Statistical Lead Professor of Mathematical Biology, University of Oxford	MBBonsall	15/03/2024
Owen Slade	Statistical Analyst Statistical Programmer, P1vital Limited	OSlade	14/03/2024
Dr Asad Malik	Statistical Advisor/Analyst Head of Statistical Sciences, P1vital Limited	Asad Malik	14/03/2024
Prof. Thomas Jaki	Expert Advisory Panel Professor of Statistics and MRC Investigator, University of Cambridge	Thomas Jaki	14/03/2024

CONTENTS

REVISION HISTORY

Version	Date	Change/Comment	Statistician
V4.0	13Mar2024	• Minor formatting and consistency changes • Finalising of previous changes	Owen Slade
V3.6	12Mar2024	• Moving of frequentist analysis to sensitivity analysis • Specification of priors throughout for the Bayesian analysis	Owen Slade
V3.5	01Mar2024	• Removal of secondary frequentist analysis sections • Detail of secondary outcome analysis • Detail of other exploratory analyses • Formatting and wording changes throughout	Owen Slade
v3.4	13Dec2023	• Changing roles from VR to OS • Addition of frequentist analysis sections • Addition of abbreviations section • Changing of wording for exploratory analysis to match protocol	Owen Slade
V3.3	29Sep2023	• Tidying of comments and minor wording changes • This version revision history has been added retrospectively by OS	Varsha Ramineni
v3.2	21Jun2023	• Adding in Job Role to Demographics • Small Corrections	Varsha Ramineni
v3.1	9Jun2023	• Removing LEC outcome measure to match protocol • Adding intrusive memory accuracy ratings	Varsha Ramineni
v3.0	10May2023	• Finalising changes	Varsha Ramineni
v2.2	28Apr2023	• Update to sections: Analysis Populations, Handling Missing Data, and Sensitivity Analyses, to reflect the Study Protocol and missing data approach. • Updating Table 1 to match Study Protocol v4.0	Varsha Ramineni
v2.1	31Mar2023	• Adding detail on handling missing data (section 5.8.2 Multiple Imputation) • Editing structure to be more clear (section 5.8.1 Intrusive Memory Diary Data)	Varsha Ramineni
v2.0	10Mar2023	• Updating text regarding first planned interim analyses in Section 2.5	Varsha Ramineni
v1.0	13Jan2023	• Removing time frame for active control adherence (section 3.2) • Removing additional analyses section as it was empty • Including detail for dropouts out in missing data section (section 5.8.1) • Finalising all changes	Varsha Ramineni
v0.6	23Dec2022	• Adding quality assurance section • Small corrections throughout • Removing Qualitative questionnaire from outcomes tables	Varsha Ramineni
v0.5	15Dec2022	• Adding missing data timeframe for interim analyses	Varsha Ramineni
v0.4	12Dec2022	• Adding adherence for active control arm • Adding revised sample size section • Small corrections throughout	Varsha Ramineni
v0.3	17Nov2022	• Corrections to outcome measures • Detailing the math model for secondary ordinal outcome • Edits to the missing data section	Varsha Ramineni
v0.2	14Nov2022	• Implementing comments from Asad Malik (P1vital) and Michael Bonsall • Expansion of secondary outcome modelling & sensitivity analysis • Correction to outcome measures • Minor edits throughout	Varsha Ramineni
v0.1	24Oct2022	• Initial draft.	Varsha Ramineni

ABBREVIATIONS

Abbreviation	Description
AC	Active control
AIC	Akaike information criterion
BF	Bayes Factor
BIC	Bayesian information criterion
CCP	Complete-Case Population
COVID-19	Coronavirus Disease 2019
DMC	Data Monitoring Committee
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, 5th Edition
EAP	End-of-Study Analysis Population
EQ-5D-5L	5-level EQ-5D
GAD-2	Generalised Anxiety Disorder – 2-item questionnaire
IAP	Interim Analysis Population
ICU	Intensive Care Unit
IMDD	Intrusive Memory Daily Diary
IMR	Intrusive Memory Rating
IRR	Incidence Rate Ratios
ITL	Intention to leave job
ITT	Intention to treat
MCAR	Missing completely at random
NHS	National Health Service
PCL-5	PTSD Checklist for DSM-5
PHQ-2	Patient Health Questionnaire – 2-item version
PL	P1vital
PPL	P1vital Products
PPP	Per-Protocol Population
PTSD	Post-traumatic Stress Disorder
RCT	Randomised control trial
RMSE	Root mean square error
SAP	Statistical Analysis Plan
SCI	Sleep Condition Indicator
SWEBO	Scale of Work Engagement and Burnout
TAU	Treatment as usual
WHODAS	World Health Organisation Disability Assessment Schedule

1 INTRODUCTION

This statistical analysis plan (SAP) details the proposed plan to be followed when analysing and reporting quantitative results from the randomised controlled trial study of a brief cognitive task intervention to support NHS staff experiencing intrusive memories of traumatic events from working in the COVID-19 pandemic.

The purpose of this SAP is to:

1. Ensure that the analysis is appropriate for the aims of the trial, reflects good statistical practice, and that interpretation the results is appropriate.
2. Explain in detail how the data will be processed and analysed for pre-specification and to enable others to perform and/or replicate these analyses.

The main analysis of the study will follow the strategy set out here, which adheres to the guidelines for the content of a SAP (Gamble et al., 2017). Additional exploratory or auxiliary analyses of data not specified in the protocol may be included in this analysis plan.

Any subsequent analyses of a more exploratory nature will not be bound by this strategy. Suggestions for subsequent analyses by oversight committees, journal editors or referees, will be considered carefully in line with the principles of this analysis plan and will be documented in a file note.

Amendments or deviations from the SAP will be described and justified in the final report of the trial and, where appropriate, in publications arising from the project.

The analysis will be carried out by identified, appropriately qualified and experienced statisticians, who will ensure the integrity of the data during their processing.

This analysis plan, or a simplified version as appropriate, will be published in a publicly available platform before data base lock in line with Wellcome Trust recommendations.

Other health economic and qualitative analysis plans are beyond the scope of this document.

This version of the SAP was developed using P1V GAINS-IN02 Protocol version 7.0, dated 12th February 2024.

1.1 Background and Rationale

There is an urgent need for scalable approaches to support the mental health of frontline healthcare workers (Holmes et al., 2020) and other groups exposed to work-related traumatic events. Amplified by the COVID-19 pandemic, frontline healthcare staff are at elevated risk of post-traumatic stress disorder (PTSD) and other mental health symptoms due to increased exposure to occupational trauma, e.g., traumatic or tragic death of a patient, and require prompt access to effective interventions. Retaining healthcare staff in their jobs and preventing work dropout is necessary for delivering critical care to NHS patients.

A significant proportion of those who experience traumatic events experience intrusive memories (or “flashbacks”) of these events that pop suddenly into mind; these imagery-based memories can disrupt functioning and contribute to post-traumatic stress disorder. Even before the COVID-19 pandemic, 65% of emergency nurses reported having intrusive memories of work-related traumatic events (Kleim et al., 2015). Intrusive memories of trauma may be a potential target for early and preventative interventions post-trauma as this symptom is distressing / impairing in its own right, and because it may prevent PTSD from developing (Bryant et al., 2017; Iyadurai et al., 2019; McNally, 2017).

Previous research has shown that a brief behavioural intervention can reduce the number of intrusive memories after a traumatic event. Positive results for the intervention’s effectiveness in reducing the number of intrusive memories has been seen in three randomised controlled trials (Horsch et al., 2017; Iyadurai et al., 2018; Kanstrup, Singh, et al., 2021) two case series studies (Kanstrup, Kontio, et al., 2021; Kessler et al., 2018), and most recently for NHS ICU staff exposed to work-related trauma including during the COVID-19 pandemic (Iyadurai et al, 2023; Ramineni et al., 2023), see also Deforges et al (2023)..

The current study is a randomised controlled trial study of a brief cognitive task intervention to support NHS staff experiencing intrusive memories of traumatic events from working in the COVID-19 pandemic. It aims to test the efficacy of a brief cognitive task intervention against an alternative brief cognitive task (active control), and treatment as usual (TAU) for NHS staff with intrusive memories of work-related traumatic events associated with the pandemic. We test the effect of task(s) on the number of intrusive memories at week 4 (primary outcome), and other clinical symptoms, work and general functioning and quality of life (secondary outcomes) at 4, 12 and 24 weeks (including number of intrusive memories at 12 and 24 weeks). Results of the study will be relevant globally to healthcare staff affected by traumatic events through the COVID-19 pandemic as well as other trauma-exposed groups.

This randomised control trial (RCT) uses an adaptive Bayesian design for efficacy evaluation of a brief cognitive task intervention. Recent advances in trial design and methodology offer more efficient alternatives to traditional RCTs to speed up the testing and thus implementation of evidence-based treatments (Dimairo et al., 2020). Adaptive designs enable smaller, more efficient trials without loss of scientific integrity, and allow a trial to be modified based on interim analyses, thereby making optimal use of all data for decision-making. Compared to traditional designs, a sequential Bayesian approach typically requires 50% to 70% smaller samples to conclude the presence of an effect and has the same or lower rate of false inference (Schönbrodt et al., 2017).

1.2 Study Objectives

1.2.1 Primary Objectives

The primary objective is to determine if access to a brief cognitive task intervention compared to an active control, and compared to TAU, reduces the number of intrusive memories in week 4 (i.e. between-groups comparison), controlling for the number of intrusive memories in the run-in/baseline week.

1.2.2 Secondary Objectives

The secondary objectives of this trial are:

- To determine if access to a brief cognitive task intervention compared to an active control, and compared to TAU, reduces the number of intrusive memories in weeks 12 and 24 (i.e. between-groups comparison), controlling for the number of intrusive memories in the run-in/baseline week.
- To determine if access to a brief cognitive task (intervention), compared to an active control, and compared to TAU, reduces the impact of intrusive memories (e.g. disruption to concentration and functioning); clinical symptoms of PTSD, insomnia, anxiety, and depression; burnout, sickness absence and intention to leave the job; and improves work satisfaction, general functioning and quality of life at 4 weeks, 12 weeks, and 24 weeks (i.e. between-groups comparisons).

1.2.3 Exploratory Objectives

The exploratory objectives of this trial are:

- To assess new traumatic events, adverse events, treatment received, and self-reported changes to job or hours worked.
- To obtain intervention usage data to inform the intervention implementation.
- To assess the acceptability and feasibility of the intervention from participants to inform the intervention implementation.
- To obtain complementary results using alternative statistical methods (i.e. frequentist analysis) for the primary outcome measure to aid comparability of results with other trials using traditional frequentist analysis.

2 STUDY METHODS

2.1 Trial Design

The study will enrol NHS staff (e.g. doctors, nurses, paramedics and clinical support staff) aged 18 and over who: a) have experienced one or more work-related traumatic events during the COVID-19 pandemic (e.g. a patient death), and b) have intrusive memories of the event(s).

The study is a three-arm, parallel-group, Bayesian adaptive, randomised controlled trial. The study randomly allocates participants using a 2:2:1 overall ratio respectively to the three arms described below.

- **Intervention arm:** one guided session and access to a brief cognitive task intervention (self-guided) for 24 weeks (first 4 weeks with optional researcher support)
- **Active control arm:** one guided session and access to an alternative brief cognitive task (self-guided) for 24 weeks (first 4 weeks with optional researcher support)
- **Treatment as usual (TAU) arm:** routine care that participants would otherwise receive if having intrusive memories of traumatic events for 24 weeks.

The total duration of the study from first participant enrolment to the last participant completing the study is expected to last approximately 11 months but will depend on the final number enrolled. The study will enrol up to approximately 150 participants (see Section 2.3 ‘Sample Size’), 60:60:30 per study arm (intervention: active control: TAU).

Please refer to the study flow chart in Figure **1** to see the overall trial design.

An adaptive trial design has been selected to allow for the trial to be altered (i.e. stopping the trial early either due to efficacy or evidence of negative effect) based on evidence from interim analyses. For further details please see Section 2.5 ‘Statistical Interim Analyses and Stopping Guidance’ and Section 5 ‘Analysis’.

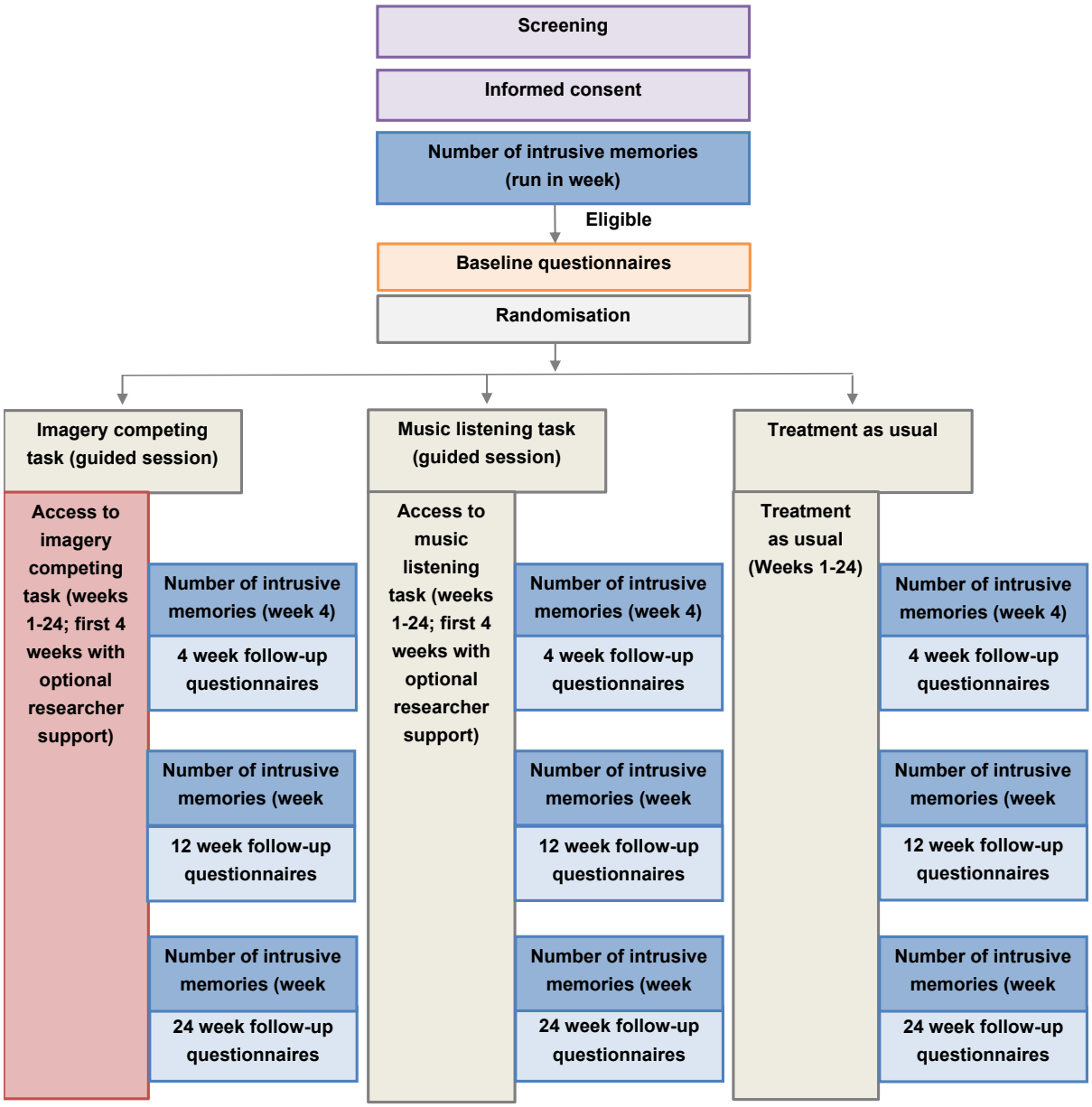

Figure 1: Study Flow Chart

2.2 Randomisation and Blinding

2.2.1 Randomisation

Participants who meet the eligibility criterion of having 3 or more intrusive memories in the run-in week (see Figure **1**) will be allocated to either the intervention, active control arm, or TAU arm using a 2:2:1 overall ratio. Participants will be randomised via a randomisation list using

block randomisation with random block sizes of multiples of 5. A randomisation program built into ePRO will be used to ensure that allocation cannot be influenced by the research team (i.e., randomisation is computerised and automated to ensure allocation concealment). The uploaded randomisation list will be validated by an independent statistician.

2.2.2 Blinding

Statisticians will be blinded to allocation, and all assessments are self-report questions administered digitally, eliminating assessor bias (i.e., to ensure blinding of outcome assessment).

Participants will be blinded to intervention or active control as both are brief cognitive tasks. Participants in the TAU arm cannot be blinded.

Researchers involved in contacting the participants and facilitating the conduct of intervention or active control arms will not be blinded. The remaining principal investigator's delegated study members will also be unblinded during the length of the trial.

Any critical trial decisions based on interim analyses (i.e., early stopping) will be made by the steering committee and the Data Management Committee (DMC), who will remain unblinded throughout the study.

2.3 Sample Size

We employ a sequential Bayesian design with maximal sample size (Schönbrodt et al., 2017; Schönbrodt & Wagenmakers, 2018); this allows for interim analyses to guide decision-making on early stopping. The maximal sample size for sequential Bayes factor designs is typically based on practical reasons (Ewings et al., 2022; Giovagnoli, 2021; Heath et al., 2021; Schönbrodt & Wagenmakers, 2018). Here, we take a maximum sample size of 150 with 2:2:1 ratio between Intervention, Active Control, and TAU arms (60:60:30). This was based on pragmatic details associated with the design, and using traditional power analysis (Freeman et al., 2020; Ryan, Lamb, et al., 2020; Ryan, Stallard, et al., 2020), assuming a two-sided

independent t-test, alpha level of 0.05, and statistical power of 90%, this sample size is sufficient to detect effect d = 0.6 for Intervention vs Active Control, and effect d = 0.75 for Intervention vs TAU.

Our previous Bayesian optimisation study showed effect size d=0.85 (Ramineni et al., 2023) using waitlist control, therefore we believe that a more conservative effect size of d=0.75 is reasonable for the Intervention vs TAU comparison. A previous pilot study using active control showed an effect size of d=0.57 (Kanstrup, Singh, et al., 2021). However, Kanstrup et al., 2021 was a prevention (rather than treatment) study, whereby the intervention was delivered only once and only on day one of trauma, with lower average number of intrusive memories at baseline than expected in the current trial. Therefore, a greater effect is anticipated under the current design (a treatment trial where the intervention can be used more than once over a 4-week period, and with higher baseline number of intrusive memories expected). From a previous lab study using active control, we estimate an effect size d=0.71 (effect of Tetris-only control scaled down by 15% to account for difference of lab-studies) (James et al., 2015). Given this evidence, an effect size of d=0.6 for Intervention vs Active Control comparison is reasonable.

Additionally, we have used the BFDA package for Bayesian design analysis (Schönbrodt & Stefan, 2019) (an R package for Bayes factor design analysis available from https://github.com/nicebread/BFDA). In this study, early stopping for effectiveness will be based on Intervention vs Active Control comparison (see Section 5.4 'Sample Size Determination and Early Stopping'). Even under the assumption of a modest effect size of d = 0.5, simulations indicate that 48% of studies would find strong evidence for the effect and could stop recruitment early (i.e., before reaching n = 60 in each group for Intervention vs Active Control). Further, the BFDA simulations show that the rates of misleading evidence are low, with virtually 0% false positive and false negative rates i.e., this design is highly unlikely to reach a false conclusion that evidence of an effect exists when there is none, and vice versa. These simulations were computed using a central Cauchy distribution (scale parameter

√2/2) as the effect size prior, between-group t-test with one-sided (positive directional) hypotheses, and symmetric Bayes factor (BF) decision boundaries of 20 (for strong evidence).

2.4 Framework

All outcomes will be analysed within a Bayesian framework based on posterior distributions.

Bayesian regression modelling will be used throughout the course of the study to evaluate the intervention based on the primary objective (between-group difference in the number of intrusive memories in week 4). The primary objective is explored for superiority (see Wang et al., 2017): the null and alternative hypotheses are arranged to test whether those in the intervention arm compared to the TAU arm, and compared to the active control arm, have fewer intrusive memories in week 4, while controlling for the baseline number of intrusive memories. Primary outcome modelling will also be conducted at the end of the study along with all secondary outcomes.

Similarly, between-groups analyses will be used to test the difference in the number of intrusive memories in weeks 12 and 24 between the intervention, active control and TAU groups. Between-groups analyses will also be used to test for differences in other secondary outcomes in week 4, week 12, and week 24 between the intervention, active control and TAU groups for intrusive memory ratings; symptoms of post-traumatic stress, insomnia, anxiety and depression; work engagement and burnout, sickness absence, and intention to leave job; and functioning and quality of life.

Exploratory analysis will be conducted for exploratory objectives, with descriptive statistics presented.

Details of analysis of all primary and secondary, and exploratory outcomes, are described in Section 5 ‘Analysis’.

2.5 Statistical Interim Analyses and Stopping Guidance

This RCT uses a Bayesian adaptive trial design for efficacy evaluation of a brief cognitive task intervention. The study aims to use accumulated data information to inform trial design, hence the blinded data will be analysed sequentially by means of sequential Bayes factor analysis (Schönbrodt et al., 2017; Schönbrodt & Wagenmakers, 2018).

Frequent planned analyses will be performed sequentially up to a maximum sample size of 150 to evaluate whether there is any evidence of a negative effect, and if there is sufficient evidence of positive treatment effect to consider early stopping of the trial. Pre-specified decision rules are defined that allow the trial to be stopped at planned analyses. The pre-defined decision rules for early stopping are described in Section 5.4 'Sample Size Determination and Early Stopping'.

The first planned analysis will be triggered following passing of primary outcome (i.e. 28 days post first intervention/active control session/post randomisation in TAU arm) of approximately 20-25 randomised trial participants and conducted as close to that as is practical given logistical constraints. If the pre-defined decision rule requirements are not met, successive planned analyses will be conducted after approximately every 4-10 participants pass the primary outcome time-point.

If the maximal sample size of 150 is reached, or a trial decision rule is met and an early stopping decision is subsequently recommended by the DMC, then all enrolled participants will be followed up to their final online assessment (completion of week 24 outcome measures i.e., Day 168) and the final analyses will be undertaken and reported on primary, secondary and exploratory outcomes.

Details of sequential analysis such as the stopping rule and time, will be described in Section 5 'Analysis'.

2.6 Timing of Final Analysis

Final analyses will be conducted at the end of the study on the final sample size which is determined from results of the interim analyses. The end of the study is defined as the date that the last participant completes their final online assessment (completion of week 24 outcome measures i.e., Day 168).

2.7 Timing of Outcome Assessments

Briefly, data from all randomised patients will be collected at baseline, 4 weeks follow-up, 12 weeks follow-up, and 24 weeks follow-up time. The schedule of study procedures for all data collection is given below in Table 1 and corresponding timings for primary, secondary, and exploratory outcomes will be given in Section 5.1 'Outcome Definitions'.

Some demographic information (job role during pandemic) will be collected via electronic source document during informed consent meeting (if missed, this can be asked any time up to the final outcome at 24 weeks).

Table 1: Time of Events Schedule

Virtual Visit Name	Pre-Screening	Screening	Baseline	Randomisation Confirmation	Week 4	Week 12	Week 24
Day	Before D-35	D-35	D-28	D1	D21-D28	D77-84	D161-168
Visit Windows		+28 days	+28 days		+49 days	+77 days	+28 days
Participant information sheet	X						
Eligibility questionnaire	X						
Consent form		X					
Inclusion/exclusion criteria		X					
Intrusive memory diary (daily for seven days ending on day indicated)			X		X	X	X
Demographics			X				
Health background			X				
Checklist of work-related traumatic events			X				
PTSD Checklist for DSM-5 (PCL-5)			X		X	X	X
Sleep Condition Indicator (SCI)			X		X	X	X
Generalized Anxiety Disorder (GAD-2)			X		X	X	X
Patient Health Questionnaire (PHQ-2)			X		X	X	X
Scale of Work Engagement and Burnout (SWEBO)			X		X	X	X
Sickness absence			X		X	X	X
Intention to leave job (ITL)			X		X	X	X

EQ- 5D-5L (5-level EuroQol 5D)			X		X	X	X
WHODAS 2.0			X		X	X	X
Intrusive memory ratings			X		X	X	X
Randomisation			X				
Credibility/Expectancy Questionnaire				X			
Intervention/active control/TAU					X		
Changes to health and work					X	X	X
Feedback questionnaire*					X		

* Feedback questionnaire only filled out by Intervention and Active Control arms

3 STATISTICAL PRINCIPLES

All primary and secondary outcomes (see Table 3) will be analysed within a Bayesian framework based on posterior distributions.

3.1 Overview

For all parameters of interest, posterior distributions will be visualised, posterior means will be reported along with 95% credible intervals.

Bayes factors (BF) will be presented to quantify the relative evidence to predict any null hypothesis or alternative hypotheses.

3.2 Adherence and Protocol Deviations

Please note that the following adherence definitions only apply to the intervention and active control arms. In cases where adherence is unclear, a case discussion will be held.

3.2.1 Adherence

Intervention Arm

During the first guided session in the intervention arm, the participant completes all key components of the intervention including: accurately identifying and briefly listing intrusive memories; choosing an intrusive memory to target; briefly bringing to mind the intrusive memory image before gameplay (so that the image is sufficiently clear, but not for so long that the participant becomes overly upset or distressed); then sufficient uninterrupted Tetris game play (c.20 min in total; range 15- 25 min); during gameplay actively using mental rotation.

Non-adherence may be indexed by in session reports or behaviours that are incompatible with the steps above e.g., deliberately bringing the memory to mind repeatedly during game play.

The following points will not be considered as protocol deviations:

- Technical problems which were resolved in some way e.g., not being able to view the video instructions for steps in the study procedure and instead reading the written transcript.
- Intervention glitches or internet connection problems which do not prevent key components from being completed.
- Minor interruptions during gameplay which do not disturb overall engagement in the game.

Active Control Arm

During the first guided session, the participant completes the control procedure, including: listening to a podcast excerpt and a 20-minute piece of music. Non-adherence is indexed by in session reports or behaviours that are incompatible with the steps above e.g., not listening to the podcast excerpt or to the music.

The following points will not be considered as protocol deviations:

- Technical problems which were resolved in some way e.g., not being able to view the video instructions for steps in the study procedure and instead reading the written transcript.
- Intervention glitches or internet connection problems which do not prevent key components from being completed.
- Minor interruptions during listening which do not disturb overall engagement.

3.2.2 Protocol Deviation

Definition of protocol deviations:

A full list of protocol deviations will be denoted prior to unblinding of treatment including:

1. Non-completion of the primary outcome measure

2. Non-completion of guided intervention session/non-adherence to the intervention (as defined above)

All protocol deviations will be summarised, using the number and percentage of participants in each arm and overall, with details of type of deviation provided. The participants that are included in the intention-to-treat (ITT) set overall and for each arm (see Section 3.3 'Analysis Populations') will be used as the denominator to calculate the percentages. Protocol deviations are classified prior to unblinding of treatment. A data review meeting will be held prior to unblinding the statistician(s) (see Section 2.2.2 'Blinding') where the Per-Protocol Population (PPP; see Section 3.3.3 'Per-Protocol Population (PPP)') will be determined according to the list of protocol deviations defined above.. The statistician(s) will not attend this meeting so that they are not unblinded at this stage. The protocol deviations list will be shared with the statistician after database lock.

3.3 Analysis Populations

Analyses will be conducted on an ITT basis, which includes all randomised participants. However, a full application of the ITT approach is possible only when complete outcome data are available for all randomised subjects (Hollis & Campbell, 1999). Therefore, where there is missing data, handling of missing responses will be explicitly described (see Section 5.9 'Handling Missing Data'), and sensitivity analyses (see Section 5.10 ' Sensitivity Analyses') will be conducted which examine the robustness of the results and assumptions (Bell et al., 2014).

3.3.1 Interim Analysis Population (IAP)

Interim analyses are conducted sequentially during the ongoing trial and are focused on the primary outcome. Therefore, the **Interim Analysis Population (IAP)** set includes all participants who were randomised and have passed the primary outcome (Day 28 post first intervention/active control session/post randomisation in TAU arm) with their primary outcome status either known or known to be missing at the time of interim analysis.

Since the 7-day daily intrusive memory diary at week 4 can also be filled in by participants retrospectively, we choose a 2-week cut off point where we declare the data as missing for the **IAP**. In other words, the **IAP** includes those who have:

- Completed all days of the 7-day daily intrusive memory diary
- Additionally, will include participants who have some/all missing days remaining post the 2-week timeframe (> 14 days have passed since the day the value was initially required to be filled in).
- Additionally, will include those who have been randomised but have dropped out/withdrew of the study, as this data will be classed as known to be missing

Any participant who has some/all missing days, but the 2-week timeframe window for each day to be declared missing has not yet been passed, will be excluded from the **IAP.**

3.3.2 End-of-Study Analysis Population (EAP)

The end-of study analysis is conducted after trial completion (following the last participant completing their final online assessment) and is focused on both primary and secondary outcomes. Therefore, the **End-of-study Analysis Population** (**EAP**) set includes all randomised participants.

3.3.3 Per-Protocol Population (PPP)

The **Per-Protocol Population** is a subset of the EAP that consists of subjects for whom there are no significant adherence and protocol deviations according to the definition in Section 3.2 ‘Adherence and Protocol Deviations’. The PPP is used for sensitivity analyses for the primary outcome (see Table 3) at the end of the study (see Section 5.10.3 ‘ Per Protocol’). The PPP will be defined at the Data Review Meeting (see Section 3.2.2 ‘Protocol Deviation’).

3.3.4 Complete-Case Population (CCP)

The **Complete-Case Population** is a subset of the EAP restricted to individuals with no missing days in their intrusive memory daily diary (IMDD) at baseline or week 4, regardless of

protocol deviations. The CCP are used for sensitivity analyses to assess impact of any missing data imputation methods used during interim analyses and end-of-study analyses of the primary outcome (see Section 5.10.4 ‘ Complete Case’).

4 TRIAL POPULATION

4.1 Screening Data

The number of patients screened will be presented in CONSORT diagrams.

4.2 Eligibility

4.2.1 Inclusion Criteria

Potential participants will be included if they meet the following criteria:

- Aged 18 or above.
- Able to read, write and speak in English.
- Worked in a clinical role with COVID-19 patients in the NHS during the COVID-19 pandemic
- Experienced at least one traumatic event related to their clinical work during the COVID-19 pandemic meeting criterion A of the Diagnostic and Statistical Manual of Mental Disorders, 5th edition (DSM-5) criteria for Post-Traumatic Stress Disorder (PTSD): “exposure to actual or threatened death, serious injury, or sexual violence” by “directly experiencing the traumatic event(s)” or “witnessing, in person, the event(s) as it occurred to others”.
- Experience intrusive memories of the traumatic event(s).
- Experienced at least three intrusive memories in the week prior to screening.
- Have internet access.
- Willing and able to provide informed consent and complete study procedures
- Willing and able to be contacted by the research team during the study period.

- Have not taken part in a previous study of this intervention from this research team (e.g. GAINS-01)

4.2.2 Exclusion Criteria

Potential participants will be excluded if they have fewer than three intrusive memories during the run-in week. We will not exclude those undergoing other treatment for post-traumatic stress disorder (PTSD) or its symptoms, so the study is as inclusive as possible to meet the challenges NHS staff are facing during their work related to the COVID-19 pandemic.

4.3 Recruitment

A CONSORT flow diagram will be used to summarise the number of patients who were:

- assessed for eligibility at screening
    - eligible at screening
    - ineligible at screening*
- eligible and randomised
- eligible but not randomised*
- received the randomised allocation
- did not receive the randomised allocation*
- completed week 4 follow up
- did not complete week 4 follow up*
- completed week 12 follow up
- did not complete week 12 follow up*
- completed week 24 follow up
- did not complete week 24 follow up*
- analysis populations

*Reason will be provided in CONSORT Diagram

4.4 Withdrawal and Follow-Up

The numbers (with reasons) of losses to follow-up (drop-outs and withdrawals) over the course of the trial will be summarised by treatment arm and presented in CONSORT diagram

4.5 Baseline Patient Characteristics

Prior to randomisation, participants complete baseline questionnaires (see Figure **1**). From these questionnaires, baseline characteristics will be summarised for each treatment arm and overall, across the arms. Discrete variables will be summarised by frequencies and percentages. Continuous variables will be summarised by use of standard measures of central tendency and dispersion using mean and standard deviation and/or median and first and third quartiles. Free text information will be assessed and assigned to a category if appropriate. No testing will be performed for differences in baseline characteristics between treatment arms. Unless otherwise stated, percentages are calculated according to the number of participants for whom data are available with the denominator stated.

Table 2 shows the baseline characteristics that will be summarised.

Table 2: Baseline measures

Baseline measure	Scale description	Items & scoring
Demographic Information	8-item questionnaire assessing:	
	1. What is your age (in years)?	1. Number (>= 18)
	2. What is your gender identity?	2. Categorical response (7 options) & free text if 'Other' category is specified
	3. What is your highest completed education level?	3. Categorical response (6 options)
	4. How would you describe your marital status:	4. Categorical response (6 options) & free text if 'Other' category is specified
	5. How would you describe your ethnicity (Asian, Black, Chines, Mixed, White or Other)?	5. Free text response (which will be categorised into 6 groups)
	6. What is your employment status?	6. Categorical response (7 options) & free text if 'Other' category is specified

	7. If you are working, how many hours per week do you currently work? 8. For how many years have you have worked as a healthcare professional? Additional demographic information (via electronic source document): Job role during pandemic	7. Number (0-168) 8. Number (0-99) Job role during pandemic: Categorical response (7 options)
PTSD Checklist for DSM-5	20-item measure assessing: symptoms of PTSD over the last week. 1. Repeated, disturbing, and unwanted memories of the stressful experience? 2. Repeated, disturbing dreams of the stressful experience? 3. Suddenly feeling or acting as if the stressful experience were actually happening again (as if you were actually back there reliving it)? 4. Feeling very upset when something reminded you of the stressful experience? 5. Having strong physical reactions when something reminded you of the stressful experience (for example, heart pounding, trouble breathing, sweating)? 6. Avoiding memories, thoughts, or feelings related to the stressful experience? 7. Avoiding external reminders of the stressful experience (for example, people, places, conversations, activities, objects, or situations)? 8. Trouble remembering important parts of the stressful experience? 9. Having strong negative beliefs about yourself, other people, or the world (for example, having thoughts such as: I am bad, there is something seriously wrong with me, no one can be trusted, the world is completely dangerous)?	Items are rated on a 5-point scale ranging from 0 (“not at all”) to 4 (“extremely”). Total Score: Scores are summed to give a total severity score (ranging 0 to 80).

	10. Blaming yourself or someone else for the stressful experience or what happened after it? 11. Having strong negative feelings such as fear, horror, anger, guilt, or shame? 12. Loss of interest in activities that you used to enjoy? 13. Feeling distant or cut off from other people? 14. Trouble experiencing positive feelings (for example, being unable to feel happiness or have loving feelings for people close to you)? 15. Irritable behavior, angry outbursts, or acting aggressively? 16. Taking too many risks or doing things that could cause you harm? 17. Being “superalert” or watchful or on guard? 18. Feeling jumpy or easily startled? 19. Having difficulty concentrating? 20. Trouble falling or staying asleep?	
Sleep Condition Indicator	2-item scale assessing: sleep problems against the DSM-5 criteria for insomnia disorder. 1. Thinking about the past month, to what extent has poor sleep troubled you in general? 2. Thinking about a typical night in the last month, how many nights a week do you have a problem with your sleep?	All items rated on 5-point scale, with scores from 0 to 2 indicating threshold criteria for insomnia disorder. 1. Rated on a 5-point scale ranging from 0 (“Very much”) to 4 (“Not at all”) 2. Rated on a 5-point scale ranging from 0 (“5-7 [nights]”) to 4 (“0-1 [nights]”) Total Score: Scores are summed to give a total score (ranging 0 to 8), with a higher score indicating better sleep.
Generalized Anxiety Disorder	2-item measure assessing: the severity of anxiety symptoms. Over the last 2 weeks, how often have you been bothered by the following problems? 1. Feeling nervous, anxious or on edge	1. Rated on a 4-point scale for how often they have bothered the respondent over the last

		two weeks, from 0 (“not at all”) to 3 (“nearly every day”). 2. Rated on a 4-point scale for how often they have bothered the respondent over the last two weeks, from 0 (“not at all”) to 3 (“nearly every day”). Total Score: Scores are summed to give a total score (ranging 0 to 6), with a cut-off score of 3 indicating a probable diagnosis of generalised anxiety disorder.
	2. Not being able to stop or control worrying	
Patient Health Questionnaire	2-item measure assessing: symptoms of depression. Over the last 2 weeks, how often have you been bothered by the following problems? 1. Little interest or pleasure in doing things 2. Feeling down, depressed, or hopeless	1. Rated for how often they have bothered the respondent over the last two weeks, from 0 (“not at all”) to 3 (“nearly every day”). 2. Rated for how often they have bothered the respondent over the last two weeks, from 0 (“not at all”) to 3 (“nearly every day”). Total Score: Scores are summed to give a total score (ranging 0 to 6), with a cut-off score of 3 indicating a probable diagnosis of depression.
Scale of work engagement and burnout	19-item measure, with 2 main subscales assessing: work engagement and burnout. 1. Engagement subscale (10 items from three dimensions: vigour, attentiveness, dedication) • Vigour subscale (3 items) • Dedication subscale (3 items) • Attentiveness subscale (4 items) 2. Burnout subscale (9 items from three dimensions: exhaustion, disengagement and inattentiveness). • Exhaustion subscale (3 items) • Disengagement subscale (3 items) • Inattentiveness subscale (3 items)	Each subscale includes multiple items rated on a 4-point scale of how often they have felt each descriptive in the past two weeks, from 1 (not at all) to 4 (all the time). The mean score of the Items under category is calculated to give the subscale score.
Sickness absence	A single item measure assessing: the number of sick days taken from work during the past 4 weeks	Number (0-28)

	During the past 4 weeks, how many sick days have you taken from work?	
Intention to leave job	3-item measure assessing: participants' intention to leave their job How much do you agree with the following? 1. I think a lot about leaving the job 2. I am actively searching for an alternative to the job 3. As soon as it is possible, I will leave the job	Each item is rated on a 5-point scale from 1 (strongly agree) to 5 (strongly disagree). Total Score: Scores are summed to give a total score (ranging 3 to 15), with a lower score indicating stronger intention to leave the job.
EQ-5D-5L	5-items assessing: general quality of life and health status 1. Mobility 2. Self-care 3. Usual activities 4. Pain/discomfort 5. Anxiety/depression VAS score: Overall health	All items 1-5 are rated on a 5-point scale 1. Items rated from 1 ("I have no problem in walking about") to 5 ("I am unable to walk about") 2. Items are rated from 1 ("I have no problems washing or dressing myself") to 5 ("I am unable to wash or dress myself") 3. Items are rated from 1 ("I have no problems doing my usual activities") to 5 ("I am unable to do my usual activities") 4. Items are rated from 1 ("I have no pain or discomfort") to 5 ("I have extreme pain or discomfort") 5. Items are rated from 1 ("I am not anxious or depressed") to 5 ("I am extremely anxious or depressed") VAS Score: rating for their overall health today from 0 (the worst health you can imagine) to 100 (the best health you can imagine).
World Health Organization Disability Assessment Schedule	12-item, self-report version of the WHODAS 2.0 assessing difficulties in relation to the impact of intrusive memories. Items 1-12 assessing how much e.g. "how much difficulty did you have in standing for long periods, such as 30 minutes?"	Items 1-12 are rated for how much difficulty they have had in each area in the past 30 days, from 0 (none) to 4 (extreme or cannot do). Overall Score: Scores are summed and divided by 48 to give overall % score (ranging 0.00% to 100.00%)

	Further 3 items assessing how many days there were difficulties in relation to the impact of intrusive memories	Further 3 items are scored as numbers (0-30)
Health background	2-item questionnaire assessing: 1. Have you been treated for/diagnosed with any mental health problems e.g. depression, anxiety, post-traumatic stress disorder? 2. [If yes]: Are you receiving any current treatments/medications for these e.g. psychological treatment, counselling, medication, alternative therapies?	1. Yes/No (binary response), and if the participant responds 'Yes', they are asked to list the mental health problems (free text response) and indicate Past/Current (binary response) for each 2. Yes/No (binary response), and if the participant responds 'Yes', they are asked to give brief details (open text response)
Checklist of work-related traumatic events	4-item checklist assessing: the types of traumatic events, they have experienced or witnessed during the COVID-19 pandemic, for which they have intrusive memories 1. List of traumatic events previously reported by healthcare professionals in research literature, including a traumatic or tragic death of a patient, increased risk of COVID-19 infection, and severe or unsuccessful resuscitation. 2. Time frame and ongoing exposure 3. How many work-related traumatic events during the COVID-19? 4. How many non-work-related traumatic events they experienced during the pandemic?	1. Yes/No (binary response) to each item on list, if 'Yes' to 'Other' then asked to give brief details (free text) 2. Yes/No (binary response) to each time frame 3. Number (0-999) 4. Number (0-999)

4.6 Additional Questionnaires

Baseline measure	Scale description	Items & scoring
Credibility/Expectancy Questionnaire	6-item questionnaire will assess participants' belief that the intervention will help reduce their intrusive memories. 1. At this point, how logical does the intervention offered to you seem?	1. Rated on 9-point scale from 1 (not at all), above 5 (somewhat) to 9 (very)

	2. At this point, how successful do you think this intervention will be in reducing your intrusive memories?	2. Rated on 9-point scale from 1 (not at all), above 5 (somewhat) to 9 (very)
	3. How confident would you be in recommending this intervention to a friend who experiences similar problems?	3. Rated on 9-point scale from 1 (not at all), above 5 (somewhat) to 9 (very)
	4. By the end of the intervention period (four weeks), how much improvement in your intrusive memories do you think will occur?	4. Rated as a percentage from 0% to 100%, in intervals of 10%.
	5. At this point, how much do you really feel that the intervention will help you to reduce your intrusive memories?	5. Rated on 9-point scale from 1 (not at all), above 5 (somewhat) to 9 (very)
	6. By the end of the intervention period (four weeks), how much improvement in your intrusive memories do you really feel will occur?	6. Rated as a percentage from 0% to 100%, in intervals of 10%.

5 ANALYSIS

5.1 Outcome Definitions

Table 3: Primary, Secondary, and Exploratory outcomes

Outcome measure	Scale description	Items & scoring	Week			
			Baseline	4	12	24
Primary outcome						
Number of intrusive memories of traumatic event(s) during week 4	Number of intrusive memories of traumatic event(s) recorded by participants in a brief diary daily for 7 days during week 4	Total number of intrusive memories of traumatic events recorded by participants in a brief daily online intrusive memory diary for 7 days during week 4 (i.e., summing from Day 22 to 28 post first intervention/control session/post randomisation in TAU arm)	X	X		
Secondary outcomes						

Number of intrusive memories of traumatic event(s) during week 12	Number of intrusive memories of traumatic event(s) recorded by participants in a brief diary daily for 7 days during week 12	Total number of intrusive memories of traumatic events recorded by participants in a brief daily online intrusive memory diary for 7 days during week 12 (i.e. summing from Day 78 to 84), Total number of intrusive memories of traumatic events recorded by participants in a brief daily online intrusive memory diary for 7 days during run-in week.	X		X	
Number of intrusive memories of traumatic event(s) during week 24	Number of intrusive memories of traumatic event(s) recorded by participants in a brief diary daily for 7 days during week 24	Total number of intrusive memories of traumatic events recorded by participants in a brief daily online intrusive memory diary for 7 days during week 24 (i.e. summing from Day 162 to 168). Total number of intrusive memories of traumatic events recorded by participants in a brief daily online intrusive memory diary for 7 days during run-in week.	X			X
PTSD Checklist for DSM-5	20-item measure assessing: symptoms of PTSD over the last week. 1. Repeated, disturbing, and unwanted memories of the stressful experience? 2. Repeated, disturbing dreams of the stressful experience? 3. Suddenly feeling or acting as if the stressful experience were actually happening again (as if you were actually back there reliving it)? 4. Feeling very upset when something reminded you of the stressful experience? 5. Having strong physical reactions when something reminded you of the stressful experience (for example, heart pounding, trouble breathing, sweating)? 6. Avoiding memories, thoughts, or feelings related to the stressful experience? 7. Avoiding external reminders of the stressful experience (for example, people, places, conversations, activities, objects, or situations)?	Items are rated on a 5-point scale ranging from 0 (“not at all”) to 4 (“extremely”). Total Score: Scores are summed to give a total severity score (ranging 0 to 80).	X	X	X	X

	8. Trouble remembering important parts of the stressful experience? 9. Having strong negative beliefs about yourself, other people, or the world (for example, having thoughts such as: I am bad, there is something seriously wrong with me, no one can be trusted, the world is completely dangerous)? 10. Blaming yourself or someone else for the stressful experience or what happened after it? 11. Having strong negative feelings such as fear, horror, anger, guilt, or shame? 12. Loss of interest in activities that you used to enjoy? 13. Feeling distant or cut off from other people? 14. Trouble experiencing positive feelings (for example, being unable to feel happiness or have loving feelings for people close to you)? 15. Irritable behavior, angry outbursts, or acting aggressively? 16. Taking too many risks or doing things that could cause you harm? 17. Being “super alert” or watchful or on guard? 18. Feeling jumpy or easily startled? 19. Having difficulty concentrating? 20. Trouble falling or staying asleep?					
Sleep Condition Indicator	2-item scale assessing: sleep problems against the DSM-5 criteria for insomnia disorder. 1. Thinking about the past month, to what extent has poor sleep troubled you in general? 2. Thinking about a typical night in the last month, how many nights a week do you have a problem with your sleep?	All items rated on 5-point scale, with scores from 0 to 2 indicating threshold criteria for insomnia disorder. 1. Rated on a 5-point scale ranging from 0 (“Very much”) to 4 (“Not at all”) 2. Rated on a 5-point scale ranging from 0 (“5-7 [nights]”) to 4 (“0-1 [nights]”)	X	X	X	X

		Total Score: Scores are summed to give a total score (ranging 0 to 8), with a higher score indicating better sleep.				
Generalized Anxiety Disorder	2-item measure assessing: the severity of anxiety symptoms. Over the last 2 weeks, how often have you been bothered by the following problems? 1. Feeling nervous, anxious or on edge 2. Not being able to stop or control worrying	1. Rated on a 4-point scale for how often they have bothered the respondent over the last two weeks, from 0 (“not at all”) to 3 (“nearly every day”). 2. Rated on a 4-point scale for how often they have bothered the respondent over the last two weeks, from 0 (“not at all”) to 3 (“nearly every day”). Total Score: Scores are summed to give a total score (ranging 0 to 6), with a cut-off score of 3 indicating a probable diagnosis of generalised anxiety disorder.	X	X	X	X
Patient Health Questionnaire	2-item measure assessing: symptoms of depression. Over the last 2 weeks, how often have you been bothered by the following problems? 1. Little interest or pleasure in doing things. 2. Feeling down, depressed, or hopeless.	1. Rated for how often they have bothered the respondent over the last two weeks, from 0 (“not at all”) to 3 (“nearly every day”). 2. Rated for how often they have bothered the respondent over the last two weeks, from 0 (“not at all”) to 3 (“nearly every day”). Total Score: Scores are summed to give a total score (ranging 0 to 6), with a cut-off score of 3 indicating a probable diagnosis of depression.	X	X	X	X
Scale of work engagement and burnout	19-item measure, with 2 main subscales assessing: work engagement and burnout. 1. Engagement subscale (10 items from three dimensions: vigour, attentiveness, dedication) • Vigour subscale (3 items) • Dedication subscale (3 items)	Each subscale includes multiple items rated on a 4-point scale of how often they have felt each descriptive in the past two weeks, from 1 (not at all) to 4 (all the time). The mean score of the items under category is calculated to give the subscale score.	X	X	X	X

	• Attentiveness subscale (4 items) 2. Burnout subscale (9 items from three dimensions: exhaustion, disengagement and inattentiveness). • Exhaustion subscale (3 items) • Disengagement subscale (3 items) • Inattentiveness subscale (3 items)					
Sickness absence	A single item measure assessing: the number of sick days taken from work during the past 4 weeks During the past 4 weeks, how many sick days have you taken from work?	Number (0-28)	X	X	X	X
Intention to leave job	3-item measure assessing: participants' intention to leave their job How much do you agree with the following? 1. I think a lot about leaving the job 2. I am actively searching for an alternative to the job As soon as it is possible, I will leave the job	Each item is rated on a 5-point scale from 1 (strongly agree) to 5 (strongly disagree). Total Score: Scores are summed to give a total score (ranging 3 to 15), with a lower score indicating stronger intention to leave the job.	X	X	X	X
EQ-5D-5L	5-items assessing: general quality of life and health status 1. Mobility 2. Self-care 3. Usual activities	All items 1-5 are rated on a 5-point scale 1. Items rated from 1 ("I have no problem in walking about") to 5 ("I am unable to walk about") 2. Items are rated from 1 ("I have no problems washing or dressing myself") to 5 ("I am unable to wash or dress myself") 3. Items are rated from 1 ("I have no problems doing my usual activities") to 5 ("I am unable to do my usual activities")	X	X	X	X

	4. Pain/discomfort 5. Anxiety/depression VAS score: Overall health	4. Items are rated from 1 ("I have no pain or discomfort") to 5 ("I have extreme pain or discomfort") 5. Items are rated from 1 ("I am not anxious or depressed") to 5 ("I am extremely anxious or depressed") VAS Score: rating for their overall health today from 0 (the worst health you can imagine) to 100 (the best health you can imagine).				
World Health Organization Disability Assessment Schedule	12-item, self-report version of the WHODAS 2.0 assessing difficulties in relation to the impact of intrusive memories. Items 1-12 assessing how much e.g. "how much difficulty did you have in standing for long periods, such as 30 minutes?" Further 3 items assessing how many days there were difficulties in relation to the impact of intrusive memories.	Items 1-12 are rated for how much difficulty they have had in each area in the past 30 days, from 0 (none) to 4 (extreme or cannot do). Overall Score: Scores are summed and divided by 48 to give overall % score (ranging 0.00% to 100.00%) Further 3 items are scored as numbers (0-30)	X		X	X
Intrusive memory ratings	8-item measure assessing: the following characteristics of their intrusive memories over the last week: 1. Frequency & how many times per day 2. Distress 3. Disruption to concentration 4. Interference with what they were doing & for how long each time 5. impact on work functioning (how much & in what ways) 6. impact on functioning in other areas of life (how much & in what ways)	Items are examined separately (not summed) and scored as follows: 1. 7-point categorical response from 'never' to 'many times a day' & Number (0-99) 2. 11-point ordinal response from 'not at all' (0) to 'extremely' (10) 3. 11-point ordinal response from 'not at all' (0) to 'very much (10) 4. 11-point ordinal response from 'not at all' (0) to 'very much (10) & 6-point categorical response from '<1min' to '>60mins' 5. 11-point ordinal response from 'not at all' (0) to 'very much (10) & free text response 6. 11-point ordinal response from 'not at all' (0) to 'very	X		X	X

		much (10) & free text response			
	Two additional items: 7. the number of days worked 8. number of night shifts worked in the last week.	7. Number (0-7) 8. Number (0-7)			
Exploratory outcomes					
Changes to Health and Work	9-item questionnaire assessing: 1. New work-related traumatic events 2. Types of new work-related traumatic events 3. The number of new work-related traumatic events 4. The number of new non-work-related traumatic events 5. Additional stressful life events 6. New treatments received 7. The occurrence of adverse events 8. Changes to the job, 9. Changes to the number of hours worked	1. Yes/No (binary response) 2. Yes/No (binary response) for a series of 10 different types of traumatic events, if Yes to 'Other' type of traumatic event, then free text response to give details 3. Number (0-99) 4. Number (0-99) 5. Yes/No (binary response), If 'Yes' free test response to give details 6. Yes/No (binary response), If 'Yes' free test response to give details 7. Yes/No (binary response), If 'Yes' free test response to give details 8. Yes/No (binary response), If 'Yes' free test response to give details 9. Yes/No (binary response), If 'Yes' free test response to give details	X	X	X
Feedback Questionnaire*	A 12-item questionnaire assessing: participants' experience of using the brief cognitive task 1. Easy participants found the brief cognitive task 2. Helpful participants found the brief cognitive task 3. Burdensome participants found the brief cognitive task 4. Distressing participants found the brief cognitive task 5. Acceptable participants found the brief cognitive task 6. What was useful 7. What was difficult 8. How willing they would be to use it in the future 9. How confident they would be in recommending it to a friend	Item 1-5, and 8-10 are rated on 11-point scale from 0 (not at all) to 10 (very) Item 6, 7, 11, and 12 are free text responses Total Score: Numerical Scores are summed to give a total score (ranging 0 to 80).	X		

	10. How much they feel it could be used to support staff within the NHS 11. How could it be improved 12. Other suggestions			
Intervention/control use	e.g. • Number of times used • Duration • Time of day	X	X	X

* Feedback questionnaire only filled out by Intervention and Active Control arms

5.2 Accuracy Ratings

The outcome, number of intrusive memories of traumatic event(s), is recorded by participants in a brief intrusive memory diary daily for 7 days (see Section 5.1 'Outcome Definitions'). Each day, the participant is asked "Have you had any intrusive memories on [date]?" (yes/no response) and if yes, "Please select how many intrusive memories you have had" (participant selects a number). At the end of the week, they will be asked to rate how accurately they think they completed the diary over the last week (on an 11-point scale from 0 = not at all accurately to 10 = extremely accurately).

The diary accuracy rating measure (see Table 4) will be summarised along with all baseline and outcome measures. This will be used for comparability with previous research.

Table 4: Intrusive Memory Diary Accuracy Rating

Compliance measure	Scale description	Items & scoring	Week			
			Baseline	4	12	24
Diary accuracy rating	Single item assessing self-reported accuracy of completing the 7-day intrusive memory diary	11-point ordinal scale from 0 (not at all accurately) to 10 (extremely accurately)	X	X	X	X

5.3 Descriptive Summaries

Descriptive summaries of all baseline measures will be presented as described in Section 4.5 'Baseline Patient Characteristics'. Similarly, all outcomes (including accuracy ratings, see Section 5.2 'Accuracy Ratings') will be summarised for each treatment arm and overall. At interim analyses, descriptive summaries of the IAP will be presented. At End-of-study

analyses, descriptive summaries of the EAP population (all randomised) will be presented (see Section 3.3 'Analysis Populations').

5.4 Sample Size Determination and Early Stopping

We employ a Sequential Bayes Factor design with maximal sample size (Schönbrodt et al., 2017; Schönbrodt & Wagenmakers, 2018). Sequential analyses are based on computation of BFs comparing the intervention arm to the TAU arm and comparing intervention arm to the active control arm for the primary outcome (total number of intrusive memories recorded over seven days in week 4, controlling for baseline number of intrusive memories). Therefore, BFs are computed repeatedly during data collection, enabling repeated comparisons during data collection, until the BFs exceed an a priori defined grade of evidence, or the maximal sample size (see Section 2.3 'Sample Size') of 150 has been reached (see Section 5.4.1 'Decision Rules').

The BF is either denoted as BF_{10} ("H_1 over H_0"), or as its inverse BF_{01} ("H_0 over H_1,"), for the alternative hypothesis H_1 and null hypothesis H_0. When the Bayes factor BF_{10} equals five, this indicates that the data are five times more likely under H_1 than under H_0, meaning that H_1 has issued a better probabilistic prediction for the observed data than did H_0 (see Schönbrodt & Wagenmakers, 2018). Conversely, if the Bayes factor BF_{01}(equivalent to $\frac{1}{BF_{10}}$ **)**, equals five, this would indicate that the data observed were five times more probable if H_0 were true than if the H_1 were true.

Guidelines given by Jeffreys (1998) and Lee & Wagenmakers, (2013) label that if 1 < BF < 3, the BF indicates anecdotal evidence, 3 < BF < 10 moderate evidence, 10 < BF < 30 strong evidence, and BF > 30 very strong evidence. Kass & Raftery (1995) suggest 20 as a threshold for 'strong evidence'. Given these guidelines, we choose 20 to be our threshold for strong evidence.

Interim analysis is conducted sequentially on the primary outcome, please refer to Section 5.5 ‘Analyses of Primary ’ for details of the modelling.

5.4.1 Decision Rules

For each sequential interim analyses, we conduct between group analyses of primary outcome for Intervention vs TAU, and between group analyses of primary outcome for Intervention vs Active Control.

We first calculate Bayes factors for a negative effect of the intervention (i.e., whether participants in the intervention arm have a greater number of intrusive memories at week 4 than the TAU arm, followed by comparison with the Active Control arm). The decision rule for stopping for negative effect will be based on the comparison with TAU arm; if this BF exceeds 20, we conclude that there is strong evidence for a negative effect of the intervention and the trial may need to be altered or stopped.

We then calculate Bayes factors for positive treatment effect (i.e., whether those in the intervention arm have fewer intrusive memories at week 4 than the TAU arm, followed by comparison with the Active Control arm). The decision rule for stopping for effectiveness will be based on the comparison with Active Control arm; if this BF exceeds 20, we conclude that there is strong evidence for the effectiveness of the intervention and consideration will be given to stopping the trial early.

A decision to discontinue recruitment will be made in consultation with the steering committee and DMC members. Please refer to Section 5.5 ‘Analyses of Primary Objective’ for details of the modelling and the exact hypothesis tests used for negative and positive effects.

5.5 Analyses of Primary Outcome

A between-groups analysis will be used to test the difference in the number of intrusive memories in week 4 (i.e., from Day 22 to 28) between the intervention, active control, and TAU groups. The parameters being tested will be β_2 and β_3 in the model below (Section 5.5.1

'Modelling Assumptions') which represent the difference in the log number of memories between the intervention and TAU or active control respectively. The analysis will control for the number of intrusive memories during the run-in week as a fixed effect and include participant as a random effect.

For the primary outcome, we use Bayesian multi-level regression models implemented using *brms* package in R (Bürkner, 2017, 2018, 2021). Bayes factors are used to compare the null and alternative hypotheses (see Section 5.5.3 'Hypotheses') pertaining to the comparisons between the intervention arm and the other arms.

5.5.1 Modelling Assumptions

The primary outcome (intrusive memory count) will be modelled using a Poisson linear mixed model (Observation Level Random Effects Poisson, Harrison, 2014). The baseline number of intrusive memories, and treatment arm are fitted as fixed effects with a random intercept effect for participant.

Let n be the number of participants which we are analysing, then for each participant $i = 1, \dots, n$ we denote:

- Y_i to be the random variable representing the primary outcome (total number of intrusive memories in week 4).
- $Baseline_i$ to be the total number of intrusive memories recorded during the run-in week.
- The three arms (Intervention, TAU, Active Control) are indicated by dichotomous variables I_i, TAU_i, AC_i which equal one if participant i was randomised to the respective arm and zero otherwise.
- $subj_i$ denotes a dichotomous variable, equal to one if the observation is from the subject i, and zero if not.

γ_{0i} is the random effect for the intercept for each participant i that accounts for the participant-specific variation in the primary outcome. The random-effects intercepts γ_{0i} are drawn from a

normal distribution with mean 0 and standard deviation σ_{γ_0}which is estimated by the model. The model is as follows, where $\alpha, \beta_1, \beta_2, \beta_3, \sigma_{\gamma_0}$ are our unknown model parameters.

$$Y_i \sim Poisson(\mu_i)$$

$$E(Y_i) = \mu_i$$

$$ln(\mu_i) = \alpha + Baseline_i\beta_1 + TAU_i\beta_2 + AC_i\beta_3 + \sum_{i=1}^{n} subj_i\gamma_{0i}$$

$$\gamma_{0i} \sim N(0, \sigma_{\gamma_0}^2)$$

with the following priors:

$$\alpha \sim t(3,2.5)$$

$$\beta_i \sim N(0,10)$$

$$\sigma_{\gamma_0} \sim t(3,2.5)$$

All fixed effect terms will use normal priors centred on zero with a standard deviation of 10. The standard deviation parameter σ_{γ_0} for the random effect will be restricted to be positive and takes a half student-t prior (with 3 degrees of freedom and a scale parameter 2.5).

Note that the arm variable I_i is taken as the reference level, therefore the estimated parameters β_2, β_3 are really contrasts comparing the respective arm to the reference level arm. In this context, the β_2, β_3 parameters represent the difference between the logs of the number of intrusive memories between the respective arms (TAU, and Active Control) in comparison to the intervention arm; specifically, the estimates of the covariates being > 0 would indicate a positive treatment effect in the direction of the intervention as per the model formulation. Testing the difference between the logs tests the primary objective pertaining to the difference in the number of intrusive memories as if log (x) > log (y) then x > y.

Through visual and quantitative analysis, the validity of any model assumptions will be examined. This will include examining plots of residuals to check for normality and homogeneity of variance, and checking for zero inflation and overdispersion in the Poisson model. When our count response variable is too over dispersed, or zero-inflated, and violates

the Poisson regression assumptions, we will consider using negative binomial or zero-inflated models as alternatives. Where other models are considered, we will perform model comparison using performance measures such as Bayes factors, AIC and BIC, and root mean squared error. The most appropriate model which corrects for any zero inflation and overdispersion present will be used, selected via comparison of performance measures (such as, BF, AIC, and RMSE). Posterior predictive checks will be used to qualitatively assess the performance of the model. This involves generating a predictive distribution of the data based on the posterior distribution of the parameters and comparing the results to the observed data.

5.5.2 Priors

Bayesian inference requires the specification of appropriate prior distributions on model parameters. Therefore, we must express our prior understanding of regression coefficients ($\alpha, \beta_1, \beta_2, \beta_3, \sigma_{\gamma_0}$).

Specifically, for the intercept α and the fixed effect estimates $\beta_1, \beta_2, \beta_3$, since these coefficients can each take on any value on the real line, it is reasonable to utilize normal priors. We assume a normal distribution with mean zero and standard deviation of σ. Assuming a mean of zero means that we do not make any assumption a priori that the effect differs from zero. If the effect should differ from zero, we want the data to tell us that. The standard deviation will be set as σ = 10; varying standard deviations will be tested as sensitivity analysis (refer to Section 5.10.1 'Prior distributions').

The default priors set within the *brms* package (v2.19.0) will used for any other model parameters such as the group-level effects, or any family specific parameters. For example, the standard deviation parameter σ_{γ_0} for the random effect will be restricted to be positive and takes a half student-t prior (with 3 degrees of freedom and a scale parameter 2.5). A negative binomial model will need a prior for the shape parameter which by default is set to a gamma distribution (shape=0.01, scale=0.01)

If issues with convergence or sampling arise, variations to these priors are permissible however, all model variations and their justification will be disclosed in all internal and external publications.

The impact of other prior forms will be examined through a sensitivity analysis. We examine how robust the results are when the priors are altered, and the model is re-estimated (see Section 5.10 ‘ Sensitivity Analyses’)

5.5.3 Hypotheses

5.5.3.1 Positive Treatment Effect

Intervention vs TAU:

$H_0: \beta_2 = 0$	*The intervention has no effect (participants in the intervention arm, compared to the TAU arm, will have an equal number of intrusive memories during week 4)*
$H_1: \beta_2 > 0$	*The intervention has a positive effect (participants in the intervention arm, compared to the TAU arm, will have fewer intrusive memories during week 4)*

Intervention vs Active Control:

$H_0: \beta_3 = 0$	*The intervention has no effect (participants in the intervention arm, compared to the active control arm, will have an equal number of intrusive memories during week 4)*
$H_1: \beta_3 > 0$	*The intervention has a positive effect (participants in the intervention arm, compared to the active control arm, will have fewer intrusive memories during week 4)*

5.5.3.2 Negative Effect

Intervention vs TAU:

H_0: $\beta_2 \geq 0$	*The intervention has no effect or a positive effect (participants in the intervention arm, compared to the TAU arm, will have equal or fewer intrusive memories)*
H_1: $\beta_2 < 0$	*The intervention has a negative effect (participants in the intervention arm, compared to the TAU arm, will have a greater number of intrusive memories in week 4)*

Intervention vs Active Control:

H_0: $\beta_3 \geq 0$	*The intervention has no effect or a positive effect (participants in the intervention arm, compared to the active control arm, will have equal or fewer intrusive memories)*
H_1: $\beta_3 < 0$	*The intervention has a negative effect (participants in the intervention arm, compared to the active control arm, will have a greater number of intrusive memories in week 4)*

The Bayes factors here represent the degree to which the posterior distribution has moved from the null relative to the prior distribution.

5.6 Analyses of Secondary Outcomes

The analysis of all secondary outcomes detailed in Table 3 will be conducted only in the final analysis and within a Bayesian inference framework. An outline of the modelling approach for each of the different data types are given below. Hypotheses will be set up to test for between group differences for Intervention and TAU, and Intervention and Active Control. Modelling assumptions will be evaluated using exploratory analysis for all secondary modelling, with alternative models used when appropriate (e.g., multivariate distributions, negative binomial or zero-inflated models for count data).

5.6.1 Discrete Outcomes

To model discrete outcomes, such as the number of intrusive memories at week 12 and week 24, we will use the same framework as detailed in Section 5.5 ‘Analyses of Primary Outcome’.

The baseline number of intrusive memories, and treatment arm are fitted as fixed effects with a random intercept effect for participant; additionally, we include fixed effects for week (week 4, 12, 24) and for interactions between treatment arm and week.

Let n be the number of participants we are analysing, then for each participant $i = 1, \dots, n$ and week j (4, 12 or 24),

- Y_i to be the random variable representing the outcome (e.g., total number of intrusive memories in week x).
- $Baseline_i$ to be the total number of intrusive memories recorded during the run-in week.
- The three arms (Intervention, TAU, Active Control) are indicated by dichotomous variables I_i, TAU_i, AC_i which equal one if participant i was randomised to the respective arm and zero otherwise.
- t_x is the dummy variable for the time point week x ($t_x = 1 \text{if } x = j$, and 0 otherwise); Week 4 is taken as a reference level, $x$and can take the value of either 12 or 24.
- $subj_i$ denotes a dichotomous variable, equal to one if the observation is from the subject, and zero if not.

  γ_{0i} is the random effect for the intercept for each participant i that accounts for the participant-specific variation in the outcomes. The random-effects intercepts γ_{0i} are drawn from a normal distribution with mean 0 and standard deviation σ_{γ_0}which is estimated by the model.

The model is as follows, where $\alpha, \beta_{1-9}, \sigma_{\gamma_0}$ are our unknown model parameters.

$$Y_{ij} \sim Poisson(\mu_{ij})$$

$$E(Y_{ij}) = \mu_{ij}$$

$$ln(\mu_{ij}) = \alpha + Baseline_i\beta_1 + TAU_i\beta_2 + AC_i\beta_3 + t_{12}\beta_4 + t_{24}\beta_5 + TAU_i t_{12}\beta_6 + TAU_i t_{24}\beta_7 + AC_i t_{12}\beta_8 + AC_i t_{24}\beta_9 + \sum_{i=1}^{n} subj_i\gamma_{0i}$$

$$\gamma_{0i} \sim N(0, \sigma_{\gamma_0}^2)$$

with the following priors:

$$\alpha \sim t(3,2.5)$$

$$\beta_i \sim N(0,10)$$

$$\sigma_{\gamma_0} \sim t(3,2.5)$$

All fixed effect terms will use normal priors centred on zero with a standard deviation of 10. The standard deviation parameter σ_{γ_0} for the random effect will be restricted to be positive and takes a half student-t prior (with 3 degrees of freedom and a scale parameter 2.5).

Posterior predictive checks will be used to qualitatively assess the performance of the model. This involves generating a predictive distribution of the data based on the posterior distribution of the parameters and comparing the results to the observed data.

This framework will also be used to model the number of sick days taken in the previous 4 weeks by subjects, as well as the further 3 items on the WHODAS 2.0 questionnaire (see Section 5.1).

5.6.2 Continuous outcomes

There are secondary outcomes which are on a continuous scale (although they may be theoretically discrete or ordinal, when the range of possibilities exceeds 10 with uniform distance between each, the data can be assumed continuous for the purpose of modelling). Linear regression using a normal distribution to model the likelihood of observations and the identity link function is a standard model to use for continuous outcomes. This will be performed in the first instance for continuous outcomes, at which point residuals will be analysed. If the residuals are normally distributed, then the model will be assumed a reasonable fit and the parameters investigated. If the residuals do not meet the assumptions for linear regression and use of the normal distribution, alternative models will be considered. This may involve transforming the data, such as log transforming, or considering alternative probability distributions, such as gamma, and/or using alternative link functions. In the case of the WHODAS 2.0 overall score from the initial 12 questions and the SWEBO subscales, a

beta distribution with a logit link will be considered if the linear model does not accurately reflect the data.

The baseline outcome score, and treatment arm are fitted as fixed effects with a random intercept effect for participant; additionally, we include fixed effects for the week (week 4, 12, 24) and for the interactions between treatment arm and week.

Let n be the number of participants we are analysing, then for each participant $i,\ldots,n$ and week j (4, 12 or 24),

- Y_{ij} to be the random variable representing the outcome for participant i at week j.
- $Baseline_i$ to be the outcome score recorded during the run-in week.
- The three arms (Intervention, TAU, Active Control) are indicated by dichotomous variables I_i, TAU_i, AC_i which equal one if participant i was randomised to the respective arm and zero otherwise.
- t_x is the dummy variable for the time point week x ($t_x = 1 \text{if } j = x$, and 0 otherwise); Week 4 is taken as a reference level, $x$and can take the value of either 12 or 24.
- $subj_i$ denotes a dichotomous variable, equal to one if the observation is from the subject, and zero if not.

γ_{0i} is the random effect for the intercept for each participant i that accounts for the participant-specific variation in the outcomes. The random-effects intercepts γ_{0i} are drawn from a normal distribution with mean 0 and standard deviation σ_{γ_0}which is estimated by the model. σ^2 is the variance of the normal distribution from which Y_{ij} is drawn from.

The model is as follows, where $\alpha, \beta_{1-9}, \sigma_{\gamma_0}, \sigma^2$ are our unknown model parameters.

$$Y_{ij} \sim N(\mu_i, \sigma^2)$$

$$E(Y_{ij}) = \mu_{ij}$$

$$\mu_{ij} = \alpha + Baseline_i\beta_1 + TAU_i\beta_2 + AC_i\beta_3 + t_{12}\beta_4 + t_{24}\beta_5 + TAU_i t_{12}\beta_6 + TAU_i t_{24}\beta_7 + AC_i t_{12}\beta_8 + AC_i t_{24}\beta_9 + \sum_{i=1}^{n} subj_i\gamma_{0i}$$

$$\gamma_{0i} \sim N\left(0, \sigma_{\gamma_0}^2\right)$$

with the following priors:

$$\alpha \sim t(3,2.5)$$

$$\beta_i \sim N(0,10)$$

$$\sigma_{\gamma_0} \sim t(3,2.5)$$

$$\sigma \sim t(3,2.5)$$

All fixed effect terms will use normal priors centred on zero with a standard deviation of 10. The standard deviation parameters σ_{γ_0} for the random effect, and σ for Y_{ij}, will be restricted to be positive and takes a half student-t prior (with 3 degrees of freedom and a scale parameter 2.5). If a gamma distribution is used, the default gamma prior for the shape and scale parameters from *brms* will be used, a gamma distribution (alpha = 0.01, beta = 0.01).

Predictive checks will be used to qualitatively assess the performance of the model. This involves generating a predictive distribution of the data based on the priors or posterior distribution of the parameters and comparing the results to the observed data.

This framework will be applied to all secondary outcomes where the score could be considered continuous. This includes: PTSD Checklist for DSM-5 total score, WHODAS 2.0 Overall score, SWEBO subscale scores, Intention to leave job, and EQ-5D-5L VAS health score.

5.6.3 Ordinal Outcomes

Many of our secondary outcomes are on an ordinal scale. In this section, we detail the modelling framework that will be used to analyses these measures. This includes questionnaires where total scores are calculated but the possible outcomes are limited enough that treating them as ordinal variables is still appropriate.

5.6.3.1 Cumulative Model

Some secondary outcome measure responses, for example the Intrusive Memory Rating (IMR) 7-point scale, are on a Likert-type scale that involves a series of statements that

respondents may choose from to rate their responses to evaluative questions. This is an example of an ordinal variable: the categories have an ordering, but the psychological distance between the categories may not be the same for all pairs (Bürkner & Vuorre, 2019). We can apply an ordered ordinal or cumulative regression model to the secondary data (and can implement this within the *brms* packages model family *cumulative* as this allows for ordered ordinal regression, Bürkner & Vuorre, 2019; Liddell & Kruschke, 2018). The model is specified as follows:

Let Y_{ij} denote a random variable with sample space $\{1,2,\dots,K\}$ for participant $i = 1,\dots,n$ and week $j = 4,12,24$, where the ordering is natural. A proportional odds model can be constructed based on the categorisation of a latent continuous variable Y_{ij} . The model assumes there are $K-1$ ordered cut-off thresholds τ_k for $k \in \{1,2,\dots,K-1\}$ which partition Y_{ij} into K categories, where $\tau_k < \tau_{k+1}$ for $k \in \{1,2,\dots,K-2\}$.

The thresholds τ_k for $k \in \{1,2,\dots,K-1\}$ and regression parameters corresponding to the treatment arm, week, and treatment arm and week interaction are our unknown model parameters.

$$Y_{ij}^{*} = \mu_{ij} + \varepsilon_{ij}$$

$$\mu_{ij} = Baseline_i\beta_1 + TAU_i\beta_2 + AC_i\beta_3 + t_{12}\beta_4 + t_{24}\beta_5 + TAU_i t_{12}\beta_6 + TAU_i t_{24}\beta_7 + AC_i t_{12}\beta_8 + AC_i t_{24}\beta_9$$

with the following priors:

$$\beta_i \sim N(0,10)$$

We assume that ε_{ij} is a random error from a logistic distribution with mean zero and constant variance, with cumulative distribution F (note that we can also use normal distribution as in

most cases the fit is quite similar). We can then derive the probability of Y_{ij} being in category k.

$$P(Y_{ij} \leq k) = F(\tau_k - \mu_{ij}) = \frac{e^{(\tau_k - \mu_{ij})}}{1 + e^{(\tau_k - \mu_{ij})}}$$

$$P(Y_{ij} = k) = P(Y_{ij} \leq k) - P(Y_{ij} \leq k - 1) = F(\tau_k - \mu_{ij}) - F(\tau_{k-1} - \mu_{ij})$$

$$\tau_k \sim N(\mu, \sigma^2)$$

All fixed effect terms will use normal priors centred on zero with a standard deviation of 10.

For the threshold parameters, we want our starting assumption to be a uniform distribution between the scores. Prior forms (starting with normal with different means, μ, and variance, σ^2) will be explored to achieve approximate uniform distribution with the priors sequentially centred in different places.

If issues with convergence or sampling arise, variations to these priors are permissible at the discretion of the statistician, however, all model variations and their justification will be reported.

This model will be applied to all outcomes with ordinal scores. This includes the SCI-02, ITL, EQ-5D-5L items 1-5, IMR items 1-6, GAD-2, and the PHQ-2.

5.6.4 Hypotheses

For different outcomes either a high or low score may be considered more favourable, and therefore a positive treatment effect may be indicated by either a higher or lower score relative to an alternative group. We will, therefore, set up alternative hypotheses for positive and negative treatment effects, depending on the outcome measure.

Secondary outcomes where a lower score is considered better:

- Number of intrusive memories at Week 12 and 24
- PTSD checklist
- GAD-2

- PHQ-2
- SWEBO
- SICAB
- EQ-5D-5L items 1-5
- IMR items 1-6
- WHODAS overall score

Secondary outcomes where a higher score is considered better:

- SCI-02
- ITL
- EQ-5D-5L: VAS score

The values of w_i in the hypotheses below describe contrasts. For example, at week 4, to test for differences between the intervention and TAU we will only look at β_2 and therefore $w_2 = 1$ and all other weights will be 0.

Lower score indicates better outcome:

$H_0: \sum_{i=1}^{9} w_i \beta_i \leq 0$	*The variable has no effect or a negative effect (equal or higher score for the outcome measure for reference compared to comparison)*
$H_1: \sum_{i=1}^{9} w_i \beta_i > 0$	*The variable has a positive effect (lower score for the outcome measure for reference compared to comparison)*

Higher score indicates better outcome:

$H_0: \sum_{i=1}^{9} w_i \beta_i \geq 0$	*The variable has no effect or a negative effect (equal or lower score for the outcome measure for reference compared to comparison)*
$H_1: \sum_{i=1}^{9} w_i \beta_i < 0$	*The variable has a positive effect (greater score for the outcome measure for reference compared to comparison)*

Taking the example of week 12, if we want to compare the effect of the active control vs. the intervention then w_3 and w_8 will be 1 and all other weights 0, i.e., :

$H_0: \beta_8 + \beta_3 = 0$	*The intervention has no effect at week 12 (participants in the intervention arm, compared to the active control arm, will have an equal outcome score at week 12)*
$H_1: \beta_8 + \beta_3 > 0$	*The intervention has a positive effect at week 12 (participants in the intervention arm, compared to the active control arm, will have lower outcome score at week 12)*

This is for a positive treatment effect of intervention vs active control where a lower score is considered favourable. For a negative effect the signs will be switched to $H_0: \beta_8 + \beta_3 \geq 0$ and $H_1: \beta_8 + \beta_3 < 0$. To compare intervention vs TAU, we will use $H_0: \beta_6 + \beta_2 = 0$ and $H_1: \beta_6 + \beta_2 > 0$. These values can be swapped around depending on what week and group we wish to compare to the intervention, and whether the outcome is favourable at a higher or lower score. The bayes factor represents the evidence of the hypothesis against its alternative.

5.7 Analyses of Exploratory Outcomes

Exploratory data analyses will be conducted on the exploratory outcomes and descriptive statistics will be presented overall and by arm.

Intervention Usage Data will be summarised with the key measures determined in discussion with clinical researchers e.g., the number of times the intervention was used, time of day, and duration.

Adverse events will be summarised, please see Section 5.12 ‘ Harms’, for further details.

The number of new traumatic events, and treatments received will be summarised overall and by arm.

Self-reported changes to job and hours worked will be summarised per question by arm and overall, with discrete variables summarised by frequencies and percentages, and continuous variables by mean and standard deviation (or median and IQR).

Qualitative interviews will be undertaken to assess the acceptability and feasibility of the intervention from participants, the analysis of this data is beyond the scope of this document.

5.7.1 Quantitative Feedback Questionnaire

The quantitative feedback questionnaire is performed at week 4 by only the intervention and active control arms. We want to identify if there is a difference between the two arms in the responses to the questionnaire. There is no total score for the questionnaire but items 1-5, and 8-10 are rated on 11-point scale from 0 (not at all) to 10 (very). Therefore, we can use a cumulative regression model like for the ordinal secondary outcomes (see Section 5.6.3.1). As there is only one week with data, we will not include time or baseline covariates, leaving only the AC covariate. The BF testing will therefore be based solely on the AC parameter.

$$Y_{ij}^{*} = \mu_{ij} + \varepsilon_{ij}$$

$$\mu_{ij} = AC_i\beta_1$$

5.8 Analyses of Additional Questionnaires

5.8.1 Credibility Expectancy Questionnaire

The credibility expectancy questionnaire is also ordinal in its responses, and we will again have to analyse individual questions as there is no total score. As with the quantitative feedback questionnaire it is only performed by the intervention and active control arms at a single timepoint. Therefore, we will set up the model identically to the quantitative feedback questionnaire.

5.9 Handling Missing Data

Exploratory data analysis will be used to assess the level of missing data, the occurrence of missingness will also be explored to identify any potential patterns. Patterns and percentage of missingness for the outcome and covariates will be reported overall and by week and arm.

When the dependent/outcome variable for a participant is missing across all visits then the incomplete cases do not contain any information about the model coefficients (Hughes et al., 2019; Jakobsen et al., 2017). In this situation, no additional information is gained from imputing the outcome therefore no specific methods will be used to handle the missing data, and instead complete cases will be analysed (Hughes et al., 2019; Jakobsen et al., 2017).

The number of intrusive memories will be computed at baseline, Week 4, Week 12, and Week 24 (see Table 3). For each of these outcomes, if they complete at least one day of the IMDD, the remaining days will be imputed (details provided in the following subsections). To assess the impact of this imputation on the analysis of the primary outcome, sensitivity analysis on the CCP will be conducted (see Section 5.10.4 ‘ Complete Case’).

Missing data will not be imputed for any other outcomes.

Additionally, sensitivity analysis will be conducted to examine the Missing Completely At Random assumption for the number of intrusive memories at baseline and week 4 (see Section 5.10.5 ‘ Imputation’).

5.9.1 Missing Data in IMDD at Interim Analysis

The primary outcome is the total number of intrusive memories of traumatic events recorded by participants in a brief daily online intrusive memory diary for 7 days at week 4. The total number of intrusive memories at baseline is also needed for the analyses as a co-variate.

Those with partial completion (at least one day of intrusive memory diary filled in) in the IAP set will have missing value(s) imputed. As the intrusive memory diary data is collected sequentially over time, we will use time series methods (Chatfield, 2003) and an expectation-maximisation algorithm (Dempster et al., 1977) to impute missing values (see also National Research Council., 2010). Initial missing values will be imputed by taking expectations across a participant's available diary data. Using Poisson likelihood and correlated errors, we maximise over this ‘full’ data set to provide updated expected values for the missing data. We

will iterate over these latter steps until convergence in values of missing data (to a pre-determined threshold) is achieved.

Those with no completion (no days of intrusive memory diary filled in) in the IAP set will not have missing value(s) imputed and will be excluded from the analysis.

The missing data approach described above is consistent with the interim analyses conducted in the previous related clinical trial (Ramineni et al., 2023).

5.9.2 Missing Data in IMDD at End-of-study Analysis

5.9.2.1 End-of-study Primary Analysis

The same missing data approach taken for interim analysis will be taken for the primary analysis (analysis of the primary outcome) at the end-of study.

5.9.2.2 End-of-study Secondary Analysis

The total number of intrusive memories of traumatic events recorded by participants in a brief daily online intrusive memory diary for 7 days during week 12, week 24 are secondary outcomes. The total number of intrusive memories at baseline is needed as a covariate, and additionally there is a fixed effect for the week (week 4, 12, 24).

Those with partial completion (at least one day of intrusive memory diary filled in) will have missing value(s) imputed using time series methods (Chatfield, 2003) and an expectation-maximisation algorithm (Dempster et al., 1977) as described above.

Those with no completion (no days of intrusive memory diary filled in) will be excluded from the analysis.

5.10 Sensitivity Analyses

Sensitivity analyses will be undertaken to determine how the results of the analysis may change depending upon modelling assumptions.

5.10.1 Prior distributions

For interim analyses of primary outcome and end-of-study analyses of both primary and secondary outcomes, we will examine how robust the original results are when the priors are altered, and the model is re-estimated. This sensitivity analysis will evaluate the posterior distributions of the model parameters when different priors are used. Example of alternative priors include:

- Priors with different variances i.e., large standard deviation which allows for very large effect size a priori, and small standard deviation which implies the assumption that we expect the effect not to be very large a priori.
- Prior of different forms (e.g., student t distribution)

Visual inspection of posterior distributions and posterior predictive checks will be used to assess the impact of varying priors at interim analysis and end of study analysis.

5.10.2 Outliers

For interim analyses and end-of-study analyses of primary outcome, potential outliers will be identified through inspection of residual plots and Cook's distance vs leverage plots of an appropriate fitted frequentist model. If any outlier(s) are identified, results using data without outliers will be presented as sensitivity analysis.

5.10.3 Per Protocol

At the end of the study, Per-Protocol Population (PPP) will be used to analyse the primary outcome as a sensitivity analysis.

5.10.4 Complete Case

In the interim and end-of-study analyses, where imputation methods have been implemented for primary outcome analyses (i.e., the time-series imputation for baseline and/or Week 4 IMDD data, described in Section 5.9 'Handling Missing Data'), analyses will also be conducted on Complete-Case Population (CCP, see section 3.3.4 'Complete-Case Population (CCP)')

as a sensitivity analysis to examine the effect of this imputation method. This will involve running the analysis for the primary outcome on only those with complete weekly diaries at baseline and Week 4, excluding those with any number of missing days at either visit.

5.10.5 Imputation

Excluding missing cases assumes that the data is missing completely at random (MCAR) (i.e., the fact that the data are missing is independent of the observed and unobserved data) Therefore, during interim analyses and end-of-study analyses, if the number of intrusive memories at baseline or week 4 is missing for any participant, analysis using a standard imputation approach, e.g. mean value substitution, will be conducted as a sensitivity analyses.

5.10.6 Simulations

For end-of-study analyses of the primary outcome, we will re-perform analysis for different artificial data sets simulated i.e. bootstrapping, or from the posterior predictive distribution (Schad et al., 2022). For each of these data sets, we can proceed in exactly the same way as we did for the real observed experimental data i.e., again fit the same model, now to the simulated data, and using the same prior as before. For each simulated dataset, we compute the Bayes factor. We can then visualise the distribution of Bayes factors. This provides a robustness check on the validity of the differences reported between the groups.

5.10.7 Frequentist Analysis

For the primary outcome, frequentist statistical testing will be performed as a supplementary sensitivity analysis. Incidence rate ratios (IRR) will be presented with 95% confidence intervals to quantify the relative treatment effect of the intervention. Single-level generalised linear modelling will be performed for the primary analysis, using Arm and Baseline intrusive memories as fixed effects. The model design will be the same as described in Section 5.5.1 ‘Modelling Assumptions’, a Poisson linear mixed model with baseline intrusive memories and treatment arm as fixed effects, and subject as a random effect. The same imputed data will be used for this modelling as in the Bayesian analysis of the primary outcome.

The primary outcome (intrusive memory count) will be modelled using a Poisson linear mixed model (Observation Level Random Effects Poisson, Harrison, 2014). The baseline number of intrusive memories, and treatment arm are fitted as fixed effects with a random intercept effect for participant.

Let n be the number of participants which we are analysing, then for each participant $i = 1, \dots, n$ we denote:

- Y_i to be the random variable representing the primary outcome (total number of intrusive memories in week 4).
- $Baseline_i$ to be the total number of intrusive memories recorded during the run-in week.
- The three arms (Intervention, TAU, Active Control) are indicated by dichotomous variables I_i, TAU_i, AC_i which equal one if participant i was randomised to the respective arm and zero otherwise.
- $subj_i$ denotes a dichotomous variable, equal to one if the observation is from the subject i, and zero if not.

γ_{0i} is the random effect for the intercept for each participant i that accounts for the participant-specific variation in the primary outcome. The random-effects intercepts γ_{0i} are drawn from a normal distribution with mean 0 and standard deviation σ_{γ_0} which is estimated by the model.The model is as follows, where $\alpha, \beta_1, \beta_2, \beta_3, \sigma_{\gamma_0}$ are our unknown model parameters.

$$Y_i \sim Poisson(\mu_i)$$

$$E(Y_i) = \mu_i$$

$$ln(\mu_i) = \alpha + Baseline_i\beta_1 + TAU_i\beta_2 + AC_i\beta_3 + \sum_{i=1}^{n} subj_i\gamma_{0i}$$

$$\gamma_{0i} \sim N(0, \sigma_{\gamma_0}^2)$$

Alternative models will be considered as appropriate, such as those assuming observations to be drawn from negative binomial or zero-inflated distributions where the response data are over dispersed or zero-inflated. Incidence Rate Ratios (IRRs), and their 95% confidence intervals, will be reported for (i) the intervention group vs. active control group comparison and (ii), the intervention group vs. TAU group comparison. The IRR represents the relative rate of occurrence for the event compared to the reference level. An IRR greater than 1 signifies an increased rate of intrusive memories in the intervention group, while an IRR less than 1 would indicate a reduced rate.

5.11 Statistical Software

The statistical analysis described will be conducted in R (R Core Team, 2021) using recent software versions available at the time. We will use R Markdown Scripts to enhance reproducibility. Additionally, SAS 9.4 or later may be used to present tables and figures.

R package ***tidyverse*** will be used for data manipulation and plotting with ***ggplot2***. Modelling in a Bayesian framework uses the R package ***brms*** (Bürkner, 2017, 2018). Model fit can also be evaluated through built-in posterior checks in ***brms***. The packages ***bayestestR*** will be used to calculate Bayes factors that compare hypotheses relating to the primary objective***.*** Exploratory modelling and model comparisons are conducted using the following R packages ***lme4, MASS, pscl, boot, performance, ZIM, lmtest.*** Bespoke R scripts using time series models implemented in an E-M approach (Dempster et al., 1977) are used to deal with imputation of missing values.

5.12 Harms

During the interim analyses, we test for negative effect allowing for early action to be taken if needed (see Section 5.5.3.2 ’Negative Effect’).

The number (percentage) of patients experiencing each adverse event / serious adverse event will be presented for each treatment arm categorised by severity (across follow-up time). For

each participant, only the maximum severity experienced of each type of adverse event will be displayed. The number (percentage) of occurrences of each adverse / serious adverse event will also be presented for each treatment arm. No formal statistical testing will be undertaken.

6 QUALITY ASSURANCE AND CHECKS

6.1 Scripts and Data

Analysis code and output will be reviewed by Prof. Michael Bonsall (Statistical Lead) and/or checked by another statistician.

The statistician is responsible for storing analytical code and outputs and making the final analytical code and results for interim and end-of-study analysis available (e.g., for internal records, and/or to upload to Open Science Framework) as required by the study team.

All datasets for analysis are created using fully validated data processing scripts and provided to the statistician by the data management team at P1vital Products via the Citrix ShareFile System. Additional dataset formatting undertaken by the statistician to conduct statistical analysis will be quality checked by the data management team at P1vital Products. All quality checks conducted by the data management team will be documented and all documentation will be stored on the PPL secure file server.

REFERENCES

Bell, M. L., Fiero, M., Horton, N. J., & Hsu, C. H. (2014). Handling missing data in RCTs; A review of the top medical journals. In *BMC Medical Research Methodology* 14(1). https://doi.org/10.1186/1471-2288-14-118

Bryant, R. A., Creamer, M., O'Donnell, M., Forbes, D., McFarlane, A. C., Silove, D., & Hadzi-Pavlovic, D. (2017). Acute and chronic posttraumatic stress symptoms in the emergence of posttraumatic stress disorder a network analysis. *JAMA Psychiatry*, 74(2). https://doi.org/10.1001/jamapsychiatry.2016.3470

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), pp.1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P. C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, 10(1), 395–411. https://doi.org/10.32614/rj-2018-017

Bürkner, P. C. (2021). Bayesian Item Response Modeling in R with brms and Stan. *Journal of Statistical Software*, 100(5). https://doi.org/10.18637/JSS.V100.I05

Bürkner, P. C., & Vuorre, M. (2019). Ordinal Regression Models in Psychology: A Tutorial. *Advances in Methods and Practices in Psychological Science*, 2(1). https://doi.org/10.1177/2515245918823199

Chatfield, C. and A. J. Collins. , & B. A. F. (2003). The analysis of time series: an introduction. In *Translated by: Noroumand, H., A., & Bozorgnia, A. Ferdousi University of Mashhad publications* (Sixth Edition). Chapman and hall/CRC. https://doi.org/10.4324/9780203491683

Deforges, C., Sandoz, V., Noël, Y., Avignon, V., Desseauve, D., Bourdin, J., Vial, Y., Ayers, S., Holmes, E.A., Epiney, M. and Horsch, A. (2023). Single-session visuospatial task procedure to prevent childbirth-related posttraumatic stress disorder: a multicentre double-blind randomised controlled trial. *Molecular Psychiatry*, 28(9), pp.3842-3850. https://doi.org/10.1038/s41380-023-02275-w

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), pp.1–22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

Dimairo, M., Pallmann, P., Wason, J., Todd, S., Jaki, T., Julious, S. A., Mander, A. P., Weir, C. J., Koenig, F., Walton, M. K., Nicholl, J. P., Coates, E., Biggs, K., Hamasaki, T., Proschan, M. A., Scott, J. A., Ando, Y., Hind, D., & Altman, D. G. (2020). The Adaptive designs CONSORT Extension (ACE) statement: A checklist with explanation and elaboration guideline for reporting randomised trials that use an adaptive design. *BMJ*, 369(m115). https://doi.org/10.1136/bmj.m115

Ewings, S., Saunders, G., Jaki, T., & Mozgunov, P. (2022). Practical recommendations for implementing a Bayesian adaptive phase I design during a pandemic. *BMC Medical Research Methodology*, 22(1). https://doi.org/10.1186/s12874-022-01512-0

European Medicines Agency. (2016). Guideline on multiplicity issues in clinical trials. *Guidance*, 44, pp. 1–15.

Freeman, T. P., Hindocha, C., Baio, G., Shaban, N. D. C., Thomas, E. M., Astbury, D., Freeman, A. M., Lees, R., Craft, S., Morrison, P. D., Bloomfield, M. A. P., O'Ryan, D., Kinghorn, J., Morgan, C. J. A., Mofeez, A., & Curran, H. V. (2020). Cannabidiol for the treatment of cannabis use disorder: a phase 2a, double-blind, placebo-controlled, randomised, adaptive Bayesian trial. *The Lancet Psychiatry*, *7*(10), pp. 865–874. https://doi.org/10.1016/S2215-0366(20)30290-X

Gamble, C., Krishan, A., Stocken, D., Lewis, S., Juszczak, E., Doré, C., Williamson, P. R., Altman, D. G., Montgomery, A., Lim, P., Berlin, J., Senn, S., Day, S., Barbachano, Y., & Loder, E. (2017). Guidelines for the content of statistical analysis plans in clinical trials. *JAMA - Journal of the American Medical Association*, 318(23). https://doi.org/10.1001/jama.2017.18556

Giovagnoli, A. (2021). The Bayesian Design of Adaptive Clinical Trials. *International Journal of Environmental Research and Public Health*, 18(2). https://doi.org/10.3390/ijerph18020530

Harrison, X. A. (2014). Using observation-level random effects to model overdispersion in count data in ecology and evolution. *PeerJ Inc*, 2(e616). https://doi.org/10.7717/peerj.616

Heath, A., Yaskina, M., Pechlivanoglou, P., Rios, D., Offringa, M., Klassen, T. P., Poonai, N., & Pullenayegum, E. (2021). A Bayesian response-adaptive dose-finding and comparative effectiveness trial. *Clinical Trials*, 18(1). https://doi.org/10.1177/1740774520965173

Hollis, S., & Campbell, F. (1999). What is meant by intention to treat analysis? Survey of published randomised controlled trials. *British Medical Journal*, 319(7211). https://doi.org/10.1136/bmj.319.7211.670

Holmes, E. A., O'Connor, R. C., Perry, V. H., Tracey, I., Wessely, S., Arseneault, L., Ballard, C., Christensen, H., Cohen Silver, R., Everall, I., Ford, T., John, A., Kabir, T., King, K., Madan, I., Michie, S., Przybylski, A. K., Shafran, R., Sweeney, A., … Bullmore, E. (2020). Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. *The Lancet Psychiatry*, *7*(6), pp. 547–560. https://doi.org/10.1016/S2215-0366(20)30168-1

Horsch, A., Vial, Y., Favrod, C., Harari, M. M., Blackwell, S. E., Watson, P., Iyadurai, L., Bonsall, M. B., & Holmes, E. A. (2017). Reducing intrusive traumatic memories after emergency caesarean section: A proof-of-principle randomized controlled study. *Behaviour Research and Therapy*, 94, pp. 36–47. https://doi.org/10.1016/j.brat.2017.03.018

Hughes, R. A., Heron, J., Sterne, J. A. C., & Tilling, K. (2019). Accounting for missing data in statistical analyses: Multiple imputation is not always the answer. *International Journal of Epidemiology*, 48(4). https://doi.org/10.1093/ije/dyz032

Iyadurai, L., Blackwell, S. E., Meiser-Stedman, R., Watson, P. C., Bonsall, M. B., Geddes, J. R., Nobre, A. C., & Holmes, E. A. (2018). Preventing intrusive memories after trauma via a brief

intervention involving Tetris computer game play in the emergency department: A proof-of-concept randomized controlled trial. *Molecular Psychiatry*, 23(3), pp. 674–682. https://doi.org/10.1038/mp.2017.23

Iyadurai, L., Highfield, J., Kanstrup, M., Markham, A., Ramineni, V., Guo, B., Jaki, T., Kingslake, J., Goodwin, G.M., Summers, C. and Bonsall, M.B., 2023. Reducing intrusive memories after trauma via an imagery-competing task intervention in COVID-19 intensive care staff: a randomised controlled trial. *Translational psychiatry*, 13(1), pp. 290. https://doi.org/10.1038/s41398-023-02578-0

Iyadurai, L., Visser, R. M., Lau-Zhu, A., Porcheret, K., Horsch, A., Holmes, E. A., & James, E. L. (2019). Intrusive memories of trauma: A target for research bridging cognitive science and its clinical application. In *Clinical Psychology Review* , 69. https://doi.org/10.1016/j.cpr.2018.08.005

Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials - A practical guide with flowcharts. *BMC Medical Research Methodology*, 17(1). https://doi.org/10.1186/s12874-017-0442-1

James, E. L., Bonsall, M. B., Hoppitt, L., Tunbridge, E. M., Geddes, J. R., Milton, A. L., & Holmes, E. A. (2015). Computer Game Play Reduces Intrusive Memories of Experimental Trauma via Reconsolidation-Update Mechanisms. *Psychological Science*, 26(8), pp. 1201–1215. https://doi.org/10.1177/0956797615583071

Jeffreys, Harold. (1998). *The theory of probability*. OUP Oxford.

Kanstrup, M., Kontio, E., Geranmayeh, A., Olofsdotter Lauri, K., Moulds, M. L., & Holmes, E. A. (2021). A single case series using visuospatial task interference to reduce the number of visual intrusive memories of trauma with refugees. *Clinical Psychology and Psychotherapy*, 28(1), pp. 109–123. https://doi.org/10.1002/cpp.2489

Kanstrup, M., Singh, L., Göransson, K. E., Widoff, J., Taylor, R. S., Gamble, B., Iyadurai, L., Moulds, M. L., & Holmes, E. A. (2021). Reducing intrusive memories after trauma via a brief cognitive task intervention in the hospital emergency department: an exploratory pilot randomised controlled trial. *Translational Psychiatry*, 11(1). https://doi.org/10.1038/s41398-020-01124-6

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90(430), pp. 773–795. https://doi.org/10.1080/01621459.1995.10476572

Kessler, H., Holmes, E. A., Blackwell, S. E., Schmidt, A. C., Schweer, J. M., Bücker, A., Herpertz, S., Axmacher, N., & Kehyayan, A. (2018). Reducing intrusive memories of trauma using a visuospatial interference intervention with inpatients with posttraumatic stress disorder (PTSD). *Journal of Consulting and Clinical Psychology*, 86(12), pp. 1076–1090. https://doi.org/10.1037/ccp0000340

Kleim, B., Bingisser, M. B., Westphal, M., & Bingisser, R. (2015). Frozen moments: Flashback memories of critical incidents in emergency personnel. *Brain and Behavior*, 5(7), e00325. https://doi.org/10.1002/brb3.325

Lee, M. D., & Wagenmakers, E. J. (2013). Bayesian cognitive modeling: A practical course. In *Bayesian Cognitive Modeling: A Practical Course*. https://doi.org/10.1017/CBO9781139087759

Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, 79. https://doi.org/10.1016/j.jesp.2018.08.009

McNally, R. J. (2017). Networks and nosology in posttraumatic stress disorder. In *JAMA Psychiatry* 74(2). https://doi.org/10.1001/jamapsychiatry.2016.3344

National Research Council. (2010). *The prevention and treatment of missing data in clinical trials.*

R Core Team. (2021). R: A language and environment for statistical computing. In *R Foundation for Statistical Computing*. R Foundation for Statistical Computing. http://www. R-project. org

Ramineni, V., Millroth, P., Iyadurai, L., Jaki, T., Kingslake, J., Highfield, J., Summers, C., Bonsall, M. B., & Holmes, E. A. (2023). Treating intrusive memories after trauma in healthcare workers: a Bayesian adaptive randomised trial developing an imagery-competing task intervention. *Molecular Psychiatry*. https://doi.org/10.1101/2022.10.06.22280777

Ryan, E. G., Lamb, S. E., Williamson, E., & Gates, S. (2020). Bayesian adaptive designs for multi-arm trials: An orthopaedic case study. *Trials*, 21(1). https://doi.org/10.1186/s13063-019-4021-0

Ryan, E. G., Stallard, N., Lall, R., Ji, C., Perkins, G. D., & Gates, S. (2020). Bayesian group sequential designs for phase III emergency medicine trials: A case study using the PARAMEDIC2 trial. *Trials*, 21(1). https://doi.org/10.1186/s13063-019-4024-x

Schad, D. J., Nicenboim, B., Bürkner, P. C., Betancourt, M., & Vasishth, S. (2022). Workflow Techniques for the Robust Use of Bayes Factors. *Psychological Methods*. https://doi.org/10.1037/met0000472

Schönbrodt, F. D., & Stefan, A. M. (2019). *BFDA: An R package for Bayes factor design analysis* (version 0.5.0). https://github.com/nicebread/BFDA

Schönbrodt, F. D., & Wagenmakers, E. J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin and Review*, 25(1), pp. 128–142. https://doi.org/10.3758/s13423-017-1230-y

Schönbrodt, F. D., Wagenmakers, E. J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, 22(2), pp. 322–339. https://doi.org/10.1037/met0000061

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3). https://doi.org/10.18637/jss.v045.i03

Wang, B., Wang, H., Tu, X. M., & Feng, C. (2017). Comparisons of Superiority, Non-inferiority, and Equivalence Trials. *Shanghai Archives of Psychiatry*, 29(6). https://doi.org/10.11919/j.issn.1002-0829.217163

Zhou, X., & Reiter, J. P. (2010). A note on Bayesian inference after multiple imputation. *American Statistician*, 64(2). https://doi.org/10.1198/tast.2010.09109