Sleep Treatment Outcome Predictors (STOP) Pilot Study

Study protocol and statistical analysis plan

Trial registration: NCT03062891

Ethical approval: 9/26/2016

Study protocol: 5/9/2017

Statistical analysis plan: 3/5/2018

Study protocol and statistical analysis plan combined together into single document on 27/4/2020

Full study protocol including planned primary, secondary, and exploratory aims:

Denis et al (2017). Sleep Treatment Outcome Predictors (STOP) pilot study: A protocol for a randomised controlled trial examining predictors of change of insomnia symptoms and associated traits following cognitive-behavioural therapy for insomnia in an unselected sample. *BMJ Open, 7.* E017177. DOI: 10.1136/bmjopen-2017-017177

List of measure names and abbreviations

SCI = Sleep condition indicator

PSQI = Pittsburgh sleep quality index

PSQI-A = Pittsburgh sleep quality index addendum

PSAS = Pre-sleep arousal scale

DBAS = Dysfunctional beliefs about sleep questionnaire

MCTQ = Munich chronotype questionnaire

STAI = State-trait anxiety index

MFQ = Mood and feelings questionnaire

ADHD = Attention deficit hyperactivity disorder

SPEQ = Specific psychotic experiences questionnaire

PMH = Positive mental health scale

PSS = Perceived stress scale

LTE = List of threatening events

TAQ = Treatment acceptability questionnaire

Study protocol

Methods and analysis

Study dates

Recruitment for the study started November 2016, and data will finish being collected by the end of September 2017. The study was retrospectively registered on 5th December 2016. The reason the trial was registered retrospectively was due to very restricted limitations on when participants could be recruited (see Participant recruitment). Unfortunately, the trial was not registered until after the first recruitment dates had passed. Rather than lose potential recruiting opportunities, we decided to register the trial retrospectively.

Design

The study is a two-group parallel randomised controlled trial in which the intervention group will receive a digital CBT-I intervention, and the control group will receive a weekly online puzzle. See the intervention section for more details.

Participants are female students (both undergraduate and postgraduate) completing a psychology programme at one of three London universities (for full details see the trial registration). After completing the baseline assessment online via the Qualtrics system, participants were randomly allocated to either the CBT-I or puzzles group. Three weeks later, participants completed a second online assessment, and then a third online assessment 6 weeks after the start of the study. Finally, a follow-up online assessment will be carried out 6 months after group allocation. Participants were also invited to give a DNA sample at the start of data collection. While we have limited statistical power to look at genetic predictors of treatment outcomes in the pilot

study, these samples could be pooled with other data collected in the future, and also provide a useful opportunity for our collaborators to collect data for another ongoing research initiative. See **Figure 1** for a detailed outline of the study timeline.

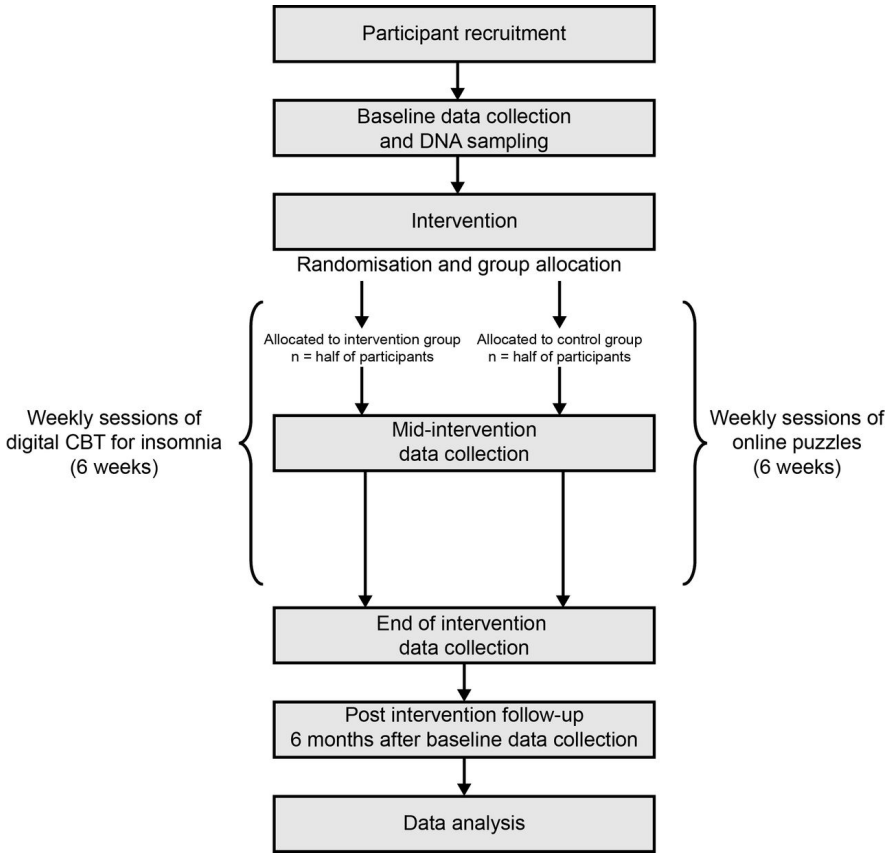

Figure 1. Study flowchart

Inclusion and exclusion criteria

Only females were eligible for participation. This is because the majority of the students on the psychology courses are female, and so adding males would create heterogeneity but without sufficient power to examine this further. Furthermore, only individuals enrolled in a psychology course from one of 3 London universities were recruited due to reasons of convenience. We focused our recruitment efforts on first year students in particular, as it is possible that a small number of students in other years may have already taken part in studies using the same digital CBT-I platform. In order to address this point explicitly, in the questionnaires given to participants, they were asked if they have had any experience with Sleepio before taking part in this study.

Participant recruitment

Participants were recruited to the study using a two-step procedure. Initially, potential participants were contacted via an e-mail that provided the study information, specific instructions as to the nature of the recruitment day, and contact information.

The second stage of recruitment involved a series of recruitment days at the three 3 sites. These recruitment days were timed to coincide with classes that potential participants were present at, to make it more likely that they would be in university. At sign-up, participants were given a paper copy of the information sheet and were given the option to ask any questions about the nature of the study. After confirming that they were happy to take part in the study, all participants were asked to give informed consent, provide a DNA sample (see DNA sample collection), and were assigned a unique participant ID number which was used for future assessments. To allow the participation of individuals who wished to take part in the study but were unable to sign-up in person, participants were given the option to contact the research team directly by e-mail in order to arrange providing consent to take part in the study.

Participants will be rewarded for their time, either in the form of course credits (offered credits + £5 online gift voucher) or online shopping voucher (£40), awarded to them upon completion of the study.

Randomisation and study automation

After collection of baseline data, participants were randomly allocated to either the CBT-I group or the puzzles group. A member of the research team randomised eligible participants using the blockrand package for R. Participants were stratified based upon age, sleep problems, and study site. Stratification on age was performed to assure similar age distributions in both groups. Stratification on sleep problems was implemented to avoid the possibility of a disproportionate number of participants with sleep problems being randomly allocated to the same group. Stratification for study site was implemented to avoid an unnecessary delay between completing the first questionnaire and being allocated to a group.

An automated e-mail was sent to participants to inform them of which group they had been assigned. Those in the CBT-I group were given further information as to the nature of the programme (see Digital CBT-I) as well as a unique code needed to log into the website. Those in the puzzles group were given information as to the nature of the tasks that they were required to complete (see Puzzles). Participants were not able to change groups once they have been allocated.

Intervention

Digital CBT-I

CBT-I participants received 6 weekly CBT-I sessions delivered by an animated 'virtual therapist' (The Prof) via the online platform 'Sleepio' (http://www.sleepio.com). The programme comprised a fully automated media-rich web application, driven dynamically by baseline, adherence, performance and progress data, and provides additional access to elements such as an online library with background information, a community of fellow users, and support, prompts and reminders sent by e-mail.

The Sleepio programme covers behavioural (e.g., sleep restriction, stimulus control) and cognitive (e.g., putting the day to rest, thought restructuring, imagery, articulatory suppression, paradoxical intention, mindfulness) strategies, as well as additional relaxation strategies (progressive muscle relaxation and autogenic training) and advice on lifestyle and bedroom factors (sleep hygiene). As part of the intervention, participants filled in a daily sleep diary. The intervention was based upon a previously validated manual. Sleepio has been shown to improve sleep and associated daytime functioning in adults diagnosed with insomnia disorder.

Puzzles

Participants in the control group were sent weekly puzzles to complete within Qualtrics. Each puzzle was designed to be cognitively engaging, and time taken to complete a puzzle has been matched as closely as possible to the time taken to complete one session of digital CBT-I. Puzzles were sent directly to participants via automated distribution e-mails sent at 7 day intervals. In order to track whether participants were completing the puzzles, they were required

to enter their participant ID number at the start of each puzzle. The types of puzzles administered to participants included word searches, crosswords, and lateral thinking problems.

Data Collection

DNA sample collection

This project was conducted in collaboration with the National Institute for Health Research (NIHR) Biomedical Research Centre (BRC) BioResource for Mental and Neurological Health in London as part of a national NIHR initiative to build up a central library of information (or "BioBank)" about people's health.

In this study, we obtained saliva samples from our participants after obtaining consent during the recruitment days. Samples were collected by a researcher from the BioResource team in compliance with their ethically approved protocol. The BRC is the custodian of the samples received. On receipt, samples were logged and prepared for extraction of DNA. The BRC ensured that genetic samples were processed in accordance with strict health and safety guidelines and under the requirements of the Human Tissue Act (HTA). King's College London holds a HTA license, number: 12293. All samples are stored in tubes labelled with a barcode that includes the participant number. The link between the participant ID and de-identified data is kept in a secure folder. The DNA samples collected as part of this study are stored by the BRC for future analysis and hypothesis testing with appropriate ethical approval in the future, and under existing BRC BioResource approvals.

Wave 1 data collection

Eligible participants were given the option of completing the baseline survey online after signing up. Participants were encouraged to complete the survey within one week from sign-up. Paper copies were made available for participants who had problems with their device or internet access.

Participants completed all measures, as shown in **Table 1**. Participants had the option to leave out any question. The survey took 30-40 minutes to complete. At the end the survey, participants were reminded that they would be contacted with regards to future data collection.

Waves 2-4 data collection

The second and third, waves of data collection were carried out 3 weeks and 6 weeks following allocation. The fourth wave will be carried out 6 months following the allocation of participants to groups. These time points correspond to mid-intervention, end-of-intervention, and post-intervention follow-up time points respectively (see Figure 1 for more detail). Automated e-mails distributed by Qualtrics will be sent to participants at the designated intervals. Not all measures are assessed at all waves, as shown in **Table 1**. Follow-up emails to non-responders will be sent each week to participants who fall behind on their tasks (i.e. CBT-I, puzzles, or surveys).

Measures

Descriptions for all measures used are provided below. For waves 2-4 some measures were adapted to ask participants to consider their answers with reference to the last 2 weeks (unless otherwise stated below), in order to ensure participants were considering only the time since the last wave of data collection when responding. Full details on each measure used can be found in **Supplementary File 3**.

Demographic information was collected at baseline. At the start of each survey, participants were asked to indicate whether it was currently term time, exam time, or holiday time. At wave 2, participants in the Sleepio group indicated whether they had ever used Sleepio before.

Sleep measures

Insomnia symptoms – *Sleep Condition Indicator* (SCI). An 8-item measure assessing symptoms of insomnia, used to identify insomnia symptoms in community samples.

Sleep Quality – *Pittsburgh Sleep Quality Index (PSQI)*. An 18-item questionnaire assessing 7 components of sleep quality and disturbances, which also yields a global score of sleep quality.

The scale has been shown to be reliable and valid in assessing sleep quality in adult community samples.

Trauma-related sleep disturbances – *Pittsburgh Sleep Quality Index Addendum*. Assesses frequency of 7 sleep disturbances typically related to trauma. The measure has been validated for use in assessing these disturbances.

Pre-sleep-arousal – *Pre-sleep Arousal Scale* Measures symptoms of cognitive (8 items) and somatic (8 items) arousal experienced around bedtime. It is has been validated with respect to objective measures of pre-sleep arousal.

Cognitions about sleep – *Dysfunctional Beliefs About Sleep Scale (DBAS)*. A 10-item questionnaire that includes items about sleep-disruptive cognitions such as faulty beliefs, worry, and attentional bias. The measure has shown to be reliable.

Chronotype – *Munich Chronotype Questionnaire (MCTQ)*. Chronotype is estimated as the midpoint of sleep on workdays and work-free days minus half of the difference between sleep duration on work-free days and average sleep duration of the work to control for sleep debt (i.e. the midpoint of sleep on work-free days, corrected for sleep duration). The MCTQ is a reliable and valid measure of chronotype.

Sleep paralysis – *Waterloo Unusual Experiences Questionnaire (WUSEQ*). Items from the WUSEQ were used to assess the frequency of sleep paralysis and associated hallucinations. The measure is valid and reliable in healthy student samples.

Sleep paralysis – *Fearful Isolated Sleep Paralysis Interview (FISPI)*. Two items from this measure were adapted to measure the amount of fear/distress typically caused by sleep paralysis episodes, and how much interference with waking life episodes have caused. The FISPI has been used as a valid and reliable measure of sleep paralysis in university samples.

Exploding head syndrome – *Munich Parasomnia Screening (MUPS)*. Lifetime prevalence of exploding head syndrome was measure using a single item from the MUPS.

Psychopathology and well-being measures

Anxiety symptoms – *State Trait Anxiety Index (STAI)* The STAI assesses both state (20 items) and trait (20 items) levels of anxiety, and is a valid and reliable measure of anxiety symptoms.

Depressed mood – *Mood and Feelings Questionnaire (MFQ)* Depressed mood was measured using the 13-item MFQ. This has been shown to be a valid measure of depressed mood.

ADHD symptomatology – Bespoke measure examined 18 symptoms of ADHD according to DSM-5 criteria. This is a valid and reliable measure of ADHD symptoms, and has been previously used to in young adults to assess ADHD symptomatology in the context of sleep quality.

Psychotic experiences – *Specific Psychotic Experiences Questionnaire (SPEQ)*. Sub-scales relating to paranoia, hallucinations, and cognitive disorganisation were used as they are strongly related with sleep disturbances. The scale has been shown to have good reliability and validity.

Positive mental health – *Positive Mental Health Scale (PMH)*. Positive aspects of health and life experiences were assessed using a 9-item questionnaire.

Life stress – *Perceived Stress Scale (PSS).* Life stress was measured with a 10-item measure. A review of articles assessing the psychometric properties of the PSS found the measure to be a reliable and valid measure of life stress.

Exposure to threatening events – *List of Threatening Experiences (LTE).* Participants were asked to indicate whether they had experienced any threatening events from a list of 24. The LTE has been shown to have high reliability and be a valid measure of exposure to potentially threatening experiences.

Lifestyle measures

At each wave, participants were asked about their sleeping arrangements and alcohol and caffeine intake. Cigarette and electronic (e) cigarette usage were assessed at baseline.

Treatment acceptability

The 6-item *Treatment Acceptability Questionnaire (TAQ)* asked specific questions regarding the degree to which they found the treatment acceptable, ethical, and effective. There were also specific questions about the nature of the virtual therapist. Only participants in the Sleepio group received the TAQ.

Sample size

For this study, the target was to have 200 participants, which should provide power to examining our primary research questions. Though we plan to over recruit to account for some attrition throughout the study. As such, 240 participants will be recruited. Power analyses are often conducted using hypothesised effect sizes based on mean differences (e.g. before and after treatment). However, as this is a pilot for a future behavioural genetics study, the main statistic of interest is not mean differences, but individual differences (i.e. variances). The decision to recruit 200 participants for this pilot study was mainly based on personal experiences of recruiting undergraduates from our institutions.

Table 1 - Study timeline and measures

Measure	Wave 1 Baseline	Wave 2 3 weeks	Wave 3 6 weeks	Wave 4 6-month
SCI	X	X	X	X
PSQI	X	X	X	X
PSQI-A	X		X	X
PSAS	X	X	X	X
DBAS	X	X	X	X
MCTQ	X		X	X
STAI	X	X	X	X
MFQ	X	X	X	X
ADHD	X		X	X
SPEQ	X		X	X
PMH	X	X	X	X

PSS	X	X	X	X
LTE	X	X	X	X
Lifestyle	X	X	X	X
TAQ		X	X	

Primary outcome
Secondary outcome
Exploratory outcome

Statistical analysis

Missing data

Given participant drop-out through the trial a two-step process which assumed data were missing at random was used. A binary variable was created to indicate whether data were missing or not. Predictors of missing data at the end of intervention assessment were then examined using logistic regression. All baseline measures and treatment acceptability at mid-intervention were investigated as potential predictors in the same regression model. In the second step, multiple imputation was used to estimate missing data, carried out in STATA using an imputation- chained-equations algorithm. Significant predictors of missingness were included in the imputation model. A total of 25 imputed datasets were created. All variables that had missing data of <30% and were deemed missing completely at random or missing at random were entered into the multiple imputation algorithm.

Given high drop-out at the fourth assessment, all primary/secondary analyses will be focused on the Wave 3 assessment.

Descriptive statistics

Discrete variables: N & %

Continuous variables: Mean and SD

Descriptive statistics for all measures at each wave where the variable is assessed (see study timeline and measures section).

Summary statistics produced both for the overall sample and by group (Sleepio, hereafter referred to intervention; and Puzzles, hereafter referred to as control).

Primary aim 1 – Insomnia symptom improvement

SCI change scores between baseline and end of intervention will be computed, and Cohen's d will be calculated as a measure of effect size for the *between-group* difference in SCI change score. Next, given the effect size obtained, the following will be calculated:

1. The percentage of participants in the intervention group who have an SCI change score above the mean SCI change score of the control group (Cohen's U_3)
2. The percentage overlap between the SCI change scores of the two groups
3. The probability that one person picked at random from the intervention group will have a higher score than a person picked at random from the control group (the probability of superiority)

Independent-samples t-tests will be used to assess significant differences in change score size between intervention and control. Changes in the percentage of participants meeting diagnostic criteria for insomnia between baseline and the end of the intervention, and the percentage of participants with change scores exceeding the SCI reliable change index will be assessed using chi-square.

The change in insomnia symptoms across the intervention will be modelled using a generalized estimating equation:

Type	Variable	Categorical/continuous
Dependent	SCI score at mid- and end-of-intervention assessment	Continuous
Predictors	**Group**: Control (1), Intervention (2)	Categorical Categorical

	Time: Dummy variable coded as 2 (mid-assessment),3 (end-assessment),4 (6-month follow-up) **Group X Time interaction**	Categorical
Covariates	**Baseline SCI score** **Age**	Continuous Continuous

Model – Generalized estimating equation. Performed using the `xtgee` command in STATA.

Model to examine whether change in SCI score across follow-up assessments is predicted by group after the effects of baseline SCI score, age, and time of year have been controlled for

Primary aim 2 – Participation rates and adherence

The number and proportion of participants who completed each wave will be calculated. For the proportions, 95% confidence intervals will be calculated. Calculation of confidence intervals for proportions will be achieved using the `proportion` command in STATA.

Group differences in participation rate will be assessed for statistical significance using chi-square tests to assess differences between groups at each assessment.

Within each group, adherence will be assessed by deriving the number and proportion of participants who completed either all six weekly sessions of Sleepio (intervention group), or all six weekly sessions of puzzles (control group). For proportions, 95% confidence intervals will be calculated using the `proportion` command in Stata. Group differences will be assessed using chi-square.

Primary aim 3 – Treatment acceptability

Treatment acceptability is assessed in the intervention group only, using the TAQ. The TAQ contains 6 items, each asking about a different aspect of treatment acceptability on a 7-point scale. A mean TAQ score will be calculated by summing responses to each of the 6 items. The scale has a theoretical range of 7-42, with a higher score indicating higher treatment acceptability.

Number of participants selecting each response option to the 6 TAQ questions at the end of the intervention will be assessed and plotted, and mean treatment acceptability will be derived.

Secondary aim 1 – Changes in associated variables

Analyses will focus on changes in scores between baseline and end-of-intervention for the following variables: STAI, MFQ, ADHD, SPEQ (3 sub-scales analysed separately), PMH, PSS. Each scale score will be used as the dependent variable in a mode as follows:

Type	Variable	Categorical/continuous
Dependent	Score at 2[1] follow-up assessments	Continuous
Predictors	**Group**: Control (1), Intervention (2) **Time**: Dummy variable coded as 2 (mid-assessment),3 (end-assessment)[1] **Group X Time interaction**	Categorical Categorical Categorical
Covariates	**Baseline STAI score** **Age**	Continuous Continuous

Model – Generalized estimating equation. Performed using the `xtgee` command in STATA.

Model to examine whether change in score across follow-up assessments is predicted by group (intervention vs control), after the effects of baseline score, age, and time of year have been

controlled for. Coefficients and 95% confidence intervals for the effect group will be plotted in a single forest plot to illustrate group effect sizes on changes in moderator variables.

[1] For ADHD and SPEQ models, time will not be entered as a predictor as there was no mid-assessment for these measures

Secondary analysis 2 – Moderators of treatment outcome

Potential moderators of SCI score at the end-of-intervention will be assessed using the following potential moderators: baseline STAI, MFQ, ADHS, SPEQ (3 sub-scales analysesd separately), PMH, PSS, LTE:

Type	Variable	Categorical/continuous
Dependent	SCI score at the end-of-intervention	Continuous
Predictors	**Group**: Control (1), Intervention (2)	Categorical
	Moderator: Baseline score of moderator variable	Continuous
	Group X moderator interaction	Categorical * continuous
Covariates	**Baseline SCI score**	Continuous
	Age	Continuous

Model – Linear regression. Preformed using the `regress` command in STATA.

Regression co-efficient and 95% confidence intervals for each of the interactions will be plotted in a single forest plot figure to illustrate effect sizes for each of the moderators. Standardized scores will be used to aid comparison between moderators.

Exploratory outcome 1 – Changes in general sleep quality

This outcome will be assessed using the same procedure as primary outcome 1, but focusing on PSQI scores rather than SCI scores.

Exploratory outcome 2 – Mediators of treatment outcome

Potential mediators of SC score at the end of the intervention will be assessed using the following potential mediators: PSQI, PSQI-A DBAS, PSAS, MCTQ (all assessed at the end of the intervention):

Type	Variable	Categorical/continuous
Dependent	SCI score at the end-of-intervention	Continuous
Predictors	**Group**: Control (1), Intervention (2) **Mediator:** Mediator score at the end of the intervention	Categorical Continuous
Covariates	**Baseline SCI score** **Age**	Continuous Continuous

Model – Bootstrapped mediated regression model with 5000 repetitions. Performed using the `sgmediation` and `bootstrap` commands in STATA.

Standardised coefficients and bias-corrected 95% confidence intervals for each indirect effect of group on SCI score via the mediator plotted in a single forest plot.