Integrated mental health care and vocational rehabilitation to improve return to work rates for people on sick leave because of common mental disorders (IBBIS)

Authors: Andreas Hoff¹ (corr. author), Jonas Fisker¹, Rie Mandrup Poulsen¹, Carsten Hjorthøj¹,², and Lene Falgaard Eplov¹

¹ Copenhagen Research Center for Mental Health [CORE], Mental Health Services Capital Region of Denmark, University of Copenhagen, Gentofte Hospitalsvej 1, Opgang 15-4, DK-2900
² University of Copenhagen, Department of Public Health, Section of Epidemiology

Correspondence: andreas.hoff@regionh.dk

Abbreviations:

RCT: Randomized controlled trial
RTW: Return to work
SDA: Study Design Article
IBBIS: Integriert Beskæftigelses- og Behandlingsindsats til Sygedagpengemodtagere (Danish)

1 ADMINISTRATIVE INFORMATION

The structure of this SAP is largely aligned with the recommendations by Gamble et. al¹.

1.1 TITLE AND TRIAL REGISTRATION

This SAP is the detailed statistical analysis plan, expanding the scientific IBBIS protocol of the two IBBIS randomized clinical trials (ClinicalTrials.gov Identifiers: NCT02872051 (RCT¹) and NCT02885519 (RCT²)):

RCT¹: “Integrated Mental Health Care and Vocational Rehabilitation to Individuals on Sick Leave Due to Anxiety and Depression (IBBIS)”

and

RCT²: “Integrated Mental Health Care and Vocational Rehabilitation to Individuals on Sick Leave Due to Stress Disorders (IBBIS)”.

Due to extensive methodological similarities between these studies this SAP applies to both, unless differences are mentioned explicitly.

1.2 SAP VERSION

This is the first version of the SAP.

1.3 PROTOCOL VERSION

Previous to publication of this SAP, plans have been described in both the protocol (published on clinicaltrials.org in the links provided), as well as in two study design articles (SDA), corresponding to the two RCTs²-³.

¹ https://clinicaltrials.gov/ct2/show/NCT02872051
² https://clinicaltrials.gov/ct2/show/NCT02885519
2 INTRODUCTION: BACKGROUND, RATIONALE AND OBJECTIVES

Described thoroughly in the SDAs\textsuperscript{2,3}. Furthermore, the protocol was published\textsuperscript{3} on the official webpage of the organization (Mental Health Services, Capital Region of Denmark).

3 STUDY METHODS

3.1 TRIAL DESIGN

See SDAs\textsuperscript{2,3}.

3.2 RANDOMIZATION

From SDA\textsuperscript{2}:

“The allocation ratio between the three arms is 1:1:1. A centralized randomization will take place according to a web-based computer-generated allocation sequence with varying block sizes kept unknown to the assessors. Odense Patient data Explorative Network (OPEN) is responsible for the randomization, administrative personnel in the IBBIS team perform the online randomization and the IBBIS team leader assign the participant to interventions and professionals.

We expect that service delivery can vary from municipality to municipality and the process of gaining a new job from unemployment will take longer time than returning to an existing job. Previous research has shown that diagnosis is a possible predictor of return to work\textsuperscript{4}. Thus, the randomization is stratified according to 1) municipality 2) employment status (on sick leave from work vs. on sick leave from unemployment) 3) diagnosis […]”

In RCT 1 diagnosis stratification is depression versus anxiety as primary diagnosis, and in RCT 2 diagnosis stratification is burnout vs. distress vs. adjustment disorder as primary diagnosis.

3.3 SAMPLE SIZE

Replicated from protocol follows:

*The sample size is based on a sample size calculation, using the ‘Power and Sample Size’ calculation programme\textsuperscript{4}.*

**Type I error (\(\alpha\)) risk**

In each of the two RCTs we wish to conduct multiple comparisons (between 3 groups), and hence significance level must be as follows, due to Bonferroni correction:

\[
\alpha = \frac{0,05}{3} = \frac{1}{60} = 0,0167
\]

**Type II error (\(\beta\)) risk**

The organizational constellation of the interventions has not yet been trialled, and thus the desired power shall be set to:

\[
\beta = 0,9
\]

If it turns out that we cannot include enough participants, the power could be set to:\(\beta = 0,8\)

**Hazard ratio (\(R\))**

The mean difference in time for return to work will be calculated as a hazard ratio. We estimate that as sufficient HR is

\[
R = 1,5
\]

\textsuperscript{3} https://www psykiatri-regionh.dk/Kvalitet-og-udvikling/udvikling/ibbis/Sider/IBBIS-forskning.aspx

\textsuperscript{4} http://ps-power-and-sample-size-calculation.software.informer.com
since just 50% faster return to work time in the intervention groups will convey a relevant economic benefit, due to the hence smaller loss of productivity.

**Mean time to return to work (M₁)**

Number of days from baseline to return to work is conservatively estimated to be 210 days, after an observed range from 104 to 210 days, in the control groups in three Dutch RCTs⁵⁻⁷, which were comparable to the control groups in the IBBIS RCTs. Hence,

\[ M₁ = 210 \]

**Inclusion time period (A)**

We will include participants through 24 months,

\[ A = 730 \text{[days]} \]

**Ratio between groups (m)**

Ratio is 1:1:1, and hence \( m = 1 \)

**Follow-up time (F)**

We will follow participants up for 365 days, in which they will contribute with risk time in the survival analysis, hence \( F = 365 \)

**Result**

In each group, due to the above-mentioned variables, we need \( 198 \) participants per group, and with three groups that yields a need for, with power = 0.9,

\[ N = 198 \frac{\text{participants}}{\text{group}} \times 3 \frac{\text{groups}}{\text{trial}} = 594 \text{ participants} \]

If, in case of insufficient inclusion possibilities, power could be lowered to 0.8. In such case we would need the following number:

\[ N = 153 \frac{\text{participants}}{\text{group}} \times 3 \frac{\text{groups}}{\text{trial}} = 459 \text{ participants} \]

### 3.4 STATISTICAL INTERIM ANALYSES AND STOPPING GUIDANCE

No interim analysis will be performed. We planned no stopping guidance.

### 3.5 TIMING OF FINAL ANALYSIS

The researchers who will perform the 6- and 12-month outcome analyses (AH and JP) will be blinded from intervention group allocation, until the primary outcome and all 12-month follow-up outcome main analyses are completed. The true randomization group allocation is concealed, with values X, Y and Z reflecting group allocation in the blinded dataset. The randomization allocation variable conversion formula is until unblinding only know and hidden by an administrative co-worker, who will not perform or assist any analysis.

At the time of publication of this SAP version, baseline distributional analyses, and unadjusted estimated marginal means-analyses of self-reported numerical secondary outcomes at 6-month follow-up (and only these) have been calculated, but will not be published, since this was not complying with the SDAs, nor this SAP.
4 STATISTICAL PRINCIPLES

4.1 CONFIDENCE INTERVALS AND P-VALUES
For all outcomes, the three randomization groups are pairwise compared. Due to these multiple comparisons, we will calculate 98.3% confidence intervals, according to Bonferroni correction of desired α-level of 0.05 in testing of 3 hypotheses:

\[ \alpha \text{-level: } 0.05 \times \frac{1}{3} \approx 0.0167 \Rightarrow \text{Confidence Interval: } 1 - 0.0167 \approx 98.33\% \]

4.2 ANALYSIS POPULATIONS
All analyses are performed as intention-to-treat, unless otherwise stated.

5 TRIAL POPULATION

5.1 WITHDRAWAL AND FOLLOW-UP
Due to legislative circumstances participants can withdraw consent, and followingly all person sensitive data on these subjects will be deleted, yet participant ID number (not CPR number, but generated for this research project) and randomization result will be stored. In sensitivity analyses these ID numbers will be included, as described in “handling of missing data”.

5.2 BASELINE PATIENT CHARACTERISTICS
The following will be reported per RCT, per randomization allocation group. For all mean values of numeric variables, standard deviations will be reported.

<table>
<thead>
<tr>
<th>Total number included in RCT and number in each randomization group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (mean, year)</td>
</tr>
<tr>
<td>Gender (%)</td>
</tr>
<tr>
<td>Bech Depr. Inventory (mean)</td>
</tr>
<tr>
<td>Bech Anxiety Inventory (mean)</td>
</tr>
<tr>
<td>WSAS (mean)</td>
</tr>
<tr>
<td>Perceived Stress Scale (mean)</td>
</tr>
<tr>
<td>Employment status (%; vacant vs. employed)</td>
</tr>
<tr>
<td>Primary diagnosis (%)</td>
</tr>
<tr>
<td>Municipality (%)</td>
</tr>
<tr>
<td>Sick leave duration at randomization (mean, days)</td>
</tr>
</tbody>
</table>

Distributional balances of these covariates will be calculated using one way-ANOVA for numerical data and \( X^2 \) for categorical data, and analyses with \( p \leq 0.05 \) will define imbalanced baseline covariates.

6 ANALYSIS
The first subsections of this section 6, describes general strategies applying to all analyses unless otherwise specifically stated. Subsection 6.8 contains the separate analysis strategies per outcome in 6.8.x.

6.1 COVARIATE ADJUSTMENT IN GENERAL
Analyses will be adjusted for the three stratification variables, and no other, complying with RCT analysis guidelines from European Medicines Agency.\(^5\)

6.2 Sensitivity analyses in general
As sensitivity analyses, all outcome analyses will be performed adjusted for any unbalanced baseline covariates, as defined in 5.2, Baseline patient characteristics.

Results of sensitivity analyses are only interpreted as supplements to the main analysis and will not substitute main results.

6.2.1 Sensitivity analyses for questionnaire based, self-reported data outcome
As sensitivity analyses, self-reported data outcomes (all other outcomes than numbered 1, 2, 3, 4, 9, 10 and 11 in tables in section 6.4, Outcome definitions) will be calculated with all missing outcome data replaced with a value equalling the mean of the outcome variable ± 2 standard deviations, and participants who withdraw themselves from the study will be included in these analyses with all their data handled as missing.

6.2.2 Sensitivity analyses for register data based outcomes
For register data-based outcomes (outcomes numbered 1, 2, 3, 4, 9, 10 and 11 in tables in section 6.4, Outcome definitions), sensitivity analyses will be performed including the participants who withdraw themselves from the study, included in these analyses with all their outcomes handled as either the worst possible (never returning to work) vs best possible (returning to work as soon as possible).

6.3 Subgroup analyses in general
All outcomes will be analysed with respect to the following subgroups:

a) per primary diagnosis (in RCT1 anxiety vs. depression; in RCT2 per distress, adjustment disorder, and burnout);
b) per employment status group at baseline (vacant vs. employed);
c) per IBBIS Team (two teams, Team North and Team Byen)

Furthermore,

d) divided in two groups by relative time of randomization: first and last temporal half of randomized participants.

Finally,

e) we will test for interaction between diagnostic group and treatment allocation group/arm.

No outcomes have other subgroup analyses planned.
### 6.4 Outcome Definitions

The outcomes are reported in the SDA and replicated in the following table, here with numbers 1 through 44 denoting the outcome numbers for reference purposes for this SAP section.

#### PRIMARY AND SECONDARY OUTCOMES and outcome numbering

<table>
<thead>
<tr>
<th>Outcome class</th>
<th>Data source</th>
<th>Outcome</th>
<th>6-month follow-up</th>
<th>12-month follow-up</th>
<th>24-month follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>DREAM database</td>
<td>Time from baseline to RTW</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary</td>
<td>DREAM database</td>
<td>Proportion in ordinary work</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DREAM database</td>
<td>Time from baseline to RTW</td>
<td>3</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DREAM database</td>
<td>Time from the first day of RTW until recurrent sick leave</td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Questionnaire</td>
<td>Depressive symptoms measured by Beck Depression Inventory (BDI)</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Questionnaire</td>
<td>Anxiety symptoms measured by Beck Anxiety Inventory (BAI)</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Questionnaire</td>
<td>Stress symptoms measured by Cohen perceived stress scale (PSS)</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Questionnaire</td>
<td>Social and work related function measured by WSAS</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcome class</td>
<td>Data source</td>
<td>Outcome</td>
<td>Follow-up</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-defined exploratory outcomes</td>
<td>DREAM database</td>
<td>Weeks of work from baseline to current follow-up</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Symptoms of Distress, anxiety, depression and somatization by Four-Dimensional Symptom Questionnaire (4DSQ)</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Depressive symptoms measured by Beck Depression Inventory (BDI)</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Anxiety symptoms measured by Beck Anxiety Inventory (BAI)</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stress-symptoms measured by Cohen perceived stress scale (PSS)</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Social and work related function measured by WSAS</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Burn-out symptoms measured by Karolinska Exhaustion Scale (KES)</td>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Health-related quality of life measured by EQ-5D-5L</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>General Quality of life scale measured by Flanagan’s’ QOLS</td>
<td>29</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Self-efficacy concerning symptoms measured by IPQ subscale on personal control</td>
<td>32</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Return to work self-efficacy measured by RTW-SE</td>
<td>35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Questionnaires</td>
<td></td>
<td>General self-efficacy measured by General Self-efficacy scale (GSS)</td>
<td>38</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Client satisfaction with treatment measure measured by CSQ-8</td>
<td>41</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Presenteeism measured by Stanford Presenteeism Scale (SPS)</td>
<td>42</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
6.5 **HYPOTHESES AND NULL-HYPOTHESES**

Stated below are the generic versions of all three hypotheses \((H_1)\) and all three null-hypotheses \((H_0)\) that apply to each outcome.

Regarding what is a “better outcome” is listed in section 6.6, defined for each outcome measure, respectively.

6.5.1 **HYPOTHESES**

This superiority trial hypothesizes that, for all outcomes,

\(H_{1A}\)  Group 3, “Integrated IBBIS mental health care treatment and vocational rehabilitation” conveys better outcomes than

Group 2, “IBBIS mental health care (and standard VR)”, and

\(H_{1B}\)  Group 2, “IBBIS mental health care (and standard VR)”, and

conveys better outcomes than

Group 1, “Control group, treatment as usual (standard MHC and standard VR)” and followingly

\(H_{1C}\)  Group 3, “Integrated IBBIS mental health care treatment and vocational rehabilitation” conveys better outcomes than

Group 1, “Control group, treatment as usual (standard MHC and standard VR)”.

and followingly

Group 3 conveys better outcomes than Group 1, since if

\[
\text{Group 3 outcome} > \text{Group 2 outcome} > \text{Group 1 outcome}
\]

then

\[
\text{Group 3 outcome} > \text{Group 1 outcome}.
\]

The groups are thoroughly described in the IBBIS Protocol and the SDAs.
6.5.2 NULL-HYPOTHESES
The corresponding null-hypotheses are

**H$_{0A}$**  Group 3, “Integrated IBBIS mental health care treatment and vocational rehabilitation”

*does not convey better outcomes than*

Group 2, “IBBIS mental health care (and standard VR)”, and

**H$_{0B}$**  Group 2, “IBBIS mental health care (and standard VR)”,

*does not convey better outcomes than*

Group 1, “Control group, treatment as usual (standard MHC and standard VR)”.

and followingly

**H$_{0C}$**  Group 3 *does not convey better outcomes than* Group 1.
### 6.6 Outcome Benefit Direction

Referring to the hypothesis section, this table describes whether a “better outcome” is a higher or lower score on the numeric outcome variables.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Is “better outcome” defined by lower or higher numbers?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time from baseline to RTW</td>
<td>Lower</td>
</tr>
<tr>
<td>Proportion in ordinary work</td>
<td>Higher</td>
</tr>
<tr>
<td>Time from baseline to RTW</td>
<td>Lower</td>
</tr>
<tr>
<td>Time from the first day of RTW until recurrent sick leave</td>
<td>Higher</td>
</tr>
<tr>
<td>Depressive symptoms measured by Beck Depression Inventory (BDI)&lt;sup&gt;8&lt;/sup&gt;</td>
<td>Lower</td>
</tr>
<tr>
<td>Anxiety symptoms measured by Beck Anxiety Inventory (BAI)&lt;sup&gt;9&lt;/sup&gt;</td>
<td>Lower</td>
</tr>
<tr>
<td>Stress symptoms measured by Cohen perceived stress scale (PSS)&lt;sup&gt;10&lt;/sup&gt;</td>
<td>Lower</td>
</tr>
<tr>
<td>Social and work related function measured by WSAS&lt;sup&gt;11&lt;/sup&gt;</td>
<td>Lower</td>
</tr>
<tr>
<td>Weeks of work from baseline to current follow-up</td>
<td>Higher</td>
</tr>
<tr>
<td>Symptoms of Distress, anxiety, depression and somatization by Four-Dimensional Symptom Questionnaire (4DSQ)&lt;sup&gt;12&lt;/sup&gt;</td>
<td>Lower</td>
</tr>
<tr>
<td>Burn-out symptoms measured by Karolinska Exhaustion Scale (KES)&lt;sup&gt;13&lt;/sup&gt;</td>
<td>Lower</td>
</tr>
<tr>
<td>Health-related quality of life measured by EQ-5D-5L&lt;sup&gt;14&lt;/sup&gt;</td>
<td>Higher&lt;sup&gt;21&lt;/sup&gt;</td>
</tr>
<tr>
<td>General Quality of life scale measured by Flanagan’s’ QOLS&lt;sup&gt;15&lt;/sup&gt;</td>
<td>Higher</td>
</tr>
<tr>
<td>Self-efficacy concerning symptoms measured by IPQ subscale on personal control&lt;sup&gt;16&lt;/sup&gt;</td>
<td>Higher&lt;sup&gt;22&lt;/sup&gt;</td>
</tr>
<tr>
<td>Return to work self-efficacy measured by RTW-SE&lt;sup&gt;17&lt;/sup&gt;</td>
<td>Higher</td>
</tr>
<tr>
<td>General self-efficacy measured by General Self-efficacy scale (GSS)&lt;sup&gt;18&lt;/sup&gt;</td>
<td>Higher</td>
</tr>
<tr>
<td>Client satisfaction with treatment measure measured by CSQ-8&lt;sup&gt;19&lt;/sup&gt;</td>
<td>Higher</td>
</tr>
<tr>
<td>Presenteeism measured by Stanford Presenteeism Scale (SPS)&lt;sup&gt;20&lt;/sup&gt;</td>
<td>Higher&lt;sup&gt;23&lt;/sup&gt;</td>
</tr>
</tbody>
</table>
6.7 Missing data in general
In general, proportion of missing data will be reported per intervention group for all outcomes.

6.7.1 Handling of missing data in registers
In RWT-outcomes (outcomes numbered 1, 2, 3, 4, 9, 10 and 11 in tables in section 6.4, Outcome definitions) we expect no missing data, due to the nature of the Dream Register. Missing data will only be in case of a participant moving out of Denmark. We consider these events to be so rare in our data that we will handle such missing data as missing completely at random. Thus, no imputation or other correction is necessary. We will report proportion of data missing.

We will report number of censored participants per treatment group.

6.7.2 Handling of missing data in questionnaire based, self-reported data outcome
For these outcomes (all other outcomes than numbered 1, 2, 3, 4, 9, 10 and 11 in tables in section 6.4, Outcome definitions), missing data will be handled as missing at random. To handle this, 100 multiple imputations will be performed, using following variables: stratification variables: diagnosis, municipality, employment status; age; gender; time to stable RTW; psychometric variables at baseline and all follow-up at outcome time: BDI, BAI, WSAS and PSS.

6.8 Analysis methods per outcome group
This section describes the details of the statistical analyses. Since several outcomes require exact same analysis methods, outcomes are grouped for the following description

6.8.1 Time to return to work-outcomes (Outcomes #1, #3 and #9)
This section describes primary outcome Time from baseline to RTW at 12-month follow-up (1), and the secondary outcomes Time from baseline to RTW at 6- (outcome 3) and 24-month follow-up (outcome 9). The 24-month follow-up outcome will be calculated no earlier than June 2020. The other two, readily after the publication of this SAP, but before unblinding of analysists.

6.8.1.1 Calculation of the outcome: Specific measurement and units (and transformation, where applicable)
Time from baseline to RTW is defined as the number of weeks from randomization date, to stable return to work. Stable return to work is defined as 4 weeks consecutively in work, i.e. with no sick leave benefit those 4 weeks in the Dream register, and a so-called “branch code” in at least some of this 4 week period (benefit codes are week-based, branch codes are month based, and hence a period of 4 weeks may represent only one month, or overlap a two month period; in the latter case, return to work will be attained if at least one of these registrations contains a branch code; a branch code means that the individual received salary from an employer in this period). Time of event is first day of the four weeks.

These events will define censoring: 1) moving out of the country, 2) death, 3) public retirement pension (Da.: “Folkepension”), and 4) voluntary early retirement scheme (Danish: “Efterløn”).

At randomization all participants are, according to inclusion criteria, on sick-leave from employment or vacancy. Some participants might be on sick-leave from an employment in a flexjob\(^6\), and hence receiving flexjob benefit during employment. This benefit is changed to flexjob sick-leave benefit similar to regular sick leave benefit for participants not granted flexjob benefit prior to randomization. In these cases (of participants granted flexjob benefit prior to randomization) RTW is defined as either not receiving flexjob sick-leave benefit for four consecutive weeks, along with a registered branch code as above mentioned (or alternatively not receiving flexjob benefit, but an ordinary salary indicated by a branch code during those four weeks).

For participants, who at baseline are on sick-leave from vacancy (but not receiving flexjob benefit), RTW can both be defined as above mentioned (four consecutive weeks without sick leave benefits and a branch code during those four weeks) or receiving flexjob benefit for four consecutive weeks and a branch code during those four weeks.

\(^6\) “Flexjob” is one of the Danish benefit schemes; it is a subsidy granted those with a chronic reduced work capacity.
6.8.1.2 Specific analysis method and result presentation
Comparisons of RTW time will be calculated as hazard rate ratios between groups (and corresponding 98.3\%CI), using a Cox-regression model.

6.8.1.3 Covariate adjustment
Only for stratification variables, see 6.1 “Covariate adjustment in general“.

6.8.1.4 Statistical method assumption control
Assumptions for the proportional hazards (~Cox-) regression model are proportional hazards; this will be controlled performing af Schoenfeld (SF) test for residuals and visual inspection.

6.8.1.5 Alternative analysis method in case of assumption fail
If the SF test is positive (p<0.05), the analysis will we performed adjusted for the interaction between time and treatment group allocation. If SF test hereafter is still positive, the analysis will instead be adjusted for the interaction between quadratic time (time\(^2\)) and treatment group allocation. If SF test hereafter is still positive, the analysis will instead be adjusted for the interaction between log(time) and treatment group allocation. If SF test hereafter is still positive, the analysis with the highest p-value will be reported.

6.8.1.6 Sensitivity analyses
See “6.2.2 Sensitivity analyses for register data based outcomes”.

6.8.1.7 Reporting and statistical methods to handle missing data
On RWT-outcomes we expect no missing data, due to the nature of the Dream Register. Missing data will only be in case of a participant dying or moving out of Denmark. We consider these events to be so rare in our data that we will handle such missing data as missing completely at random. Thus, no imputation or other correction is necessary. We will report proportion of data missing.

We will report number of censored participants per treatment group.

6.8.2 Proportion in ordinary work at 12-month follow-up, secondary outcome (outcome #2)

6.8.2.1 Calculation of the outcome: specific measurement and units (and transformation, where applicable)
This outcome is calculated as the share of the treatment allocation group that on the 12-month follow-up date was in stable RTW (≥ 4 weeks). Stable RTW if defined exactly as in the primary outcome, see 6.8.1.1.

6.8.2.2 Specific analysis method and result presentation
Pairwise odds ratios will be calculated using logistic regression.

6.8.2.3 Covariate adjustment
Only for stratification variables, see 6.1 “Covariate adjustment in general“.

6.8.2.4 Statistical method assumption control
The assumptions of the model are assumed to be acceptable, due to large sample, binary outcome, categorical independent variable.

6.8.2.5 Alternative analysis method in case of assumption fail
No alternative methods are planned, since assumptions are assumed to hold.

6.8.2.6 Sensitivity analyses
See “6.2.2 Sensitivity analyses for register data based outcomes”.

6.8.2.7 Reporting and statistical methods to handle missing data
Same as 6.8.1.7.
6.8.3 All self-reported, numerical outcomes, at 6-, 12-, and 24-month follow-up at
(secondary outcomes ##5-8 and predefined exploratory outcomes ##12-44)

6.8.3.1 Calculation of the outcome: specific measurement and units (and transformation, where applicable)
All outcomes are calculated as the sum of scores on the respective measurement scales.
All 6-month follow-up outcome analyses are calculating using baseline and 6-month follow-up observations.
All 12-month follow-up outcome analyses are calculating using baseline and 6- and 12-month follow-up observations.
All 24-month follow-up outcome analyses are calculating using baseline and 6-, 12-, and 24-month follow-up observations.

6.8.3.2 Specific analysis method and result presentation
Linear mixed-effects model with unstructured covariance. Results will be presented in pairwise group differences between outcomes, from the estimated marginal means from the model, and the confidence intervals of these differences.

6.8.3.3 Covariate adjustment
Only for stratification variables, see 6.1 “Covariate adjustment in general“.

6.8.3.4 Statistical method assumption control
Assumption: normal distribution of scores. Control: Visual inspection by plotting the score residuals.
Assumption: normal distribution of individuals’ score differences between baseline and follow-up. Control: Visual inspection by plotting the score difference residuals.
Assumption: Equality and homogeneity of variance. Control: Breusch Pagan test and Bartlett’s test are used to identify violations of these assumptions.

6.8.3.5 Alternative analysis method in case of assumption fail
In case of positive tests or visual inspections a robust variance estimator is used to correct standard errors.

6.8.3.6 Sensitivity analyses
See “6.2.1 Sensitivity analyses for questionnaire based, self-reported data outcome”.

6.8.3.7 Reporting and statistical methods to handle missing data
Proportion and amount of missing data per outcome variable per follow-up event per treatment group will be reported.
To handle missing data, 100 multiple imputations will be performed, using following variables:
stratification variables: diagnosis, municipality, employment status; age; gender; time to stable RTW;
psychometric variables at baseline and all follow-up at outcome time: BDI, BAI, WSAS and PSS.

6.8.4 Weeks of work from baseline to 12- and 24-month follow-up (outcomes ##10-11)

6.8.4.1 Calculation of the outcome: specific measurement and units (and transformation, where applicable)
From baseline to follow-up, the number of weeks in work per participant is calculated. A week is noted as being in work, if no sick leave benefit has been received, and if a branch code is registered in the month of that week (branch codes are registered on monthly basis, if an individual has received salary from an ordinary job during that month).
For participants receiving flexible job benefit prior to randomization, and participants on sick leave from vacancy, the same principles apply, as described in 6.8.1.1, in the section “Time to return to work-outcomes (outcomes #1, #3 and #9)"
6.8.4.2 **Specific analysis method and result presentation**

Severely skewed data is expected in for this outcome, why a robust Poisson regression model will be used to test the differences between groups.

6.8.4.3 **Covariate adjustment**

Only for stratification variables, see 6.1 “Covariate adjustment in general”.

6.8.4.4 **Statistical method assumption control**


6.8.4.5 **Alternative analysis method in case of assumption fail**

If X² goodness-of-fit test is significant, negative binomial regression model will be used instead. If X² goodness-of-fit test is significant for this distribution, zero inflated poisson regression will be used.

6.8.4.6 **Sensitivity analyses**

See “6.2.2 Sensitivity analyses for register data based outcomes”.

6.8.4.7 **Reporting and statistical methods to handle missing data**

See 6.8.1.7

6.8.5 **At 24 months: Time from the first day of RTW until recurrent sick leave (outcome #4)**

*The calculation and analysis of this outcome has been cancelled, after the publication of the SDA. No analysis attempts have been made prior to this change (decision made June 27th, 2019)!*

Reason: During the development of this SAP, the authors realized, through thorough discussions that this outcome cannot validly be calculated and analysed. The intention with the outcome was to examine whether the IBBIS interventions convey a lower risk of recurrent sick leave after RTW, compared to the control group.

The intention was to perform a Cox-regression analysis, with risk time starting after RTW. The duration between randomization and RTW would vary largely between individuals, and since no other randomization takes place, this covariate would not be equally distributed between groups, as well as possibly a long range of other covariates influencing the risk of recurrent sick-leave. Adjusting for these covariates would not be sufficient to attribute risk rates in the groups to the interventions.

The authors consider including other strategies for examining risk for recurrent sick leave in an update of this SAP.
7 REFERENCES


