BSOCIAL RCT
Main Data File Documentation
de jure Estimand

David M. Erceg-Hurn

2019-11-27

Overview

This file contains an overview of the variables in the main data file for the BSOCIAL RCT.

There are over 2000 variables in this dataset. You will only need a relatively small subset of these for the analyses that will be conducted for any particular paper.

Sample

There were 107 patients randomly allocated to treatment in the trial. The main data file should contain data for 105 these patients. Two were removed for providing invalid responses, as noted on page 10 of the statistical anlaysis plan.

Data Format

The data are arranged in 'wide' rather than 'long' format. This means there is one row of data per patient. In order to perform some longitudinal analyses, you will likely need to reshape the data into 'long' format, using a tool such as the R package tidyr.

Estimand

The estimand specified in the statistical analysis plan for the main paper is the 'de jure' estimand. Data collected for patients who missed three or more consecutive treatment sessions (defined as treatment session 1 to session 13) has been set to missing (NA). If you wish to conduct research using a different estimand, contact me and I will prepare an appropriate data file for you.

I will now provide an overview of the variables in the data file.

ID and Administrative Variables

Patient Identifier

The first variable in the dataset is an identifier for each randomised patient. This is an integer that can take values between 1 and 107.

Variable	Description
bsocial_id	Random patient identifier variable (used for de-identified data sharing and analysis)

Other Administrative Variables

Variable	Description
bsocial_studyid	Trial id variable (created by matt)
bsocial_tx	Did the patient start treatment in the bsocial trial?
bsocial_treat	Treatment code for all patients allocated to treatment (real labels)
bsocial_treat2	Treatment code for all patients allocated to treatment (random letters)
bsocial_gid	Group identifier for each patient allocated to treatment (real labels)
bsocial_gid2	Group identifier for each patient allocated to treatment (random letters)
bsocial_ntx	Number of sessions attended (out of 13) for patients who were allocated to treatment
bsocial_ntx2	Number of sessions attended (out of 13) for patients who started treatment
date_a1	Date of A1 (initial assessment)
ax_clin	Assessing clinician
therapist1	Treating clinician 1
therapist2	Treating clinician 2

`bsocial_studyid` is an alternative patient identifier that was used by clinicians and the trial diagnostician during trial data collection (e.g., `4-140`, `5-148`). The class of this variable in *R* is a character ("string" in SPSS terminology). You are unlikely to need this variable in your analyses. It is included in the data file for two reasons: (a) in order to link the measures in the main data file with other data sources (e.g,. TIC-P and Psychophys) at a later date, and (b) in case there is a need to lookup hard copies of a patient's questionnaires. Both the hard copies and other data sources use alternative patient identifier, rather than the `bsocial_id`, to identify patients.

`bsocial_treat` contains the actual names of the two treatments (i.e., `verbal` and `imagery`). The `bsocial_treat2` variable is a recode of that, with the labels changed randomly to `A` or `B`.

`bsocial_gid` contains the real group identifiers, in the format `SAG -- Start Date -- Start Time`. For example, `SAG 2016-08-17 09:30`. The `bsocial_gid2` variable is a recode of `bsocial_gid`, with the group

labels changed randomly to letters between A and L.

To keep the analyst blind to treatment condition (and group, which could potentially be used to determine the treatment type), the `bsocial_treat` and `bsocial_gid` are NOT being supplied to the data analyst (Andrew) until after all analyses specified in the statistical analysis plan have been conducted.

bsocial_ntxandbsocial_ntx2contain information about the number of treatment sessions attended. The difference between them is that clients allocated to treatment, but who didn't start treatment (bsocial_treat == "No") have had their scores set to0in 'bsocial_ntx'; in `bsocial_ntx2` they are set to NA (missing). The `bsocial_ntx2` variable should be used when computing the average number of sessions per group for the trial, to keep the estimate consistent with how the number of treatment sessions attended has been computed in previous publications. The `bsocial_ntx` variable should be used if you want to include number of sessions included in imputation models. Setting the missing values to zero for patients who did not start treatment, as `bsocial_ntx` does, will prevent them from being erroneously imputed.

Session Attendance

Variable	Description
bsocial_t1	Treatment session 1 attendance code (from pete's log)
bsocial_t2	Treatment session 2 attendance code (from pete's log)
bsocial_t3	Treatment session 3 attendance code (from pete's log)
bsocial_t4	Treatment session 4 attendance code (from pete's log)
bsocial_t5	Treatment session 5 attendance code (from pete's log)
bsocial_t6	Treatment session 6 attendance code (from pete's log)
bsocial_t7	Treatment session 7 attendance code (from pete's log)
bsocial_t8	Treatment session 8 attendance code (from pete's log)
bsocial_t9	Treatment session 9 attendance code (from pete's log)
bsocial_t10	Treatment session 10 attendance code (from pete's log)
bsocial_t11	Treatment session 11 attendance code (from pete's log)
bsocial_post	Post-treatment attendance code (from pete's log and cci clinician info)
bsocial_fu	1-month follow up attendance code (from pete's log and cci clinician info)
bsocial_post_original	Post-treatment attendance code (from pete's log only)
bsocial_fu_original	1-month follow up attendance code (from pete's log only)

These variables indicate whether patients attended, cancelled, did not attend etc each treatment session. The codes were entered by the chief investigator (Prof. McEvoy) during the trial.

Occasionally patients who cancelled or did not attend a group session had a 'catch up' session with a trial clinician, however this has not been taken into account in the coding of these variables (except for `bsocial_post` and

bsocial_fu, for which CCI's treating clinicians recorded information which could be used to determine if a client attended a catch up session). Thus, for bsocial_t1 though bsocial_t11 there are some patients who attended catch up sessions, but have a code other than 'Attended'.

A consequence of this is that if you sum the number of sessions coded as 'Attended' for a particular patient, you will find there are 16 cases where the number differs from that in the bsocial_ntx2 variable. This is because bsocial_ntx2 variable was entered by CCI clinicians counting catch up appointments as a treatment session, whereas the bsocial_t1 to bsocial_t11 variables do not do this.

In order to calculate the average number of sessions attended, use the bsocial_ntx2 variable – do not sum and then average the data in bsocial_t1 through bsocial_fu.

DSM-5 Diagnoses

The next set of variables relate to DSM-5 diagnoses collected as part of the pre-treatment diagnostic interview.

Clinician-Entered Variables

These are variables for which the data was entered by the assessing clinician.

Diagnoses

Variable	Description
dsm5_pri	Principal diagnosis
dsm5_ad1	Additional diagnosis 1
dsm5_ad2	Additional diagnosis 2
dsm5_ad3	Additional diagnosis 3
dsm5_ad4	Additional diagnosis 4
dsm5_ad5	Additional diagnosis 5
dsm5_ad6	Additional diagnosis 6

These variables contain each patient's diagnoses (e.g., Social Anxiety Disorder, Panic Disorder, Agoraphobia).

One principal diagnosis and up to six additional ("comorbid") diagnoses were recorded. The ordering of the additional diagnoses is largely arbitrary, and usually a reflection of the order the diagnoses appear in the structured diagnostic interview (SCID). Do not interpret the ordering of the additional diagnoses as a hierarchy.

Be aware that each of the diagnostic variables has 48 possible levels, however only a handful of these were actually observed in the trial. For example, for dsm5_pri, every patient has the same diagnosis – Social Anxiety Disorder.

For dsm_ad6, only two of the levels are used (1 patient has a diagnosis of Obsessive Compulsive Disorder, and 104 a diagnosis of None).

All the diagnoses were current at the time of assessment (i.e. the patient was in an episode, not in remission), with the exception of Major Depressive Disorder, for which we record diagnoses that are Current, In partial remission, and In full remission. The same is true of Bipolar disorder, however, no patients in this data set should have a Bipolar diagnosis as that was a trial exclusion criterion.

Generally you won't need to use these variables in your analyses – you will instead use other variables (discussed later) that were derived from them.

Severity Ratings

Variable	Description
sev_pri	Severity rating - principal diagnosis
sev_ad1	Severity rating - additional diagnosis 1
sev_ad2	Severity rating - additional diagnosis 2
sev_ad3	Severity rating - additional diagnosis 3
sev_ad4	Severity rating - additional diagnosis 4
sev_ad5	Severity rating - additional diagnosis 5
sev_ad6	Severity rating - additional diagnosis 6

These are the severity ratings from the SCID.

If a client has for example has only one additional diagnosis, then the severity rating for the variables sev_ad2 through sev_ad6 will be missing. This is by design. **Do not impute these missing values.**

Age at Onset

Variable	Description
onset_pri	Age at onset - principal diagnosis
onset_ad1	Age at onset - additional diagnosis 1
onset_ad2	Age at onset - additional diagnosis 2
onset_ad3	Age at onset - additional diagnosis 3
onset_ad4	Age at onset - additional diagnosis 4
onset_ad5	Age at onset - additional diagnosis 5
onset_ad6	Age at onset - additional diagnosis 6

These variables record each client's age (in years) at the time symptoms of the each diagnosis first became a problem (i.e. age at onset of the first episode).

Once again, these variables by design have missing values if the client has less than six additional diagnoses. **Do not impute the missing values.**

Duration of the Current Episode

Variable	Description
duration_pri	Duration of the current episode - principal diagnosis
duration_ad1	Duration of the current episode - additional diagnosis 1
duration_ad2	Duration of the current episode - additional diagnosis 2
duration_ad3	Duration of the current episode - additional diagnosis 3
duration_ad4	Duration of the current episode - additional diagnosis 4
duration_ad5	Duration of the current episode - additional diagnosis 5
duration_ad6	Duration of the current episode - additional diagnosis 6

These variables store the number of months since the onset of the current episode of each diagnosis.

This will be missing for `Major Depression in Full Remission`. This is by design. **Do not impute these missing values**

DSM5 Specifiers, Episode Details, etc

The next set of variables store additional information about select diagnoses (e.g., number of depressive episodes). Once again, most patients will have missing data on these variables by design. **Do not impute these missing values.**

Variable	Description
mde	Number of major depressive episodes
mde_type	Mde recurrence specifier
pdd	Persistent depressive disorder - specifier for most recent 2 years
sad_spec	Sad - performance only subtype specifier
specific1	Specific phobia - subtype
specific4	Specific phobia - text string describing other subtype

Note that the variables `specific2` and `specific3` have been excluded deliberately (these are used for patients with more than one specific phobia, and no patient in the BSOCIAL trial was diagnosed with more than one type of specific phobia).

Derived Variables

The next set of variables I *derived* from the aforementioned diagnostic variables. Derived variables are also known as *synthetic variables*. They are useful for doing certain analyses (e.g., calculating the average number of diagnoses) or filtering the dataset (e.g., to cases with comorbid depression). These variables will be useful for reporting the clinical characteristics of the sample, and also useful in analyses focused on predictors and moderators of change.

Years Since Onset

Variable	Description
onset_yrs_pri	Years since onset - principal diagnosis
onset_yrs_ad1	Years since onset - additional diagnosis 1
onset_yrs_ad2	Years since onset - additional diagnosis 2
onset_yrs_ad3	Years since onset - additional diagnosis 3
onset_yrs_ad4	Years since onset - additional diagnosis 4
onset_yrs_ad5	Years since onset - additional diagnosis 5
onset_yrs_ad6	Years since onset - additional diagnosis 6

These variables were derived by subtracting age at onset of each disorder from age at the time of assessment.

e.g., `onset_yrs_pri = age - onset_pri`

Be aware that the number of years is a bit fuzzy. For example, if someone is 18 years and 11 months, and the age of onset was when they were 18 years and 1 month, the number of years since onset will be computed as 0, whereas if the person is 19 years 1 month, and the age of onset was two months prior (18 years 11 months), then the age at onset will be computed as 1 year.

These variables will by design have missing values if the client has less than six additional diagnoses. **Do not impute these missing values.**

Number of Diagnoses

Variable	Description
dsm5_num1	Number of diagnoses (current diagnoses plus those in partial remission plus those in full remission)
dsm5_num2	Number of diagnoses (current diagnoses plus those in partial remission)
dsm5_num3	Number of diagnoses (current diagnoses only)

These variables indicate the number of diagnoses each patient has.

The only difference between the variables is whether `Major Depression in Partial Remission`, and `Major Depression in Full Remission` are counted as diagnoses.

Presence or Absence of Comorbidity

Variable	Description
dsm5_comorb	Does the client have at least one comorbidity (current diagnoses only)?
dsm5_nocomorb	Does the client have only a single diagnosis (ie no comorbidities)(current diagnoses only)?

`dsm5_comorb` indicates whether the patient has at least one current comorbidity. `dsm5_nocomorb` is the opposite – it indicates whether the client does not have a comorbidity.

For the purpose of computing these variables, Major Depression in partial and full remission were regarded as ***not*** being current diagnoses. In other words, if a client had only one comorbid diagnoses – `Major Depression in Partial Remission` – the client would be coded as not having any (current) comorbidities.

Presence or Absence of Specific Diagnoses (Narrow Groupings)

The next set of variables indicate whether a patient has (or does not have) specific disorders, such as panic disorder, agoraphobia, etc.

Variable	Description
dsm5_mdd_cur	Was the client diagnosed with current major depressive disorder?
dsm5_mdd_par	Was the client diagnosed with major depressive disorder in partial remission?
dsm5_mdd_full	Was the client diagnosed with major depressive disorder in full remission?
dsm5_pdd	Was the client diagnosed with persistent depressive disorder?
dsm5_agor	Was the client diagnosed with agoraphobia?
dsm5_panic	Was the client diagnosed with panic disorder?
dsm5_specific	Was the client diagnosed with specific phobia?
dsm5_gad	Was the client diagnosed with generalised anxiety disorder?
dsm5_ocd	Was the client diagnosed with obsessive-compulsive disorder?
dsm5_bdd	Was the client diagnosed with body dysmorphic disorder?
dsm5_ptsd	Was the client diagnosed with post-traumatic stress disorder?
dsm5_iad	Was the client diagnosed with illness anxiety disorder?
dsm5_ss	Was the client diagnosed with somatic symptom disorder?

Presence of Absence of Specific Diagnoses (Broader Groupings)

These variables are higher-order clusterings of the diagnoses. For example, `dsm5_dep` indicates whether the patient has "depression", which is defined broadly as a diagnosis of Major Depression and Persistent Depression (Dysthymia).

Variable	Description
dsm5_dep	Was the client diagnosed with a current depressive disorder (mdd or pdd)?
dsm5_mdd_ever	Has the client ever been diagnosed with major depression (mdd current, mdd in partial remission, mdd in full remission)?
dsm5_double_dep	Does the client have double depression (current mdd and pdd)?
dsm5_anx	Was the client diagnosed with a comorbid anxiety disorder (agoraphobia, panic disorder, gad, specific phobia)?
dsm5_ocd_related	Was the client diagnosed with an obsessive-compulsive related disorder (ocd, bdd, trichotillomania, excoriation, other specified ocd)?
dsm5_trauma	Was the client diagnosed with a trauma- and stressor-related disorder (asd, ptsd)?
dsm5_somatic	Was the client diagnosed with a somatic symptom and related disorder (illness anxiety or somatic symptom disorder)?

Demographics and Related Information

General Demographics

These variables are common demographics such as age and sex.

Variable	Description
age	Age in years at time of assessment (a1 for mad; a0 or a1 for ed)
sex	Sex
birthplace	Place of birth
cultural	Cultural background
employed	Employment status (current system - with a clear definition of employment)
student	Currently a student?
education	Highest level of education completed (current coding system)
relationship	Relationship status (current coding system)

The class of these variables is either *factor* (e.g., `sex`, `relationship`) or *numeric* (`age`), except for `cultural` which is a *character* (string). If you wish to use the `cultural` variable, you will need to categorise the responses yourself. It

would probably be easiest just to report the percentage of Anglo/European Australians, as that is the most common cultural background.

Prior Treatment History

These variables provide information about whether patients had previously tried a professional treatment for their mental health problems, and whether they have been admitted to a psychiatric hospital.

Variable	Description
prev_tx	Prior tx - has the client received any prior treatment for their psychological problems
hosp	Prior tx - has the client been hospitalised for psychiatric problems
invol	Prior tx - has the client been hospitalised involuntarily for psychiatric problems

Suicide Attempts and Self-Harm

These variables provide information about each patient's history of suicide and self-harm attempts.

Variable	Description
suicide_num	Suicide attempts - number (ordinal categories)
selfharm_num	Self harm attempts - number (ordinal categories)
suicide	Ever attempted suicide?
selfharm	Ever engaged in any self-harm (besides suicide attempts)?
sosh	Ever attempted suicide or self-harmed?

The categories for the `suicide_num` and `selfharm_num` variables are: `none`, `1 attempt`, `2 - 4 attempts` and `>4 attempts`.

The `suicide` variable is derived from the `suicide_num` variable. The `1 attempt`, `2 - 4 attempts` and `>4 attempts` categories have been recoded as `Yes`, and none recoded as `No`. The `self_harm` variable uses the same strategy to recode the `selfharm_num` variable.

`sosh` is a recoded version of `suicide` and `selfharm`. Any client who has attempted suicide **or** self-harmed is coded as `Yes`; clients who have not self-harmed ***and*** not attempted suicide are coded as `No`.

Psychotropic Medication

Medications Being Used at Pre-treatment Assessment

Variable	Description
medicated	Using any psychotropic meds at time of assessment?
ad	Using antidepressants at assessment?
anx	Using anxiolytics at assessment?
ms	Using mood stabilizers at assessment?
ap	Binary - using antipsychotics at assessment
ad1	Antidepressant 1
ad2	Antidepressant 2
anx1	Anxiolytic 1
anx2	Anxiolytic 2
ms1	Mood stabilizer 1
ap1	Antipsychotic 1
med_start	How long since first started taking meds? (Ordinal)
med_stable	How long since last started taking meds? (Ordinal)

There is no missing data for the variables `medicated` through `ap`.

For the other variables, some data will be missing (e.g., if patients were not taking any medications, they will not have any values recorded for the other variables). **The data are missing by design and should not be imputed**.

Note that for a handful of patients, the time since last change medication variable (`med_stable`) is less than one month. However by the time of the first treatment session, all of these patients would have been stable on medication for at least a month as this was a trial entry criterion.

Change in Medication Use During Treatment

These variables can be used to ascertain whether patient's stopped, started, or changed medications or dosages during treatment.

Variable	Description
pmudt_nomeds	Use during tx - were no medications used during treatment?
pmudt_nochange	Use during tx - was there no change to the medications the client was using at pre-treatment?
pmudt_increase	Use during tx - was the dosage of a medication increased?
pmudt_increase_text	Use during tx - string specifying which medication(s) increased in dosage?
pmudt_reduced	Use during tx - was the dosage of any medication reduced?
pmudt_reduced_text	Use during tx - string specifying which medication(s) decreased in dosage

Follow-Up Diagnoses

Are Pre-treatment Diagnoses Still Present?

Variable	Description
sp_pri_fu	Still present at 1 month follow up - principal diagnosis
sp_ad1_fu	Still present at 1 month follow up - additional diagnosis 1
sp_ad2_fu	Still present at 1 month follow up - additional diagnosis 2
sp_ad3_fu	Still present at 1 month follow up - additional diagnosis 3
sp_ad4_fu	Still present at 1 month follow up - additional diagnosis 4
sp_ad5_fu	Still present at 1 month follow up - additional diagnosis 5
sp_ad6_fu	Still present at 1 month follow up - additional diagnosis 6
sp_pri_fu6m	Still present at 6 month follow up - principal diagnosis

For these variables, data can be missing because (a) the patient did not have additional diagnoses at pre-treatment, or (b) because they never attended the follow up assessments. **You should NOT impute missing values for clients who are missing data because they did not have additional diagnoses at pre-treatment**.

Severity Ratings

Variable	Description
sev_pri_fu	Severity rating at 1 month follow up - principal diagnosis
sev_ad1_fu	Severity rating at 1 month follow up - additional diagnosis 1
sev_ad2_fu	Severity rating at 1 month follow up - additional diagnosis 2
sev_ad3_fu	Severity rating at 1 month follow up - additional diagnosis 3
sev_ad4_fu	Severity rating at 1 month follow up - additional diagnosis 4
sev_ad5_fu	Severity rating at 1 month follow up - additional diagnosis 5
sev_ad6_fu	Severity rating at 1 month follow up - additional diagnosis 6

For these variables, data can be missing because (a) the patient did not have additional diagnoses at pre-treatment, or (b) because they never attended the follow up assessments. **You should NOT impute missing values for clients who are missing data because they did not have additional diagnoses at pre-treatment**.

Composite Scores

The composite score names are in the format: `measure_(subscale)_timecode`.

Measure Codes

Abbreviation	Name of Measure
bfnes	Brief Fear of Negative Evaluation – Straightforwardly Worded
caq_tot	Cognitive Avoidance Questionnaire – Total Score
caq_supp	Cognitive Avoidance Questionnaire – Thought Suppression
caq_subs	Cognitive Avoidance Questionnaire – Thought Substitution
caq_dist	Cognitive Avoidance Questionnaire – Distraction
caq_avoid	Cognitive Avoidance Questionnaire – Avoidance of Threatening Stimuli
caq_trans	Cognitive Avoidance Questionnaire – Transformation of Images into Thoughts
euroqol	EuroQol ED-5D-5L Index Value
fpe	Fear of Positive Evaluation
gcs	Gross Cohensiveness Scale
hrs_client	Homework Rating Scale 2 – Client Factors
hrs_therapist	Homework Rating Scale 2 – Therapist Competence
hrs_task	Homework Rating Scale 2 – Task Characteristics
nsps_tot	Negative Self Portrayal Scale – Total
nsps_social	Negative Self Portrayal Scale – Concerns About Social Competence
nsps_anx	Negative Self Portrayal Scale – Concerns About Signs of Anxiety
nsps_phys	Negative Self Portrayal Scale – Concerns about Physical Appearance
promis_anx	Patient-Reported Outcomes Measurement Information System – Anxiety (Short Form 8A)
promis_dep	Patient-Reported Outcomes Measurement Information System – Depression (Short Form 8A)
rtq	Repetitive Thinking Questionnaire (10 item trait version)
safe_tot	Subtle Avoidance Frequency Examination – Total
safe_inh	Subtle Avoidance Frequency Examination – Inhibiting & Restricting Behaviours
safe_act	Subtle Avoidance Frequency Examination – Active Behaviours (to present well)
safe_phys	Subtle Avoidance Frequency Examination – Physical Symptoms
sbsa_tot	Self-Beliefs Releated to Social Anxiety Scale – Total
sbsa_hs	Self-Beliefs Releated to Social Anxiety Scale – High Standard Beliefs
sbsa_unc	Self-Beliefs Releated to Social Anxiety Scale – Unconditional Beliefs
sbsa_con	Self-Beliefs Releated to Social Anxiety Scale – Physical Symptoms
scon_priv	Self-Consciousness Scale – High Standard Beliefs

(continued)

Abbreviation	Name of Measure
scon_pub	Self-Consciousness Scale – Unconditional Beliefs
scon_soc	Self-Consciousness Scale – Conditional Beliefs
sias	Social Interaction Anxiety Scale (20 item version)
sps	Social Phobia Scale
tracking_desc	Tracking Measure – Description
tracking_prob	Tracking Measure – Probability ratings
tracking_cost	Tracking Measure – Cost Ratings
vviq	Vividness of visual Imagery – Short Form
wai_tot	Working Alliance Inventory Short Revised (Hatcher & Gillespie, 2006 version) – Total
wai_goal	Working Alliance Inventory Short Revised (Hatcher & Gillespie, 2006 version) – Goal
wai_task	Working Alliance Inventory Short Revised (Hatcher & Gillespie, 2006 version) – Task
wai_bond	Working Alliance Inventory Short Revised (Hatcher & Gillespie, 2006 version) – Bond

Timecodes

Time Code	Corresponding Session
0	Baseline assessment (week 0)
1	Treatment Session 2 (1 week post-baseline)
2	Treatment Session 3 (2 weeks post-baseline)
3	Treatment Session 4 (3 weeks post-baseline)
4	Treatment Session 5 (4 weeks post-baseline)
5	Treatment Session 6 (5 weeks post-baseline)
6	Treatment Session 7 (6 weeks post-baseline)
7	Treatment Session 8 (7 weeks post-baseline)
8	Treatment Session 9 (8 weeks post-baseline)
9	Treatment Session 10 (9 weeks post-baseline)
10	Treatment Session 11 (10 weeks post-baseline)
11	Treatment Session 12 (11 weeks post-baseline)
15	Treatment Session 13 (One month follow Up session; 15 weeks post-baseline)
37	Six Month Follow Up (37 weeks post-baseline)

Examples

`sias_15` = Social Interaction Anxiety Scale score at 15 weeks post-baseline.

`caq_supp_11` = Cognitive Avoidance Questionnaire – Thought Suppression at 11 weeks post-baseline.

Meaning of Baseline (0) Timecode

Most measures were completed once before treatment started. Some measures were completed at the pre-treatment assessment ("A1") only, such as the SAFE, SCON and RTQ. Others were completed immediately before the first treatment session ("T1"), such as the BFNE, FPE, SPS, NSPS, CAQ and VVIQ. Three measures (SIAS, PROMIS Anxiety, PROMIS Depression) were completed at both A1 and T1. For the measures that were collected twice, the scores with a timecode of 0 is the average score across the two assessments. For the measures that were only collected once, the scores with a timecode of 0 are simply the scores that was observed on the one occasion (either A1 or T1) the measure was filled in.

Handling of Missing Item Data

Sometimes clients fill in most, but not all items in a scale. When this occurs, the total or subscale score are pro-rated, which is also known as *person-mean imputation*. For example, if a client filled in 19 of 20 SIAS items, then the total score is computed by averaging the 19 completed items, and multiplying by the number of items in the scale (20). This is why the scores for some scales have decimals. An exception is the scoring of the `EuroQol ED-5D-5L Index Value`. These cannot be computed when there is missing data, so if any items are missing, then the index value will also be missing.

HRS Single-Item Subscales

In addition to the HRS subscales listed in the Measure Codes table, there are also two single-item subscales ("quality" and "quantity"). These can be found in the item-level data section (discussed next). The HRS quality is item 1 (e.g., hrs_timecode_1) and the quantity question is item 2 (e.g., `hrs_timecode_2`).

Item-level Data

Time Code	Corresponding Session
a1	Pre-Treatment Assessment
t1	Treatment Session 1
t2	Treatment Session 2
t3	Treatment Session 3
t4	Treatment Session 4
t5	Treatment Session 5
t6	Treatment Session 6
t7	Treatment Session 7
t8	Treatment Session 8
t9	Treatment Session 9
t10	Treatment Session 10
t11	Treatment Session 11

(continued)

Time Code	Corresponding Session
post	Treatment Session 12 (final weekly treatment session, aka 'post-treatment')
fu	Treatment Session 13 (one month follow Up session)
fu6m	Six month follow up

Some measures contain a mix of positively and negatively-keyed items. The negatively-keyed items would ordinarily need to be reverse-scored before composite scores can be computed, or inter-rater reliability coefficients calculated. I have already done the reverse-scoring for you, therefore there is no need for you to reverse-score anything prior to undertaking analyses. (e.g., SIAS item 5 ordinarily requires reverse-scoring, however because I have already done this, a patient who answered this item with a response of 3 will have had his/her response stored as 1 in the data file).

Known Issues

In the section below I describe various issues relating to the data that you should familiarise yourself with before undertaking analyses.

Missing Data

- I have tried to minimise the amount of missing data but extracting it from patient files where possible. Where data are still missing, it is because the data were not collected or is irretrievable.
- The *n* at treatment session 4 for most measures is 82 or 84; for the BFNE and FPE it is 74. This is because the BFNE and FPE appear on the back of a sheet (with the PROMIS scales on the front) and at week 4, several patients appear to have only filled in the front page of the sheet.
- For the tracking measure, the *n* at 6-month follow up is 51, whereas for most other measures it is 63. This is because the tracking questions were not administered to the first cohort of groups (which commenced on 17 August 2016).
- The trial protocol only called for the EuroQol to be collected at A1, 1- and 6-month follow up. In between the first and second cohort of trial groups, CCI started routinely collecting the EuroQol in all of our programs at post-treatment. Therefore post-treatment data are available for all RCT patients except the first cohort. As the post-treatment data were not collected for the first cohort, the *n* (51) is smaller for the EuroQol than most other measures (where it is 69 or 70). I recommend using the post-treatment data as part of the trial analyses as it should improve the precision of the imputations and estimated treatment effect at the follow-ups, as the post-treatment scores should be quite strong predictors of missing follow up scores. Also we know for the first cohort missing post-treatment EuroQol scores, the missingness mechanism is *missing completely as random* as the reason the data is missing is because by design never administered the questionnaire to them.
- The first cohort of BSOCIAL groups were erroneously administered only 10 of 20 SPS items at the one-month follow up session.

- During the first half or so of the trial, it was standard CCI practice for non-trial group participants to only be administered 10/20 SPS items (the subset of items from the SPS-6 and SIPS). Therefore it would be a good idea in any study comparing outcomes for trial and non-trial subjects to recalculate SPS scores using only those 10 items for all patients, to ensure that any differences on the SPS are not due to differences in the version of the questionnaire administered.

Estimand

- I noted in the section on *Session Attendance* that there were 16 cases where the number of treatment sessions recorded by the clinician did not match the number coded as `Attended` by Prof. McEvoy. I reviewed the codes entered by Prof McEvoy for these cases. There was one client for whom three sessions in a row were coded as not attended. It is possible that there could have been a catch up session conducted for one of those sessions. If that were the case, some of the patient's data could have been erroneously wiped (due to use of the de jure estimand). I reviewed this patient's file and confirmed that she did miss three consecutive sessions and that there were no catch up sessions – thus no data will have been wiped erroneously.

Scoring

- Country-specific norms are used to compute `EurQol ED-5D-5L` index Values. To the best of my knowledge, there are no Australian norms. Therefore I used English norms to compute EurQol index values. This is the same approach that a recent South Australian study used. The R package eq5d was used to calculate the index values.

Potentially Problematic Cases

These are cases which, depending on the goals of the analyses, may need to be removed:

- `bsocial_id == 20`: after dropping out of the BSOCIAL trial, this patient took part in a non-trial group. Thus in any analysis comparing trial and non-trial groups, this patient needs to be filtered out of the analysis, or the non-trial group scores will need to be discarded.
- `bsocial_id == 83`: after dropping out the BSOCIAL trial, this patient was allocated to a non-trial group but never commenced treatment. Depending on the nature of the analysis comparing trial and non-trial groups, this patient may also need to be filtered out.
- `bsocial_id == 72`: this patient attended 4 individual therapy sessions for Body Dysmorphic Disorder at CCI in between the one and six month BSOCIAL follow ups.
- `bsocial_id == 56` : this patient's tracking measure probability rating may be unreliable. The client rated the probability of her fear as zero (i.e., at the floor) at the first two sessions (variables `track_prob_t1`, `track_prob_t2`), and as one at session 4 (`track_prob_t4`). I spoke to Sam (therapist for all trial groups) about this – she said the client was very concrete and during the first few treatment sessions she likely misunderstood the question she was being asked to rate, leading to her ratings likely being lower than they truly should have been. Sam said the client likely adjusted her style of responding as therapy progressed and she came to better understand what she was meant to be rating. This is an example of measurement invariance being violated.

In case it is relevant:

- the `bsocial_id` values for the two patients randomised to treatment whose scores were removed for invalid responding are 24 and 100
- `bsocial_id == 105` is the ID of the patient who was removed from the study after the one-month follow up due to high suicide risk.

Contact Details

If you have any questions please contact:

David Erceg-Hurn
Senior Research Scientist, Centre for Clinical Interventions
david.erceg-hurn@health.wa.gov.au
https://www.cci.health.wa.gov.au/