Update on Target-D: Statistical analysis plan for the stratified individually randomised controlled trial of the *diamond* clinical prediction tool to triage and target treatment for depressive symptoms in general practice.

Date and version number

22.03.2019 Version 1

Trial registration and ethics approval

Retrospectively registered with the Australian New Zealand Clinical Trials Registry (ANZCTR 12616000537459) on 27 April 2016. Ethics approval provided by the Human Research Ethics Committee at the University of Melbourne (ID number 1543648). The Australian Government Department of Human Services Information Services Branch has approved the collection of MBS and PBS data (ID: MI3794).

Funding acknowledgement

This study is funded by a grant from the National Health and Medical Research Council (NHMRC) (ID: 1059863). The funding source had no role in the design of this study and will not have any role during its execution, analyses, interpretation of the data, or decision to submit results.

Protocol reference

Gunn J, Wachtler C, Fletcher S, Davidson S, Mihalopoulos C, Palmer V, Hegarty K, Coe A, Murray E, Dowrick C, Andrews G, Chondros P. Target-D: a stratified individually randomized controlled trial of the diamond clinical prediction tool to triage and target treatment for depressive symptoms in general practice: study protocol for a randomized controlled trial. Trials 2017; 18: 342. DOI 10.1186/s13063-017-2089-y

Roles and responsibilities – non-signatory names and contribution

	Role	Affiliations
Dr Patty Chondros	Lead statistician	Department of General Practice, University of Melbourne, Melbourne, VIC, Australia
Dr Susan Fletcher	Project manager/Trial co-ordinator	Department of General Practice, University of Melbourne, Melbourne, VIC, Australia
Prof Cathrine Mihalopoulos	Lead health economist	Deakin Health Economics, Institute for Health Transformation, Deakin University, Melbourne, VIC, Australia
Konstancja Densley	Data manager and analyst	Department of General Practice, University of Melbourne, Melbourne, VIC, Australia
Dr Lidia Engel	Health economist (health economic methods)	Deakin Health Economics, Institute for Health Transformation, Deakin University, Melbourne, VIC, Australia
Yong Yi Lee	Health economist (health economic analysis)	Deakin Health Economics, Centre for Population Health Research (CPHR), Deakin University, Melbourne, VIC, Australia School of Public Health, The University of Queensland, Brisbane, QLD, Australia (Honorary)
Prof Jane Gunn	Chief investigator/Clinical lead	Department of General Practice, University of Melbourne, Melbourne, VIC, Australia

Contact person: Patty Chondros
Department of General Practice
The University of Melbourne
e-mail: p.chondros@unimelb.edu.au

Dr Patty Chondros

Dr Susan Fletcher

SAP revision history

Version 1 of the statistical analysis plan was finalised and approved on the **22nd March 2019**.

Protocol version	Updated SAP version no.	Section number changed	Description and reason for change	Date changed

Table of contents

Glossary and abbreviations

AQoL-8D	Assessment of Quality of Life scale (8-dimension version)
CPT	Clinical Prediction Tool
CSV	Comma Separated Value
CACE	Complier average causal effect
DSM	Diagnostic and Statistical Manual
GAD-7	Generalised Anxiety Disorder scale (7-item version)
GLM	Generalised linear models
GP	General practitioner
iCBT	Internet-based cognitive behavioural therapy
ICERs	Incremental cost-effectiveness ratios
ITT	Intention to treat
MAR	Missing at random
MHSES	Mental Health Self-Efficacy Scale
MBS	Medicare Benefit Schedule
PBS	Pharmaceutical Benefit Scheme
PHQ-2 / PHQ-9	Patient Health Questionnaire (2-item / 9-item version)
QALY	Quality Adjusted Life Year
RCT	Randomised controlled trial
RUQ	Resource Use Questionnaire
SD	Standard deviation
UC+	Usual care plus Target-D attention control

1 Background

Target-D study is a stratified individually randomised controlled trial (RCT) with two parallel arms that aims to test whether using the *diamond* clinical prediction tool (CPT) to tailor treatment recommendations to an individual's predicted depressive symptom severity is a clinically effective and economically efficient way of reducing depressive symptoms, relative to usual care plus Target-D attention control (UC+). Full details of the trial design, including setting, recruitment, eligibility criteria, the intervention, randomisation, sample size calculations and statistical analysis are detailed in the published study trial protocol (1). **This paper provides a more detailed statistical analysis plan, which includes the health economic evaluation, to complement the study protocol and expands on the secondary and sensitivity analyses (2).**

1.1 Primary hypothesis

The null hypothesis is that there is no difference in mean depressive symptom severity at 3 months between the intervention (symptom feedback, reflection, tailored treatment recommendation) and comparison (UC+) arms.

1.2 Study objectives

The **primary objective** of the Target-D trial is to determine whether the mean depressive symptom severity at 3 months of using the *diamond* CPT to triage individuals with depressive symptoms into symptom severity-appropriate treatment (intervention arm), compared to usual care plus Target-D attention control (comparison arm).

Secondary objectives are to:

a) test whether individuals in the intervention and comparison arms differ in mean depressive symptoms at 12 months, and quality of life, anxiety symptoms, self-efficacy, and health service use at both three and 12 months;

b) determine whether the outcomes differ between the two study arms within each of the three depressive symptom severity groups; and

c) evaluate the cost-effectiveness of the new model of care compared to UC+.

1.3 Trial design

In brief, participants aged 18 to 65 years old are recruited in the general practice waiting room from a minimum of ten general practices in Victoria, Australia. Patients attending general practices are assessed for eligibility using a self-report survey delivered via an iPad. Eligible and consenting participants complete the *diamond* CPT (3) online and are randomised to the intervention or comparison arm with 1:1 allocation, stratified by general practice and predicted depression group. Participants in the intervention arm are categorised into one of three treatment groups according to their *diamond* CPT results. They receive feedback on CPT responses, an opportunity for reflection, and a treatment recommendation specific to their predicted depressive symptom severity (self-help, guided self-help, or nurse-delivered collaborative care for minimal/mild, moderate, and severe symptoms respectively). Participants in the comparison arm do not receive feedback, reflection, or tailored treatment recommendations. Instead, all participants in this arm receive usual GP care plus Target-D attention control in the form of a telephone interview in which they are asked to reflect on the value of mental health research and the pros and cons of their GP being involved. Participants in this arm are also advised that they will be asked for feedback on how they manage their emotional health and wellbeing.

Participant recruitment began on 4 April 2016 and was completed on 21 December 2017. End of data collection for 12-month outcomes occurred on 15 February 2019.

1.4 Outcomes

Outcomes are measured at baseline prior to randomisation, at three- and 12-months post randomisation.

The **primary outcome** is depressive symptom severity at three months post-randomisation, assessed using the Primary Health Questionnaire (9-item version) (PHQ-9: 4). Depressive symptom severity is also measured at 12 months as a **secondary** outcome. The PHQ-9 assesses the nine Diagnostic and Statistical Manual (DSM) symptoms of depression over the last two weeks using a four-point Likert scale (where 0 = 'not at all' and 3 = 'nearly every day'). Total scores will be calculated by summing the nine items, with scores ranging between zero and 27, where low scores indicate minimal/mild symptoms for depression and higher scores indicate increasing depressive symptom severity. If two or fewer items on the PHQ-9 are missing responses, the missing values will be substituted with the mean response of the completed items, otherwise the total score will be coded as missing (5). The PHQ-9 is a validated

diagnostic measure in primary care (6), with demonstrated efficacy and sensitivity as an outcome measure for treatment trials with a recommended Reliable Change Index (7).

Secondary outcomes include mental health self-efficacy, anxiety symptom severity, and quality of life at three and 12 months.

Self-efficacy is measured using the Mental Health Self-Efficacy Scale (MHSES: 8). The MHSES comprises of six items that asks the respondent to rate on a 10 point-Likert scale (ranging from 1 – 'not at all confident' to 10 – 'totally confident') on how confident on an average day in the next month they will be able to perform behaviours related to mental health self-care. The six items are summed to create a total score, ranging from six to 60, where higher scores indicate higher self-efficacy in mental health self-care. If one or two items of the MHSES are missing values, the missing values will be substituted with the mean score of the non-missing items, otherwise the total MHSES score will be coded as missing. The MHSES displays high internal consistency (Cronbach's alpha = .91) and good construct validity, correlating well with measures of depression, anxiety, and functional impairment.

Anxiety symptom severity is measured using the 7-item Generalized Anxiety Disorder scale (GAD-7: 9). The GAD-7 assesses the presence of anxiety symptoms over the past two weeks using a four-point Likert scale (0=Not at all, 1=Several days, 2=More than half the days, 3=Nearly every day). The seven items will be summed to create a total score, ranging between zero and 21, where the higher scores indicate more severe anxiety symptoms. If one or two items on the GAD-7 are missing a response, the missing values of these items will be substituted with the mean response of the completed items, otherwise the total score will be coded as missing. The GAD-7 has excellent internal consistency (Cronbach's alpha = .92) and test-retest reliability. Its construct, convergent, and discriminant validity are high; it correlates well with measures of depression and functioning (while assessing a distinct construct), as well as with other measures of anxiety.

Quality of life is assessed at each time point using the Assessment of Quality of Life (AQoL-8D: 10). This is a validated, reliable measure that comprises 35 items across eight dimensions (independent living, senses, pain, mental health, happiness, self-worth, coping, and relationships) that can be used to calculate Quality Adjusted Life Years (QALYs) via a utility algorithm. The AQoL-8D utility scores will be calculated according to the published instructions (11), and then scaled such that the worst health state is 0.087 on a 0-1 scale,

where 1.0 is full health and zero is equivalent to being dead (12). The AQoL-8D has been shown to be sensitive to depressive symptom severity levels (13, 14).

Cost effectiveness of the intervention will be measured through assessment of health service use, effects on productivity, and calculation of QALYs. Health service use will be tracked using data routinely collected by the Australian Government Department of Health; specifically, the Medicare Benefits Schedule (MBS: 15) that maintains information about visits to health care providers and diagnostic tests; and the Pharmaceutical Benefits Scheme (PBS: 16) database of medications supplied on prescription. Participants are asked to provide additional, separate consent to access their MBS and PBS data. Other resource use not captured by these national databases, including the use of broader health and welfare services and effects on productivity (i.e., education and workforce participation), will be assessed via self-report using an adapted questionnaire developed by members of the research team and used in numerous other Australian mental health intervention trials (e.g., 17, 18). Since not all participants may provide consent to MBS/PBS data extraction, this resource use questionnaire (RUQ) also covers items captured by these administrative data sets (e.g. general practitioner [GP] consultations).

1.5 Screening and baseline data collection

General practice and GP characteristics are collected at the time of practice recruitment, prior the commencement of patient recruitment in each practice. Participant characteristics and baseline outcome measures are collected during screening and in the baseline survey prior to randomly allocating eligible individuals to the study arms. Data collected include (but are not limited to):

1.5.1 General practice characteristics

- Postcode
- Estimated number of patients seen per day
- Estimated number of patients aged between 18 and 65 years old
- Number of full-time equivalent staff (GPs, nurses, psychologists, allied health)

1.5.2 General practitioners' demographic characteristics

- Age in years
- Gender (male, female)
- Country of graduation
- Types of qualifications
- Years in general practice (in Australia and overseas)
- Bulk billing practice
- Usual approach to depression care
- In what proportion of your adult patients in the past 12 months would you estimate depression to be a significant part of the clinical picture?
- What percentage of your consultations are conducted:
    - In English
    - In a language other than English with the use of an interpreter (trained or untrained)
    - In a language other than English that you speak

1.5.3 Demographic characteristics collected at screening for individuals who completed the eligibility screening survey in their GP waiting room[1]

Item description	Questions in screening survey	Responses
Age in years*	How old are you (in years)?	Range between 18 and 65
Gender	Do you identify as:	Male Female Other
Highest level of education completed	What is the highest level of education you have completed?	Left school before completing Year 10 Year 10 or equivalent Year 11 or equivalent Year 12 or equivalent Certificate/Diploma Bachelor degree or higher
Employment status*	In a usual week, which of the following best describes you?	Employed/working for profit or pay Sheltered employment Unemployed and looking for work Not in paid employment or not looking for work
Current attendance at school or any other education institute	Are you attending school or any other education institution (including distance education?	Yes – Full-time student Yes – Part-time student No
Usual living situation (who with)	Who do you usually live with? (Please select all that apply)	I live alone Husband or wife Defacto partner My child/ren My partner's child/ren My parent/s Unrelated flatmate or co-tenant Other
Postcode*	What is your postcode?	Range between 3000 and 3999

[1] Variables with asterisks were collected to exclude patients who did not meet the trial eligibility criteria.

Item description	Questions in screening survey	Responses
Self-rated health	In general, would you say your health is…	Excellent Very good Good Fair Poor
Holds a health care card	Do you hold any of the following health care cards in addition to your Medicare Card? (Please select all that apply)	Health Care Card (Centrelink) Pensioner Concession Card (Centrelink) Commonwealth Seniors Health Card (Centrelink) Department of Veterans' Affairs Card None
PHQ-2 score*	Over the last 2 weeks, how often have you been bothered by little interest or pleasure in doing things? Over the last 2 weeks, how often have you been bothered by feeling down, depressed or hopeless?	For each item the responses are: Not at all Several days More than half the days Nearly every day
Current use of any Internet-based cognitive behavioural therapy (iCBT) programs*	Are you currently using any of the following programs:	Upside This Way Up Direct Blue MyCompass None
Number of times visited a psychologist or counsellor in the past 12 months	In the last 12 months, how many times have you visited a psychologist or counsellor to talk about your emotional wellbeing?	0 times 1-6 times 7-12 times 13 times or more
Appointment to see a psychologist or counsellor for your emotional health in the next three months*	Do you have an appointment to see a psychologist or counsellor for your emotional health in the next three months?	Yes No
Current use of antidepressants	Are you currently taking any antidepressants (eg., Zoloft, Cipramil, Prozac)?	Yes No

Item description	Questions in screening survey	Responses
Length of time taking current antidepressant*	How long have you been taking your current antidepressant for?	Less than one month 1 month to less than 3 months 3 months to less than 6 months 6 months to less than 1 year 1 year to less than 2 years 2 years or more Don't know
Current use of antipsychotics*	Are you currently taking any antipsychotics (eg. Seroquel, Risperdal, Zyprexa, etc)?	Yes No
Frequency of internet use	How often do you use the internet?	Daily Weekly Fortnightly Monthly Less often
Regular access to computer with internet access*	Do you have regular access to a computer with internet access?	No Yes

1.5.4 Demographic characteristics collected at baseline, for eligible individuals that consented to participate in the trial

Item description	Questions in baseline survey	Responses
GP seen on day recruited to Target-D	Which doctor did you see on the day we first spoke to you?	
Usual GP	Is this your 'usual' doctor? If no - Who is your 'usual' doctor?	Yes No
Holds a health care card	Do you hold any of the following health care cards?	None Health Care Card Pensioner Concession Card Commonwealth Seniors Health Card Department of Veterans' Affairs
Receiving any kind of benefit or disability support	Are you receiving any kind of benefit or disability support? If yes - What is the name of the benefit you are receiving?	Yes No
Current employment status	Are you currently In a usual week, which of the following best describes you?	Employed/working for profit or pay Unemployed Neither working nor looking for work Unable to work due to sickness or disability Looking after ill or disabled person Home duties/child care Retired/voluntarily inactive Studying Other, please specify
Current voluntary work	Do you currently work in an unpaid voluntary job?	Yes No
Attending school or any other education institution	Are you attending school or any other education institution (including distance education)?	No Yes, full time student Yes, part-time student

1.5.5 Health resource use collected at baseline, for eligible individuals that consented to participate in the trial

Item description	Questions in baseline survey	Responses
Number of visits to health professionals for mental health	In the last month, how many times have you seen the following health professionals for your mental health? GP Hospital outpatient doctor Specialist doctor Physiotherapist Chiropractor Psychologist Counsellor Psychiatrist Nurse Social Worker Domestic violence worker Alcohol and drug worker Family therapist Complementary/alternative therapist Support group Pharmacist Other natural therapist	For each item the responses are: 0 times 1-2 times 3-4 times 5-6 times 7-11 times 12 or more times
Use of self-help strategies and community services	In the last month, how many times have you used the following for your mental health? Internet sites Read a self-help book Watched a self-help DVD Telephone helpline Community rehabilitation Care worker	For each item the responses are: 0 times 1-2 times 3-4 times 5-6 times 7-11 times 12 or more times

Item description	Questions in baseline survey	Responses
Medications for mental health	We'd like to know about any medication you may be taking for your mental health *(Leave section blank in not taking anything, allows for up to 5 medicines)*	
	Name of medicine (example: Citalobell; Valpam; Seroquel; Rivotril; Pain relief; St John's Wart; Valerian; Vitamins / Minerals)	Selection provided from a pull-down list
	What is the dosage per day in milligrams (mg)?	Free text
	How long have you been taking the medicine?	Free text
Emergency or casualty visits in last month	In the last month have you had any visits to any emergency or casualty services without staying overnight? (Allows up to 5 visits) *If "Yes", next section asks following details for up to 5 visits*	Yes No
	Broad reason for attendance (e.g. felt dizzy)	Free text
	Type of Hospital	Public Private
	On average how much did you pay out of pocket each time you used this service?	$0-$9 $10-$19 $20-$29 $30-$39 $40-$49 $50-$59 $60-$69 $70-$79 $80-$89 $90-$99 $100+

Item description	Questions in baseline survey	Responses
Overnight hospital visits in last month	In the last month, have you had any overnight visits in a hospital, or had treatments in a hospital day surgery/care facility? *If "Yes", next section asks following details for up to 5 visits*	Yes No
	Reason for admission	Free text
	Type of hospital	Public Private Other
	Number of nights in the hospital	Range between 1 and 10+
	Method of transport to hospital	Ambulance Bus Bicycle Car Taxi Train Tram Walk
	Source of payment	Medicare Private Out of pocket
Private health insurance	Are you currently covered by private health insurance?	Yes No
Hospital cover	Do you have hospital cover? *Skipped if no private health insurance*	Yes No
Extras cover	Do you have extras cover? *Skipped if no private health insurance*	Yes No

1.5.6 Items in the *diamond* CPT (3)

Item description	Questions in Target-D website	Responses
Current depressive symptom severity (sum of PHQ-9 items; score ranges from 0 - 27)	As per PHQ-9	For each item the responses are: Not at all Several days More than half the days Nearly every day
History of depression	Have you ever been bothered by feeling down, depressed or hopeless for longer than 2 weeks? Have you ever been bothered by little interest or pleasure in doing things for longer than 2 weeks?	Combined responses of the two items to create a new binary variable: 1 if responded yes to both items and 0 (no) otherwise
Current anxiety (19)	Over the last 4 weeks, how often have you been bothered by feeling nervous, anxious, on edge or worrying a lot about different things?	Not at all Several days More than half these days
Long term illness	Do you have any long-term illness, health problem, which limits your daily activities or the work you can do (including problems that are due to old age)?	No Yes
Self-rated health	In general, would you say your health is...	Excellent Very Good Good Fair Poor
Live alone	Do you live alone?	No Yes
Managing on available income	How do you manage on your available income?	Easily Not too bad Difficult some of the time Difficult all of the time Impossible
Gender	Are you male or female?	Male Female

2 Data management and workflow

All participant responses at screening, baseline, and three and 12 months follow up are collected within the purpose-built Target-D website, which was developed using JavaScript. Data integrity is enforced using forced or multiple-choice items wherever possible; valid value and range checks are also built into the website for free text fields where appropriate.

Participants receive an automated email with a link to their follow-up survey two weeks before the due date. The survey due date is calculated as three and 12 months respectively from the date the participants were randomised to their study arm. They receive an automated email reminder two days later and are contacted by phone if they have not completed the survey within a week.

To minimise non-response for the primary outcome, participants who do not complete the survey after five reminders (via phone, text, or email) are sent a second link to a shorter version of the survey created in Qualtrics (20) that consists of the PHQ-9 (primary outcome) and GAD-7 only. Participants are also offered the option of being sent a hard copy of the survey with a reply-paid envelope to return via post. Returned hard copy surveys are entered into the Target-D study website by a research assistant blinded to the participant's study arm status.

The Research Electronic Data Capture (REDCap) secure software application (21) is used to manage contact with participants and track their progress through the study. Trained research assistants manually transfer participant information each day from the Target-D website into REDCap. Both REDCap (22) and Qualtrics (20) databases are password-protected and housed on secure servers. The Target-D website is hosted on the NeCTAR Research Cloud, a collaborative Australian research platform supported by the National Collaborative Research Infrastructure Strategy. Only researchers named on the ethics approval have access to the identified data.

The data manager will be responsible for exporting the data from the Target-D website and Qualtrics. The data will be exported as Comma Separated Value (CSV) files weekly, stored securely and backed up regularly on a central password-protected University system. These CSV data files will be imported to Stata Statistical Software (v15.1: 23) for data processing. The data manager blinded to study arm status will check all the data to identify and where possible resolve errors prior to statistical analyses being conducted. Variables will be coded, labelled and scales scored according to each instrument's guidelines. Datasets will be merged as required for analysis with the unique record identifier.

Final data transfer from the online data collection systems will occur after the end of data collection for 12-month outcomes and all the data will be exported as CSV files and imported to Stata Statistical Software (v15.1: 23) for statistical and economic analysis. The online data collection systems will be locked after the end of data collection for 12-month outcomes and the final statistical analysis plan has been approved. De-identified data will be stored on the University server for future use in accordance with the National Statement on Ethical Conduct in Human Research (24)[2].

3 Harms

Given that interventions delivered in the Target-D trial were selected based on randomised controlled trial evidence of effectiveness, the potential for harm is expected to be low. Reports of suicidal ideation (as indicated by a response of "nearly every day" to the question "Over the last 2 weeks, how often have you been bothered by thoughts that you would be better off dead or of hurting yourself in some way?") regardless of study arm allocation are considered adverse events but are unlikely to result in protocol deviations (i.e., treatment discontinuation or modification). All adverse events will be summarised using counts and frequencies by study arm and depressive symptom severity group.

[2] Note that this data retention plan differs from the published protocol due to the release of new guidance by the National Health and Medical Research Council in 2018.

4 Statistical methods

Statistical analyses for the three- and 12-month outcome data will be conducted together, at the end of the 12-month data collection period. It will begin after the statistical analysis plan (Version 1) is approved, and after the online data collection systems are locked. Study investigators, together with the lead statistician and health economists who will conduct and interpret the statistical and economic analysis, will remain blinded to the study arms status of participants until the intention to treat analysis is completed and interpreted. The study arm status will be coded with the letters A and B, and the key kept by the trial co-ordinator. No interim analysis is planned. All analysis will be conducted using Stata Statistical Software (23). If required, we may also use R statistical package (version 3.5.3) (25).

4.1 Descriptive analyses

Data collected at screening will be used to describe how many patients were ineligible and reasons why, noting that not all patients completed the screening items. Patients are exited from the survey when identified as ineligible for the Target-D trial in the following order: age <18 or >65 years; in sheltered employment; does not reside in Victoria; sum of the first two PHQ-9 items is <2 (i.e., participants respond as both not at all, or one not at all and one several days to the two items); currently using an online mental health program; visited a psychologist and or counsellor for emotional wellbeing at least seven times in the last 12 months and has an appointment to see psychologist and or counsellor in the next three months; taking current antidepressant for less than one month; taking any antipsychotics; does not have computer access.

A flow chart will be created to show the participant flow for Target-D (Figure 1). The flow chart will show the recruitment rate (including the number of participants screened, met eligibility criteria, consented), the number randomised to the two study arms, attrition rates, and the number of participants that contributed outcome data at each measurement time point by each study arm. When provided, the reasons participants withdrew or lost to follow up will be reported by study arm and time point. Frequency of the missing data patterns at each of the follow-up times (baseline, three and 12 months) will be described.

Participant characteristics collected at baseline will be summarised using means and standard deviations (SD) for continuous variables or counts and percentages for categorical variables between the study arms, for the entire sample and stratified according to depressive symptom severity group as shown in Table 1. For continuous data with a skewed distribution, medians and quartiles will be used instead. Note, some questions, such as employment status, self-rated health, possession of a health care card, living alone, and student status were asked more than once either when screened for eligibility, within the *diamond* CPT or in the baseline survey. For these questions, responses to the *diamond* CPT will take precedence, followed by the responses to the screening items and then responses to the baseline items. The health resource use questions asked at baseline will be used be used cost-effectiveness and cost-utility analysis (Section 4.5 below).

4.2 Intention to treat analysis

Main analyses for primary (depressive symptom severity at three months) and secondary outcomes will use an intention to treat (ITT) approach, where participants will be analysed in the study arm to which they were randomly allocated (26). Linear mixed-effects model using restricted maximum likelihood with random intercepts for individuals will be used to estimate the difference in mean outcome between study arms at three and 12 months for the primary and secondary outcomes. The variance-covariance matrix will be unstructured and will be allowed to differ between the two study arms. All regression models will adjust for baseline outcome measure, stratification factors (general practice, depressive symptom severity group) and time (baseline, three and 12 months), with a two-way interaction between study arm and time, except baseline where means in the study arms will be constrained to be equal (27). No other baseline variables will be considered for adjustment in the analysis (28).

4.2.1 Sub-group analysis

For primary and secondary outcomes, a sub-group analysis using the same regression model described for the main ITT analyses will be conducted separately for each depressive symptom severity groups. The sample size was based on detecting clinically important differences within these sub-groups. No corrections will be made for multiple testing.

4.2.2 Estimated intervention effects

Estimated intervention effects will be reported as the difference in the means of the outcome between study arms (intervention-comparison), with 95% confidence intervals and p-values. These results will be presented for the primary and secondary outcomes for all participants, and by each depressive symptom severity groups as shown in Table 2. The standardised effect size will also be presented for the primary and secondary outcomes.

For the primary outcome, the estimated means for each study arm with respective 95% confidence intervals will be plotted (*y*-axis) against the follow-up time (baseline, 3 and 12 months; *x*-axis), for all participants and by each depressive symptom severity group.

4.3 Sensitivity analyses

Three sensitivity analysis will be conducted for the main and sub-group ITT analyses of the primary and secondary outcomes:

4.3.1 Sensitivity analysis for clustering effect by nurse

Participants with severe depressive symptoms in the intervention arm are assigned to one of five nurses who deliver the collaborative care intervention. There is the potential for nurses to differ in their effectiveness in delivering this intervention, thus the outcomes in participants seen by the same nurse may be correlated (29). A sensitivity analysis will include nurse as a random effect in the linear mixed-effects model described for the main and sub-group ITT analyses to allow for possible clustering effect by nurse (29, 30). If appropriate, estimates of the intra-cluster correlation coefficient, used to quantify the degree to which outcomes of participants who see the same nurse are correlated, will also be reported (30).

4.3.2 Sensitivity analysis adjusting for baseline variables associated with non-response

A sensitivity analysis will include additional variables measured at baseline associated with non-response at 3- and 12-months follow-up as fixed effects to the mixed-effects model used in the main ITT analysis. These variables will be identified in an ancillary analysis where the baseline participant characteristics will be compared between responders (trial participants with outcome data at both follow-up periods) and non-responders using descriptive statistics by each study arm. Logistic regression will be used to investigate the association between baseline variables (independent variable) and non-response (dependent variable) (31). Adding variables associated with non-response in the model may make the MAR assumption more plausible (32).

4.3.3 Sensitivity analysis to assess robustness of missing data assumption using pattern-mixture model

Under the mixed-effects model used for the main and sub-group ITT analyses, data are assumed to be missing at random (MAR), conditional on the covariates included in the model (26). A pattern–mixture model will be used for the primary outcome, depressive symptom severity at three months, to assess whether estimates were robust to departures from the missing data assumption for all participants. Analysis for departures from MAR for the mixed effects model will be assessed by adding the quantity $\Delta = p_1\delta_1 - p_0\delta_0$ to the estimated treatment effect for depression severity score at three months in the main analysis, where p_i is the proportion of missing data at three months and δ_i the difference in mean depressive symptom score between the individuals with missing and those observed responses in the intervention ($i = 1$) and comparison ($i = 0$) arms (33).

A range of values for δ_i will be considered for the difference in mean depressive symptom scores between the participants with missing data and those observed at three months. Given that higher depressive symptom scores indicate poorer outcome, negative values of δ_i assume that individuals with missing data have lower (better) depressive symptom scores on average than observed individuals and positive values of δ_i assume that individuals with missing data have higher (worse) mean depressive symptom scores than the observed score mean.

The main analysis under MAR assumes that individuals with missing data have the same mean depressive symptom scores as those observed, that is $\delta_i = 0$ in both study arms. For the sensitivity analyses, the difference between missing and observed depressive symptom scores will be varied over the specified range of values for δ_i in the same way in both arms (that is, $\Delta = (p_1 - p_0)\delta$), vary in the intervention arm only and fixed at zero for the comparison arm ($\Delta = p_1\delta$) , and vary in the comparison arm and fixed at zero for the intervention arm ($\Delta = -p_0\delta$).

The estimated intervention effect adjusted for baseline measure of depressive symptom score with respective 95% confidence intervals will be plotted on the *y*-axis in both study arms, for selected parameter values of the difference between missing and observed mean score for depressive symptom score (δ) at three months on the *x*-axis. A horizontal reference line will be plotted at zero on the *y*-axis, where positive values of the estimated intervention effect will indicate that the mean depressive symptom score in the comparison arm is lower (better) than the intervention arm and negative values indicate that the intervention arm have lower (better) mean depressive symptoms than the comparison arm.

The analysis to assess robustness of missing data assumption may be repeated, as appropriate, for the secondary outcomes and by depressive symptom severity group.

4.4 Adherence-adjusted analysis

In a secondary analysis, complier average causal effect (CACE) analysis will be used for primary and secondary outcomes to investigate the intervention effect on individuals who adhere to their assigned intervention (34). This analysis retains and recognises the randomisation whilst adjusting by whether the intended treatment was received (35). The CACE analyses will be performed for all the participant and for each depressive symptom severity group.

Adherence to the treatment will be defined separately for each depressive symptom severity group. The interventions are flexible by design and no protocol deviations are anticipated. A participant will be considered a **complier** (binary variable) if:

- **Minimal/mild:** Participant completed at least one module of the myCompass program, as indicated by website analytics provided by the Black Dog Institute (who manage the myCompass program) (36, 37).

- **Moderate:** Participant completed the Worry and Sadness course in the This Way Up program (6 lessons in total), as tracked using website analytics provided by the This Way Up team at the University of Sydney (38).
- **Severe/complex:** Participant completed eight appointments of collaborative care, as indicated by appointment logs completed by the nurses delivering the intervention.

Quantitative measures of treatment adherence will also be examined, as follows:

- Count of number of modules/lessons/appointments completed
- Count of number of modules/lessons/appointments started
- Treatment match
    - **Minimal/mild:** modules started match priorities set by participant at time of CPT completion (using module-priority matching matrix developed prior to participant recruitment)
    - **Moderate**: priority was mood, anxiety, or loss of interest in usual activities
    - **Severe/Complex:** fidelity to collaborative care model, including individual delivering the intervention and score calculated using checklist currently under development.

Study investigators and researchers will code the variables for **treatment adherence** blinded to the study results**.**

Structural mean models will be used to estimate the CACE for treatment for the binary (compliers vs non-compliers) and quantitative variables for treatment adherence (32). The CACE analysis will be performed using a two stage-least squares instrumental regression (`ivregress` command in Stata Statistical Software (23)), with study arm used as the instrumental variable for adherence to treatment. Like the main analyses, the models will include baseline outcome measure, stratification factors (general practice, depressive symptom severity group) and time (baseline, three and 12 months) as covariates. We will also conduct a sensitivity analysis that does not adjust for any of the covariates. The CACE estimates will be reported with 95% confidence intervals and p-values.

4.5 Cost-effectiveness and cost-utility analysis

The cost-effectiveness of Target-D will be assessed alongside the trial from a health care perspective and a partial societal perspective using primarily a cost-utility framework whereby the main outcomes are QALYs. The health sector perspective includes all cost borne by the government as well as out of pocket cost incurred by patients for the direct costs of medical care. The partial societal perspective will also take productivity impacts into account. The economic evaluation will consider the cost to deliver the intervention and the cost of health care and related resources utilised by study participants during the trial. Resources used will be obtained from the financial records of the study, MBS/PBS data and self-report RUQ data, followed by a valuation, where unit costs will be taken from Australian-specific sources. The unit costs that will be applied to the resource use will be derived from a variety of sources including the MBS (medical, allied health, pathology and diagnostics), the PBS (medicines), the Independent Hospital Pricing Authority (hospitalisation and outpatient care) (39), mental health services in Australia (other types of medical care) (40) and the manual of resource use items and their associated costs (41).

Incremental cost-effectiveness ratios (ICERs) will be determined (cost of intervention – costs of comparison / outcome of intervention – outcome of comparison) using the AQoL-8D to determine QALYs. We will use the AQoL-8D utility algorithm to determine the utility score for each person at each follow up point and then the area under the curve method to determine the total QALYs for each person (42). The cost-utility analysis will be complemented with a cost-consequences analysis, whereby the differences in costs between the intervention and the comparison arm will be compared with differences in the full suite of study outcomes. The base case analysis will be undertaken as ITT and the ICE multiple imputation technique in Stata Statistical Software (23) will be used to account for missing values. Multivariate analysis of total health sector and societal costs will be undertaken with generalised linear models (GLM), supplemented by two-part multivariable models for the evaluation of the between group differences for individual categories of resource use (e.g., health professional visits). QALYs will be also analysed using GLM. Variation will be determined by bootstrap and regression analyses and results presented in cost-effectiveness planes and acceptability curves. Sensitivity analyses will also be used to determine the impact of important study parameters (such as unit cost price variation). Dependent on trial results, modelling may also be used to extrapolate beyond the trial time horizon and to evaluate the population-level costs and impacts of a potential national roll-out.

5 Table shells and figures

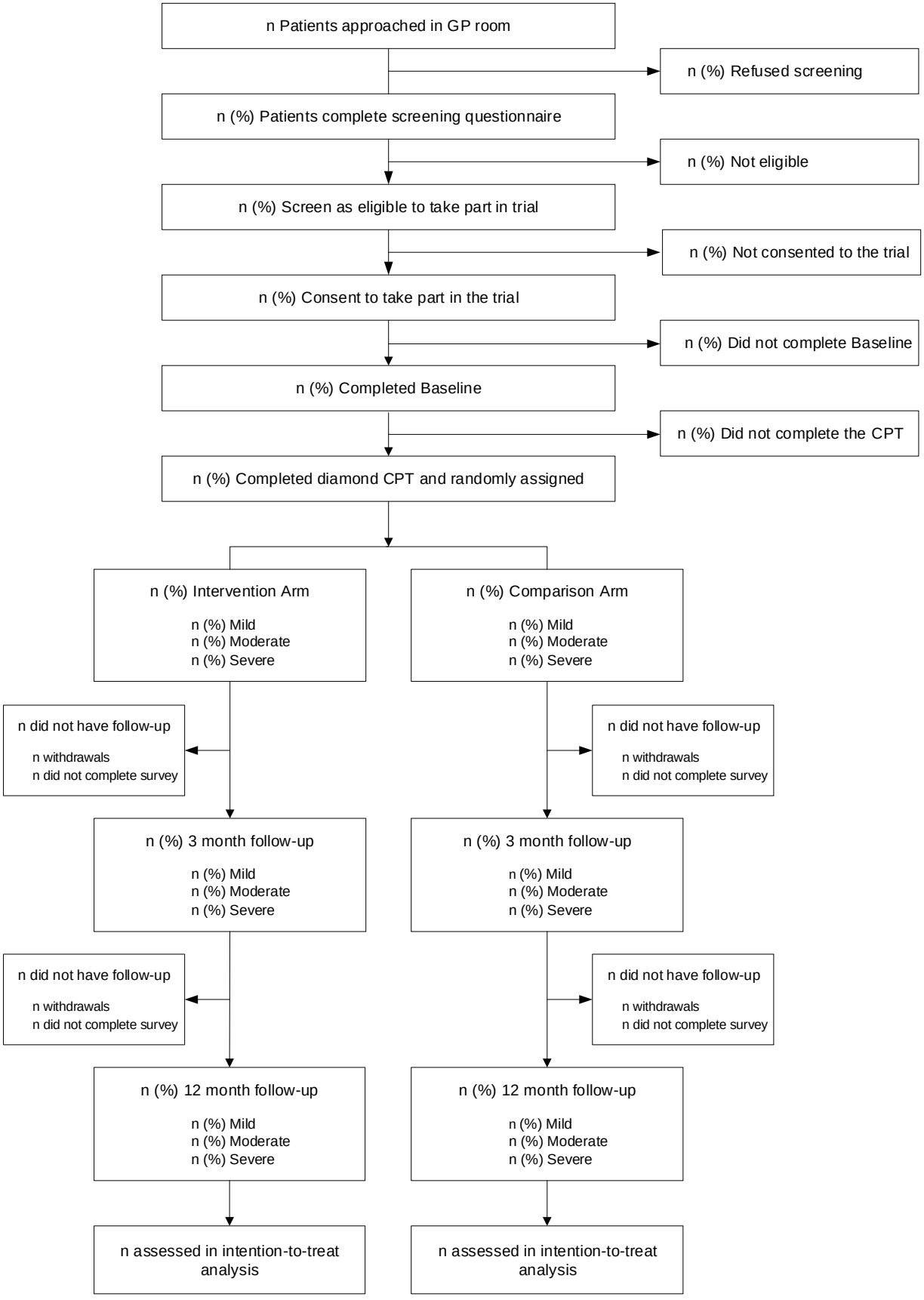

Figure 1: Trial profile

Table 1: Baseline characteristics of participant according to study arm, in total and stratified by depressive symptom severity group (N=…)

	All participants				**Mild depressive symptoms**				**Moderate depressive symptoms**				**Severe depressive symptoms**			
Study arm	Intervention (n=)		Comparison (n=)		Intervention (n=)		Comparison (n=)		Intervention (n=)		Comparison (n=)		Intervention (n=)		Comparison (n=)	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Age in years																
Depressive symptoms severity score (PHQ-9)																
Anxiety symptom severity (GAD-7)																
Mental Health Self-efficacy (MHSE)																
Quality of life (AQoL-8D)																
	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)
Gender																
Male																
Female																
Other																
Highest level of education completed																
Left school before completing Year 10																
Year 10 or equivalent																
Year 11 or equivalent																
Year 12 or equivalent																
Certificate/Diploma																
Bachelor degree or higher																
Attending school or other education institution																
Yes, full-time student																
Yes, part-time student																
No																
Are you currently																
Employed/working for profit or pay																
Unemployed																
Neither working nor looking for work																
Currently work in unpaid voluntary job																
Receiving benefit or disability support																
Holds a health care card																
Ever depressed and/or ever had little interest in doing things for greater than 2 weeks																

	All participants		Mild depressive symptoms		Moderate depressive symptoms		Severe depressive symptoms	
Study arm	Intervention (n=)	Comparison (n=)	Intervention (n=)	Comparison (n=)	Intervention (n=)	Comparison (n=)	Intervention (n=)	Comparison (n=)
PHQ current anxiety in past 4 weeks								
Not at all								
Several days								
More than half these days								
Long term illness								
Self-rated health								
Excellent								
Very good								
Good								
Fair								
Poor								
Live alone								
Managing on available income								
Easily/Not too bad/Difficult some of the time								
Difficult all of the time or impossible								
Number of times visited a psychologist/counsellor in past 12 months								
0 times								
1-6 times								
7-12 times								
13 time or more								
Current use of antidepressants								
Frequency of internet use								
Daily								
Weekly								
Fortnightly								
Monthly								
Less often								

Mean and Standard deviation (SD); Count (n) and percentage (%)
Note: Percentage of missing responses will be reported; Sub-categories may be collapsed in final table published;

Table 2: Depressive symptom severity (PHQ-9) according to study arm, in total and stratified by depressive symptom severity group

	All participants	p-value	Mild depressive symptoms	p-value	Moderate depressive symptoms	p-value	Severe depressive symptoms	p-value
Intervention arm	n		n		n		n	
Comparison arm	n		n		n		n	
Baseline estimated mean[1]	mean (SE)		mean (SE)		mean (SE)		mean (SE)	
3 months								
Mean outcome score								
Intervention arm	mean (SD)		mean (SD)		mean (SD)		mean (SD)	
Comparison arm	mean (SD)		mean (SD)		mean (SD)		mean (SD)	
Difference in mean outcome between arms (95% CI)[1]	estimate (95% CI)	p-value	estimate (95% CI)	p-value	estimate (95% CI)	p-value	estimate (95% CI)	p-value
Sensitivity analysis[2]	estimate (95% CI)	p-value	--		--		estimate (95% CI)	p-value
Sensitivity analysis[3]	estimate (95% CI)	p-value	estimate (95% CI)	p-value	estimate (95% CI)	p-value	estimate (95% CI)	p-value
CACE analysis[4]	estimate (95% CI)	p-value	estimate (95% CI)	p-value	estimate (95% CI)	p-value	estimate (95% CI)	p-value
12 months								
Mean outcome score								
Intervention arm	mean (SD)		mean (SD)		mean (SD)		mean (SD)	
Comparison arm	mean (SD)		mean (SD)		mean (SD)		mean (SD)	
Difference in mean outcome between arms (95% CI)[1]	estimate (95% CI)	p-value	estimate (95% CI)	p-value	estimate (95% CI)	p-value	estimate (95% CI)	p-value
Sensitivity analysis[2]	estimate (95% CI)	p-value	--		--		estimate (95% CI)	p-value
Sensitivity analysis[3]	estimate (95% CI)	p-value	estimate (95% CI)	p-value	estimate (95% CI)	p-value	estimate (95% CI)	p-value
CACE analysis[4]	estimate (95% CI)	p-value	estimate (95% CI)	p-value	estimate (95% CI)	p-value	estimate (95% CI)	p-value

SE – Standard error; SD - Standard deviation; CI - Confidence Interval

1 Baseline mean and mean for intervention arm minus mean for comparison arm estimated using linear mixed-effects regression with random intercepts for individuals and adjusted for baseline outcome measure, general practice, time and depressive symtom severity group (for all partcipants only); Mean outcome is constrained to be equal at baseline.

2 Same as 1, with random effects for case manager;

3 Same as 1, adjusted for baseline variables associated with non-reponse;

4 CACE analysis

Note: Table will be repeated for secondary outcomes anxiety symtom severity (GAD-7) and Mental Health Self-Efficacy Scale (MHSES)

6 References

1. Gunn J, Wachtler C, Fletcher S, Davidson S, Mihalopoulos C, Palmer V, et al. Target-D: A stratified individually randomized controlled trial of the diamond clinical prediction tool to triage and target treatment for depressive symptoms in general practice: Study protocol for a randomized controlled trial. Trials 2017; 18. DOI: 10.1186/s13063-017-2089-y
2. Gamble C, Krishan A, Stocken D, Lewis S, Juszczak E, Doré C, et al. Guidelines for the content of statistical analysis plans in clinical trials. Journal of the American Medical Association 2017; 318: 2337-43. DOI: 10.1001/jama.2017.18556
3. Chondros P, Davidson S, Wolfe R, Gilchrist G, Dowrick C, Griffiths F, et al. Development of a prognostic model for predicting depression severity in adult primary patients with depressive symptoms using the diamond longitudinal study. Journal of Affective Disorders 2018; 227: 854-60. DOI: 10.1016/j.jad.2017.11.042
4. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a brief depression severity measure. Journal of General Internal Medicine 2001; 16: 606-13. DOI: 10.1046/j.1525-1497.2001.016009606.x
5. Kroenke K, Spitzer RL, Williams JBW, Löwe B. Psychiatry and primary care: The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: A systematic review. General Hospital Psychiatry 2010; 32: 345-59. DOI: 10.1016/j.genhosppsych.2010.03.006
6. Arroll B, Goodyear-Smith F, Crengle S, Gunn J, Kerse N, Fishman T, et al. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. Annals of Family Medicine 2010; 8: 348-53. DOI: 10.1370/afm.1139
7. McMillan D, Gilbody S, Richards D. Defining successful treatment outcome in depression using the PHQ-9: A comparison of methods. Journal of Affective Disorders 2010; 127: 122-9. DOI: 10.1016/j.jad.2010.04.030
8. Clarke J, Proudfoot J, Birch M, Whitton A, Parker G, Manicavasagar V, et al. Effects of mental health self-efficacy on outcomes of a mobile phone and web intervention for mild-to-moderate depression, anxiety and stress: Secondary analysis of a randomised controlled trial. BMC Psychiatry 2014; 14: 1-10. DOI: 10.1186/s12888-014-0272-1
9. Spitzer RL, Kroenke K, Williams JBW, Lowe B. A brief measure for assessing generalized anxiety disorder: The GAD-7. Archives of Internal Medicine 2006; 166: 1092-7. DOI: 10.1001/archinte.166.10.1092

10. Richardson J, Iezzi A, Khan MA, Maxwell A. Validity and reliability of the Assessment of Quality of Life (AQoL)-8D multi-attribute utility instrument. Patient Centered Outcomes Research 2014; 7: 85-96. DOI: 10.1007/s40271-013-0036-x

11. The Centre for Health Economics MU. AQoL-8D utility algorithms. Available from: https://www.monash.edu/business/che/aqol/using-aqol/scoring; cited 2019 March

12. Richardson J, Sinha K, Iezzi A, Khan MA. Modelling utility weights for the Assessment of Quality of Life (AQoL)-8D. Quality of Life Research 2014; 23: 2395-404.

13. Mihalopoulos C, Chen G, Iezzi A, Khan MA, Richardson J. Assessing outcomes for cost-utility analysis in depression: Comparison of five multi-attribute utility instruments with two depression-specific outcome measures. British Journal of Psychiatry 2014; 205: 390-7. DOI: 10.1192/bjp.bp.113.136036

14. Engel L, Chen G, Richardson J, Mihalopoulos C. The impact of depression on health-related quality of life and wellbeing: Identifying important dimensions and assessing their inclusion in multi-attribute utility instruments. Quality of Life Research 2018; 27: 2873-84.

15. Australian Government Department of Health. MBS Online Available from: http://www.mbsonline.gov.au/internet/mbsonline/publishing.nsf/Content/Home; cited 2019 March

16. Australian Government Department of Health. Pharmaceutical Benefits Scheme (PBS). Available from: http://www.pbs.gov.au/pbs/home; cited 2019 March

17. Chatterton ML, Chambers S, Occhipinti S, Girgis A, Dunn J, Carter R, et al. Economic evaluation of a psychological intervention for high distress cancer patients and carers: Costs and quality-adjusted life years. Psycho-Oncology 2016; 25: 857-64. DOI: 10.1002/pon.4020

18. Chatterton ML, Mihalopoulos C, O'Neil A, Itsiopoulos C, Opie R, Castle D, et al. Economic evaluation of a dietary intervention for adults with major depression (the "SMILES" trial). BMC Public Health 2018; 18.

19. Spitzer RL, Kroenke K, Williams JBW. Validation and utility of a self-report version of PRIME-MD: The PHQ Primary Care Study. Journal of the American Medical Association 1999; 282: 1737.

20. Qualtrics Labs Inc. Qualtrics. Provo, Utah2018.

21. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap): A metadata-driven methodology and workflow process for providing translational research informatics support. Journal of Biomedical Informatics 2009; 42: 377-81. DOI: 10.1016/j.jbi.2008.08.010

22. Fransen J, Moens HB, Speyer I, van Riel PLCM. Effectiveness of systematic monitoring of rheumatoid arthritis disease activity in daily practice: a multicentre, cluster randomised controlled trial. Ann Rheum Dis 2005; 64.
23. StataCorp. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC; 2017.
24. National Health and Medical Research Council. National statement on ethical conduct in human research 2007 (Updated 2018). Canberra: Commonwealth of Australia; 2007.
25. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2019.
26. White IR, Carpenter J, Horton NJ. Including all individuals is not enough: Lessons for intention-to-treat analysis. Clinical Trials 2012; 9: 396-407. DOI: 10.1177/1740774512450098
27. Fitzmaurice GM, Laird NM, Ware JH. Applied Longitudinal Analysis. 2nd ed. Hoboken, N.J: Wiley; 2011.
28. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: Current practice and problems. Statistics in Medicine 2002; 21: 2917-30.
29. Lee KJ, Thompson SG. The use of random effects models to allow for clustering in individually randomized trials. Clinical Trials 2005; 2: 163-73.
30. Candlish J, Teare MD, Dimairo M, Flight L, Mandefield L, Walters SJ. Appropriate statistical methods for analysing partially nested randomised controlled trials with continuous outcomes: A simulation study. BMC Medical Research Methodology 2018; 18: 105. DOI: 10.1186/s12874-018-0559-x
31. Carpenter J, Pocock S, Lamm CJ. Coping with missing data in clinical trials: A model-based approach applied to asthma trials. Stat Med 2002; 21: 1043-66.
32. White IR, Kalaitzaki E, Thompson SG. Allowing for missing outcome data and incomplete uptake of randomised interventions, with application to an Internet-based alcohol trial. Statistics in Medicine 2011; 30: 3192-207. DOI: 10.1002/sim.4360
33. Little RJA. Pattern-mixture models for multivariate incomplete data. Journal of the American Statistical Association 1993; 88: 125-34. DOI: 10.1080/01621459.1993.10594302
34. Dunn G, Maracy M, Dowrick C, Ayuso-Mateos JL, Dalgard OS, Page H, et al. Estimating psychological treatment effects from a randomised controlled trial with both non-compliance and loss to follow-up. British Journal of Psychiatry 2003; 183: 323.

35. Sussman JB, Hayward RA. An IV for the RCT: Using instrumental variables to adjust for treatment contamination in randomised controlled trials. British Medical Journal 2010; 340.

36. Anderson L, Lewis G, Araya R, Elgie R, Harrison G, Proudfoot J, et al. Self-help books for depression: How can practitioners and patients make the right choice? British Journal of General Practice 2005; 55: 387-92.

37. Proudfoot J, Goldberg D, Mann A, Everitt B, Marks IM, Gray JA. Computerized, interactive, multimedia cognitive-behavioural program for anxiety and depression in general practice. Psychological Medicine 2003; 33: 217-27.

38. Newby J, McKenzie A, Williams A, McIntyre K, Watts S, Wong N, et al. Internet cognitive behavioural therapy for mixed anxiety and depression: A randomized controlled trial and evidence of effectiveness in primary care. Psychological Medicine 2013; 43: 2635-48. DOI: 10.1017/S0033291713000111

39. Independent Hospital Pricing Authority. National Hospital Cost Data Collection (NHCDC). Available from: https://www.ihpa.gov.au/what-we-do/nhcdc; cited 2019 March

40. Australian Institute for Health and Welfare. Mental health services in Australia. Available from: https://www.aihw.gov.au/reports/mental-health-services/mental-health-services-in-australia/report-contents/general-practice; cited 2018 August

41. Department of Health. Manual of resource items and their associated unit costs. Canberra: Australian Government; 2016.

42. Glick HA, Doshi JA, Sonnad SS, Polsky D. Economic evaluation in clinical trials (2nd ed.). Oxford: Oxford University Press; 2014.