An online randomised controlled trial to evaluate the clinical and cost effectiveness of a peer supported self-management intervention for relatives of people with psychosis or bipolar disorder: Relatives Education And Coping Toolkit (REACT)

Statistical Analysis Plan version 2.0 20/12/17

	ORIGINATED BY	QC PERFORMED BY	APPROVED BY
Name	Dr Susanna Dodd	NA	NA
Title	Senior statistician	NA	NA
Date	December 20, 2017		
Protocol Version and Date	Protocol version v1.8 Dated 06.12.17		

1. Change Control

Protocol version	Updated SAP version no.	Section number changed	Description of change	Date changed
v1.8 06.12.17	2.0	17.3	The time spent on the final webpage of a given login session will be unavailable; therefore, it is necessary to impute this time. Details for this imputation have been added.	02.11.17
		17.4	Number of decimal places reported for p-values has been simplified.	02.11.17
		18.1	Methods for imputation of missing postal GHQ-28 questionnaire data have been added, as postal questionnaires are now being used to aid follow up.	02.11.17
		19.1	An additional IV regression sensitivity analysis has been added, on request of the IDSMC, including all follow up (i.e. beyond 30 weeks). The interaction between randomisation and home internet access as a potential second instrument has been removed, as the vast majority of the intervention arm have home internet access; the lack of variation in this variable means that it will not be suitable as a second instrument.	02.11.17
		19.1	An additional exploratory IV regression analysis relating to the impact of “lurking” has been added, on request of the CI, as there is increased interest in the hypothesis that “lurking” leads to a worse impact than use/non-use of social forums.	02.11.17

2. Approval and agreement

SAP Version Number being approved: v2.0

Trial Statistician

Name _____

Signed _____ Date _____

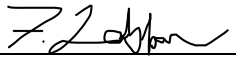
Senior Statistician

Name _____

Signed _____ Date _____

Chief Investigator/clinical lead

Name Fiona Lobban

Signed  Date 30/01/2018

OR Electronic approval attached ☐

3. Roles and responsibilities

N Bacon (Department of Biostatistics, University of Liverpool) – trial statistician

S Dodd (Department of Biostatistics, University of Liverpool) – statistics team leader

P Williamson (Department of Biostatistics, University of Liverpool) – CTRC director

F Lobban (Spectrum Centre, Division of Health Research, Lancaster University) – CI

H Robinson (Spectrum Centre, Division of Health Research, Lancaster University) – trial manager

Author's contributions

S Dodd proposed and drafted the statistical analysis plan. N Bacon checked that the report plan was executable and consistent with the protocol. P Williamson read the report plan to ensure compliance with CTRC processes and policy. F Lobban read, amended and approved the statistical analysis plan. H Robinson ensured the statistical plan was consistent with the study protocol and ethical approval and amendments.

4. List of abreviations and definitions of terms

AR	Adverse reaction
BD	Bipolar Disorder
CRF	Case report form
HE	Health Economic
IDSMC	Independent Data and Safety Monitoring Committee
IQR	Inter-quartile range
RD	Resource Directory
REACT	Relatives' Education And Coping Toolkit
SAE	Serious adverse event
SD	Standard deviation
TAU	Treatment as usual

Contents

5. Statement of Compliance

This Statistical Analysis Plan (SAP) provides a detailed and comprehensive description of the pre-planned final analyses for the study “REACT” . The planned statistical analyses described within this document are compliant with those specified in brief within the REACT protocol (v1.8 06.12.17).

This study is carried out in accordance with the World Medical Association Declaration of Helsinki (1964) and the Tokyo (1975), Venice (1983), Hong Kong (1989) and South Africa (1996) amendments and will be conducted in compliance with the protocol, Clinical Trials Research Centre (CTRC) Clinical Trials Unit (CTU) Standard Operating Procedures (SOPs) and EU Directive 2001/20/EC, transposed into UK law as the UK Statutory Instrument 2004 No 1031: Medicines for Human Use (Clinical Trials) Regulations 2004.

These planned analyses will be performed by the trial statistician. The results of the final analysis described within this statistical analysis plan will be contained within a statistical analysis report. This report will be used as the basis of the primary research publications according to the study publication plan.

All analyses are performed with standard statistical software (SAS/Stata version 13 or later; SAS will be used unless stated otherwise). The finalised analysis datasets used for all IDSMC/final analyses, programs and outputs will be archived following Good Clinical Practice guidelines and SOP TM021 Archiving procedure in CTRC. (Note that the trial sponsor, Lancaster University, will also be responsible for archiving the final dataset.) The testing and validation of the statistical analysis programs will be performed following SOP ST001. Data sharing will follow CTRC policy (guidelines in development).

6. Background and Rationale

Relatives of people with psychosis or bipolar disorder (BD) provide a large amount of vital unpaid care, but at huge personal cost in terms of high levels of distress, a significant practical/financial/emotional burden, and increased use of healthcare services [5, 6]. A recent review of the quality of mental health services has identified improving support for relatives as a national priority. REACT (Relatives’ Education And Coping Toolkit) is an online supported self-management toolkit for relatives of people with psychosis or BD that has been developed to meet this need. If effective, REACT could meet the support needs of relatives of people with psychosis or BD across the UK, who currently provide free informal

care that saves the NHS over £1.2 billion per year. This trial will test the clinical and cost effectiveness of REACT to achieve this aim.

7. REACT Study Objectives

The key objective of this trial is to determine the effectiveness of REACT compared to the standard Resource Directory (RD) in reducing distress and stress of those in a caring role (relative/friend) for service users. In particular the

1) Clinical effectiveness of REACT on the following outcomes:
    a. Primary outcome is relatives' distress at 24 weeks assessed using the General Health Questionnaire (GHQ-28 [32])
    b. Secondary outcomes are
        i. relatives' distress at 12 weeks assessed using the GHQ-28, and
        ii. relatives' wellbeing at 12 and 24 weeks assessed using the Carers' Well-Being and Support Measure (CWS [33])

2) Cost effectiveness of REACT in reducing distress for relatives using the following measures:
    a. Costs of delivering the intervention versus NHS and productivity cost savings in use of health services and paid work (adapted version of the CSRI [34])
    b. Cost effectiveness – cost of significant unit change (defined as 3 point reduction) in primary outcome (GHQ-28)
    c. Cost utility – marginal cost of any changes in marginal change in quality adjusted life years (QALYs) (making use of the EQ-5D-5L [35], as recommended by NICE)

Note that analysis of the cost effectiveness measures recorded in (2) above will be undertaken (and the corresponding health economics (HE) analysis plan will be prepared) by the health economist with delegated responsibility for this analysis.

8. Investigational Plan and Study Design

8.1 Overall study design and plan – description

REACT is a two-arm, single blind, online parallel group trial comparing two forms of online intervention for relatives/friends providing a caring or supportive role for a person with psychosis or BD (refered to as service users). Treatment is randomly allocated in a 1:1 ratio.

8.2 Treatments studied

This two arm pragmatic online RCT compares the "REACT (Relatives' Education And Coping Toolkit) including Resource Directory + Treatment As Usual (TAU)" versus "Resource Directory + TAU". See relevant sections of the protocol for more detail ("The intervention – REACT", "Resource Directory" and "Treatment as Usual (both arms of the trial)").

8.3 Treatment compliance

Participants are not required to adhere to a prescribed level of resource use; instead they are advised to use the intervention (REACT or RD) according to their level of need.

8.4 Patient population studied

This study will take place online with 666 relatives/friends of people with psychosis or BD living in the UK (confirmed via valid postal address).

8.5 Inclusion criteria

See relevant section of the protocol ("Inclusion/exclusion criteria").

8.6 Exclusion criteria

See relevant section of the protocol ("Inclusion/exclusion criteria").

8.7 Withdrawal of patients from therapy or assessment

There are no predetermined reasons for which participants will be removed from their randomised intervention. Therefore all randomised participants will be included in trial analyses and no participants will be removed from analysis (unless they withdraw their questionnaire data within 2 weeks of completion).

8.8 Consent process

See relevant section of the protocol ("Screening and online informed consent").

8.9 Blinding

Trial staff (including Chief Investigator (CI) and Trial Manager (TM)) and assessors (including trial statisticians) will be blinded to treatment assignment. Participants will be unblinded, as well as the REACT supporters (and co-applicants providing guidance to

REACT supporters in the event of the risk protocol being triggered) and IS/website personnel involved in administering and maintaining the intervention (website).

8.10 Method of assignment to treatment

Web-based blocked randomisation with varying block sizes will be implemented at baseline following eligibility and consent checks. An additional (2x2 factorial) randomisation will take place at 24 weeks to assess the impact of two incentive interventions (monetary reward of £10 or £20 offered conditionally or unconditionally on providing follow up data) to enhance completion of 24 weeks' follow up data. See "Randomisation" section of the protocol for more details.

8.11 Sequence and duration of all study periods

Participants will be followed up for 24 weeks following randomisation, as per the following flow chart:

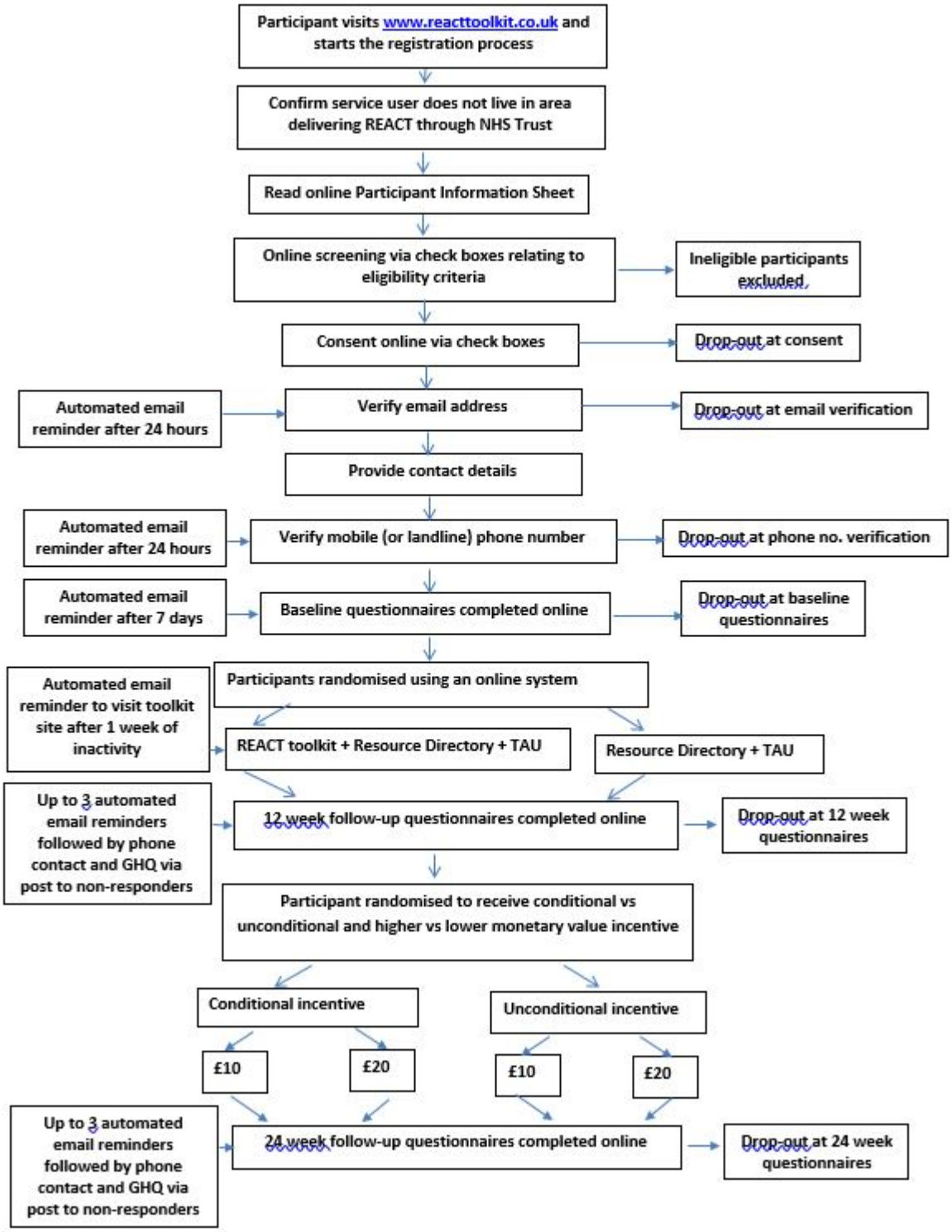


FIGURE 1 FLOW CHART

8.12 Schedule of assessments

Participants will be assessed at baseline, 12 weeks and 24 weeks following randomisation, as per the flow chart above.

9. Listing of Outcomes

17.1 Primary outcome(s)

The primary outcome is relatives'/friends' distress at 24 weeks assessed using the General Health Questionnaire (GHQ-28).

17.2 Secondary outcomes

The secondary outcomes are

- Relatives'/friends' distress at 12 weeks assessed using the General Health Questionnaire (GHQ-28)
- Relatives'/friends' wellbeing at 12 and 24 weeks assessed using the Carers' Well-Being and Support Measure (CWS)
- Costs of delivering the intervention versus NHS and productivity cost savings in use of health services and paid work (adapted version of the CSRI)
- Cost effectiveness – cost of significant unit change (defined as 3 point reduction) in primary outcome (GHQ-28)
- Cost utility – marginal cost of any changes in marginal change in quality adjusted life years (QALYs) (making use of the EQ-5D-5L, as recommended by NICE)

10. Determination of Sample Size

The null hypothesis for the REACT trial is that, on average, there will be no difference in GHQ scores between intervention groups at 24 weeks. A conventional power calculation demonstrates that a total sample of 466 participants would provide 90% power to reject the null hypothesis (at the 5% significance level) when the true effect size is 5.0 units, assuming a standard deviation of 16.6 (estimate obtained from feasibility study). Therefore this study aims to recruit 666 participants in order to allow for an estimate of 30% drop out.

Our feasibility study showed a mean difference in GHQ scores between groups at 6 months (controlling for baseline) of 6.59 units (standard deviation 16.6 units) in favour of the REACT arm. To build a degree of protection against pilot results proving optimistic, and to accommodate adaptations to the design of the study, we reduce our estimate of the mean difference in this trial to 5.0 units (s.d. 16.6). A detailed qualitative and quantitative analysis

of our feasibility data suggests that a (within-patient) reduction of 3 units on the GHQ can be used to indicate clinically meaningful change; however, the minimum difference required for the between-group comparison (of change in GHQ-28 from baseline) has been set at 5 units, in order to justify the staff and resource costs associated with delivery of the trial intervention. The sample size required to detect a significant between-group difference of 5 units with 90% power at the 5% significance level (233 per group) will provide 49% power to detect a difference of 3 units between groups.

The estimates of standard deviation and drop out rates will be assessed as part of the internal pilot to ensure that the target sample size is appropriate (see section 13.2).

11. Study Framework

The overall objective of this trial for each of the study outcomes (primary and secondary) is to test the superiority of the REACT intervention (used in combination with standard support services) compared with the Resource Directory (RD) (used in combination with standard support services).

12. Confidence Intervals, p-values and Multiplicity

The standard confidence interval level of 95% and corresponding 5% significance level will be used for the analysis of all primary, secondary and exploratory outcomes. There is no need to adjust significance levels for multiplicity, as there is only one primary outcome and no interim analyses are planned.

13. Timing and Objectives of Interim and Final Analyses

13.1 Interim monitoring and analyses

IDSMC reports will be prepared for IDSMC meetings to be held after 4 months and 9 months following the start of the recruitment period, in order to assess recruitment and retention in the internal pilot. Following the internal pilot, IDSMC reports will be produced for annual meetings of the IDSMC. There are no safety or futility concerns requiring outcome assessments using formal stopping rules, and thus there will be no formal analysis of outcome data for IDSMC reports beyond what is specified for the interim pilot.

13.2 Internal pilot

An estimated average monthly rate of over 37 relatives per month was assumed over the 18 months recruitment period in order to achieve the required sample size. An internal pilot will be undertaken by the IDSMC after 4 months and 9 months into the recruitment period, in

order to assess whether this target recruitment rate is being achieved (as well as the target retention rate of 70%). See IDSMC analysis report plan for more details.

If the internal pilot suggests that the trial should be stopped because of poor recruitment, this analysis plan will be applied to the data collected up to this point only.

13.3 Final analysis

Final analysis will take place once all participants have been followed up for 24 weeks. All outcomes will be analysed after the date of database lock. This is the date on which data modification privileges are withdrawn from the trial database.

14. Disposition of Participants

14.1 Screening, eligibility and recruitment

Completeness of follow-up will be presented in the form of a consolidated standards of reporting trials (CONSORT) flow diagram. The number lost to follow up and with missing data within each treatment group will be reported and the reasons where known will be documented. Monthly recruitment summaries will be presented against expected recruitment in a recruitment graph.

Participant progress through the recruitment process (obtained from web access data) will be summarised, from completion of eligibility screening, consent and baseline measures through to 12 and 24 week follow up.

The CONSORT flow diagram will summarise the number of participants in each intervention group (with reasons, where * indicated) who

- completed eligibility screening
- completed consent process
- completed baseline measures
- were randomised
- received their randomised allocation (i.e. accessed REACT)
- did not receive their randomised allocation/received alternative intervention*
- lost to follow-up*
- withdrew prematurely from the trial*
- randomised and included in the primary analysis
- randomised and excluded from the primary analysis*

Given that participants confirm eligibility and provide consent independently online, without input from any site staff, it is not possible to assess screening and eligibility in detail. However it will be possible to tabulate the eligibility criteria on which participants failed, as well as the consent items to which participants failed to agree. Key categories for premature withdrawal reasons can be found on the trial withdrawal database.

The number of randomised participants who found out about the trial via each of the possible sources (provided in a drop down list) will be tabulated and presented in a graph.

14.2 Post randomisation discontinuations

Withdrawal from intervention: The number of participants who withdraw from intervention use or follow up (at 12 and/or 24 weeks) will be presented.

Withdrawal from trial: Patients who fail to provide follow up information will be contacted via email reminding them to provide follow up. Participants who do not wish to provide follow up information will be requested to provide a reason. The number of participants who formally withdraw their consent from the trial will be presented, along with the number with missing follow up information at 12 weeks and 24 weeks. A breakdown of withdrawal/missing data rates according to the recruitment route will be provided.

15. Protocol Deviations

Possible protocol deviations are specified as minor or major in the REACT Monitoring Plan (finalised prior to any treatment allocations being revealed or any analyses performed) and are recorded within the protocol deviation database, including:

- Eligibility criteria violated (* see “Screening and online informed consent” section of the protocol)
- Nonadherence to randomised intervention in terms of cross-contamination of control arm (i.e. number of control arm participants who accessed the REACT intervention site) (* see “Screening and online informed consent” section of the protocol).
- Multiple registrations per participant (* see “Identity check” section of the protocol)
- Multiple participants per service user (* see “Clustering” section of the protocol)
- Failure to follow risk protocol (* see “Reporting Procedures” section of the protocol)
- Errors relating to online/email contact with participants

- Downtime of the online intervention (for example, due to routine maintenance or bug-fixing)

Note that there are measures in place to prevent the occurrence of those protocol deviations indicated by * (see the specified protocol section for more details).

The number (and percentage) of patients with at least one major/minor/all protocol deviation will be summarised by treatment group. The patients that are included in the intention to treat (ITT) analysis data set will be used as the denominator to calculate the percentages. No formal statistical testing will be undertaken.

16. Unblinding

See section 8.9 for information on which parties will be un/blinded in this trial.
It is possible that the TM may become unblinded if/when they make contact with participants to obtain follow up data or if any problems arise regarding the intervention (website), but this is not considered to be a problem, as this will occur on an individual case basis only. If the TM becomes unblinded during follow up contact with participants, the responsibility for future follow up of that individual will be passed on to the trial administrator (who is blinded to treatment assignment).

The number of participants whose assignment becomes unblinded to the TM will be reported for each treatment group, along with the time point and reason for unblinding.

17. Efficacy Evaluations

17.1 Data Sets Analysed

The principle of intention-to-treat, as far as practically possible, will be the main strategy of the analysis adopted for the primary outcome and all the secondary outcomes. These analyses will be conducted on all randomised participants, in the group to which they were allocated, and for whom the outcome(s) of interest have been completed. No imputations will be made; instead joint modelling will be used to assess the impact of missingness on outcome.

Per protocol analysis will not be considered (i.e. participants will not be excluded from analysis on the basis of any protocol deviations listed in section 15).

17.2 Demographic and Other Baseline Characteristics

Patient baseline data, as categorised within the CRFs, will be presented descriptively with respect to:

- age
- gender
- ethnicity
- marital status
- living arrangements
- dependents
- highest education level
- employment status
- home internet access
- caring role

Baseline data will be presented for the following questionnaires

- The General Health Questionnaire (GHQ 28)
- Relatives (Carers) Well-being and Support (CWS)
- Brief IPQ
- Brief COPE
- Questionnaire about caring/support role
- Questionnaire about income

These will be presented overall and separately for the two randomised groups.

Health economic measures (EQ-5D-5L and Client Service Receipt Inventory, CSRI) will be presented as part of the HE analysis (see HE analysis plan).

Categorical data will be summarised by numbers and percentages. Continuous data will be summarised by mean, standard deviation (SD) and range if data are normally distributed, and median, inter-quartile range (IQR) and range if data are skewed. Minimum and maximum values will also be presented for continuous data. Tests of statistical significance will not be undertaken for baseline characteristics; rather the clinical importance of any imbalance will be noted.

The mean, median and standard deviation of baseline variables will be presented to one decimal point beyond that of the original scale of measurement (e.g. age is measured in years, and therefore the mean, median and standard deviation of the ages in each treatment arm will be presented to one decimal point) while the range will be presented to the same number of decimal places as the original scale of measurement.

17.3 Compliance with treatment

Website usage data downloaded from the intervention site will be summarised for participants randomised to both intervention groups. These website use data will be condensed into a small number of summary covariates to be used to determine causal effects of the intervention (see section 19). These summaries will be presented overall and for each module of the REACT intervention separately (i.e. split into the 12 information modules, the forum and direct messaging) as well as the resource directory, in order to determine the most commonly used features of the intervention.

Intervention use will be summarised overall and split according to module (using either mean and SD, or median and IQR, depending on distribution) for participants in the REACT group, in terms of

1) each participants' total number of webpage downloads from the REACT intervention site over 24 weeks of follow up
2) each participants' total number of logins to the REACT intervention site over 24 weeks of follow up
3) each participants' total time spent logged on to REACT intervention site over 24 weeks follow up*.

* Note that inactivity time on a given page is capped at 20 minutes to allow for prolonged periods of inactivity when participants do not actively log off from the intervention. Given that these capped values are not likely to reflect the true time spent on a given page and are likely to skew the data, values including a capped inactivity period of 20 minutes for a given webpage will be replaced with the mean total time spent on that webpage for all participants randomised to the REACT intervention (excluding those with capped values for that webpage). Note that the time spent on the final webpage of a given login session for a participant is not available; therefore it will be necessary to impute the time spent on this webpage. If there is a video on this webpage, video feedback data will allow calculation of the time spent on this page accurate to within 5 seconds. If there is no video on this page, it

will be assumed that the time spent on this page is equal to the mean time that they have spent on all previous webpages to date.

Similar data will be summarised for control participants' access of the RD site. The number (%) of participants who do not login into their assigned intervention will also be presented for each randomised group, as well as the number (%) of participants in each randomised group who do not log in after their initial login.

In order to assess the hypothesis that digital interventions provide a distinct advantage in providing "out of hours" access, the timing of web access will be summarised in each treatment group, in particular relative to the working week (defined as 9am to 5pm Monday to Friday (excluding Bank Holidays) UK time i.e. BST or GMT depending to the time of year).

Additional analysis will explore whether reminders led to an increase in intervention use, by simply comparing participants' patterns of intervention use (described using measures of frequency and time spent on intervention) within 1 day, 3 days and 7 days of reminders being sent compared to their intervention use during the remainder of their 24 week follow up period.

17.4 Analysis of outcomes

Note that, as for baseline variables, the mean, median and standard deviation of all outcome variables will be presented to one decimal point beyond that of the original scale of measurement, while the range will be presented to the same number of decimal places as the original scale of measurement. Percentages will be presented to one decimal point, and p-values will be reported to four decimal places.

17.5 Primary Outcome: General Health Questionnaire (24 weeks)

Relatives'/friends' distress at 24 weeks assessed using the General Health Questionnaire (GHQ-28)

Note that data entered up to 18 weeks post-randomisation will be considered 12 week data; data entered beyond 18 weeks post-randomisation will be considered 24 week data, up to a maximum of 30 weeks post-randomisation. For the purposes of analysis, any duplication of 24 week data entered by a given participant (i.e. if participant provides "12" week follow up data beyond 18 weeks, as well as their "24" week follow up data) will be addressed by choosing the data entered closest to the participant's "24 weeks post-randomisation" date.

The distribution of times at which participants completed their 24 week outcome data will be presented. If >5% of participants completed their 24 week outcomes beyond 27 weeks post randomisation (beyond which no further follow up reminders are sent), a sensitivity analysis excluding those with follow up data completed beyond 27 weeks post randomisation will be carried out in order to assess the impact of these "late" data completions.

17.6 Derivation

The GHQ-28 score will be derived using the Likert scoring method as recommended by the author (http://www.gl-assessment.co.uk/products/general-health-questionnaire/faqs), assigning a value between 0 to 3 to each possible multiple choice answer (with higher scores indicating more severe distress) and resulting in a total score between 0 and 84:

- 0 assigned to "Not at all" (or "Better/more so/quicker than usual", "Definitely not" or "More satisfied")
- 1 assigned to "No more than usual" (or "Same as usual" or "About the same (as usual)" or "I don't think so")
- 2 assigned to "Rather more than usual" (or "Worse than usual" or "Rather less than usual" or "Less well than usual" or "Longer than usual" or "Less satisfied" or "Has crossed my mind")
- 3 assigned to "Much more than usual" (or "Much longer than usual" or "Much less satisfied" or "Much less than usual" or "Definitely have")

The alternative "caseness" scoring method (assigning either a 0 or 1 to each answer) will be presented as a secondary analysis:

- 0 assigned to the first two categories listed above (assigned a score of 0 or 1 in the Likert scale)
- 1 assigned to the last two categories listed above (assigned a score of 2 or 3 in the Likert scale)

Furthermore the total 28-item score can be divided into 4 subscales of 7 items each. These will be explored for secondary analyses:

- A – somatic symptoms (items 1-7)
- B – anxiety/insomnia (items 8-14)
- C – social dysfunction (items 15-21)
- D – severe depression (items 22-28)

17.7 Analysis

The distribution of the total GHQ-28 score at 24 weeks will be assessed visually by a histogram and q-q plot. If normally distributed, total GHQ-28 score will be summarised using means and standard deviations for each randomised group separately, and will be compared between randomised groups using analysis of covariance, adjusting for baseline GHQ-28. If the scores are not normally distributed, the median and interquartile range (IQR) will be presented for each randomised group, and will be compared using the Mann Whitney U test. An appropriate transformation (e.g. log transformation) will be applied, and analysis of covariance will be applied to the transformed data, adjusting for baseline GHQ-28 score.

A joint modelling approach will be used to check whether there is any difference in GHQ-28 (of the longitudinal outcome rather than the outcome at 24 weeks alone) between the randomised arms adjusted for missingness (at 12 week or 24 week follow up) (see section 18.2).

Data collected from participants at 12 week and 24 weeks follow up will be summarised for each treatment group. The baseline characteristics of those who do/do not provide 12 week and 24 week follow up data will be compared, in order to demonstrate whether or not 12 week and 24 week follow up can be assumed to be missing at random (at least with respect to recorded baseline characteristics). Any clinically important differences between those who do/do not provide follow up data will be noted.

Each of the four subscales (somatic symptoms, anxiety/insomnia, social dysfunction, severe depression) will be summarised for each randomised group using means and standard deviations (if normally distributed) or medians and IQRs (if not normally distributed), and compared between randomised groups using ANCOVA, adjusting for the corresponding baseline subscale score, (if normally distributed) or Mann Whitney U tests (if not normally distributed). Multivariate ANCOVA (“manova” in Stata) will be used to assess the impact of intervention group on each of these subscales, while taking into account correlation between the subscales. Four multivariate statistics (commonly reported in MANOVA) will be presented: Wilks’ lambda, Pillai’s trace, Lawley–Hotelling trace and Roy’s largest root. If the subscales are not normally distributed, an appropriate transformation (e.g. log transformation) will be applied, and multivariate ANCOVA will be applied to the transformed data, adjusting for the corresponding baseline subscale score.

The “stjm” code in Stata will be used to fit multivariate joint models, in order to assess the impact of missing outcome data on each of the four subscales, allowing for correlation between subscales.

17.8 Secondary outcome: General Health Questionnaire (12 weeks)

Relatives’/friends’ distress at 12 weeks assessed using the General Health Questionnaire (GHQ-28)

Note that data entered up to 18 weeks post-randomisation will be considered 12 week data; data entered beyond 18 weeks post-randomisation will be considered 24 week data, up to a maximum of 30 weeks post-randomisation. The distribution of times at which participants completed their 12 week outcome data will be presented. If >5% of participants completed their 12 week outcomes beyond 15 weeks post randomisation (beyond which no further follow up reminders are sent), a sensitivity analysis excluding those with follow up data completed beyond 15 weeks post randomisation will be carried out in order to assess the impact of these “late” data completions.

17.9 Derivation

GHQ-28 at 12 weeks will be derived as for GHQ-28 at 24 weeks (see section 17.6).

17.10 Analysis

The distribution of the total GHQ-28 score at 12 weeks will be assessed visually by a histogram and q-q plot. If normally distributed, total GHQ-28 score will be summarised using means and standard deviations for each randomised group separately, and will be compared between randomised groups using analysis of covariance, adjusting for baseline GHQ-28. If the scores are not normally distributed, the median and interquartile range (IQR) will be presented for each randomised group, and will be compared using the Mann Whitney U test. An appropriate transformation (e.g. log transformation) will be applied, and linear regression will be applied to the transformed data, adjusting for baseline GHQ-28 score.

Each of the four subscales (somatic symptoms, anxiety/insomnia, social dysfunction, severe depression) will be summarised for each randomised group using means and standard deviations (if normally distributed) or medians and IQRs (if not normally distributed), and compared between randomised groups using ANCOVA, adjusting for the corresponding baseline subscale score, (if normally distributed) or Mann Whitney U tests (if not normally distributed). Multivariate ANCOVA (“manova” in Stata) will be used to assess the impact of intervention group on each of these subscales, while taking into account correlation between

the subscales. Four multivariate statistics (commonly reported in MANOVA) will be presented: Wilks' lambda, Pillai's trace, Lawley–Hotelling trace and Roy's largest root. If the subscales are not normally distributed, an appropriate transformation (e.g. log transformation) will be applied, and multivariate ANCOVA will be applied to the transformed data, adjusting for the corresponding baseline subscale score.

The "stjm" code in Stata will be used to fit multivariate joint models, in order to assess the impact of missing outcome data on each of the four subscales, allowing for correlation between subscales.

17.11 Secondary outcome: Carers' Well-Being and Support Measure (CWS)

Carers' wellbeing at 12 and 24 weeks assessed using the Carers' Well-Being and Support Measure (CWS)

Note that data entered up to 18 weeks post-randomisation will be considered 12 week data; data entered beyond 18 weeks post-randomisation will be considered 24 week data, up to a maximum of 30 weeks post-randomisation. For the purposes of analysis, any duplication of 24 week data entered by a given participant (i.e. if participant provides "12" week follow up data beyond 18 weeks, as well as their "24" week follow up data) will be addressed by choosing the data entered closest to the participant's "24 weeks post-randomisation" date.

The distribution of times at which participants completed their 12 and 24 week outcome data will be presented. If >5% of participants completed their 12/24 week outcomes beyond 15/27 weeks post randomisation (beyond which no further follow up reminders are sent), a sensitivity analysis excluding those with follow up data completed beyond 15/27 weeks post randomisation (for 12/24 week outcomes respectively) will be carried out in order to assess the impact of these "late" data completions.

17.12 Derivation

The CWS consists of two separate scales measuring Carers' Well-Being and Support.

The well-being scale consists of 32 questions relating to their caring/support role, relationship with the person they care for/support, their relationship with family and friends, their financial situation, their physical health, their emotional well-being, stigma and discrimination, their safety and the safety of the person they care for/support. The carer rates

their level of concern in each question using a 5-point likert scale (between “A lot” and “Not at all”). The overall well-being score is derived by assigning a score of 0 to the most negative answer (“A lot of concern”) and a score of 4 to the most positive answer (“Not at all”) to the answer of each question, and summing scores from all 32 questions, creating an overall score between 0 and 128 (such that a higher score indicates better well-being).

- Assign a score of 0 to each question that the carer has ticked “A lot”
- Assign a score of 1 to each question that the carer has ticked “Quite a bit”
- Assign a score of 2 to each question that the carer has ticked “Moderately”
- Assign a score of 3 to each question that the carer has ticked “A little”
- Assign a score of 4 to each question that the carer has ticked “Not at all”
- **Total well-being score is equal to the sum of these scores from all 32 questions.**

Similarly, the support scale consists of 17 questions relating to the carers’ level of satisfaction with the information and advice they received, their involvement in treatment and care planning and support from medical and care staff, each measured on a 4-point likert scale (between “very satisfied” and “very dissatisfied”). The overall support score is derived by summing scores from these 17 questions, with a score of 0 assigned to the most negative answer (“very dissatisfied”) and a score of 3 to the most positive answer (“very satisfied”), creating an overall score between 0 and 51 (such that a higher score indicates better support).

- Assign a score of 0 to each question that the carer has ticked “Very dissatisfied”
- Assign a score of 1 to each question that the carer has ticked “Somewhat dissatisfied”
- Assign a score of 2 to each question that the carer has ticked “Somewhat satisfied”
- Assign a score of 3 to each question that the carer has ticked “Very satisfied”
- **Total support score is equal to the sum of these scores from all 17 questions.**

17.13 Analysis

The distribution of the two CWS scores (well-being score and support score) at both 12 and 24 weeks will be assessed visually by a histogram and q-q plot. If normally distributed, each score (measured at 12 and 24 weeks separately) will be summarised using means and standard deviations for each randomised group separately and compared using analysis of covariance, adjusting for baseline CWS score. If the scores are not normally distributed, the median and interquartile range (IQR) will be presented for each randomised group, and will

be compared using the Mann Whitney U test. An appropriate transformation (e.g. log transformation) will be applied, and analysis of covariance will be applied to the transformed data, adjusting for baseline scores.

A joint modelling approach will be used to check whether there is any difference in CWS (of the longitudinal outcome rather than the outcome at 24 weeks alone) between the randomised arms adjusted for missingness (at 12 week or 24 week follow up) (see section 18.2).

Data collected from participants at 12 week and 24 weeks follow up will be summarised for each treatment group. The baseline characteristics of those who do/do not provide 12 week and 24 week follow up data will be compared, in order to demonstrate whether or not 12 week and 24 week follow up can be assumed to be missing at random (at least with respect to recorded baseline characteristics). Any important differences between those who do/do not provide follow up data will be noted.

18. Missing data and withdrawals

Missing data will be monitored and strategies developed to minimise its occurrence; however as much data as possible will be collected about the reasons for missing data and this will be used to inform the handling of missing data. Participants are asked for reasons for withdrawal from follow up (at 12 weeks, 24 weeks or both) and for withdrawal from the trial.

The numbers (with reasons if available) of losses to follow-up and withdrawals over the course of the trial will be summarised by treatment arm. This will be presented in a CONSORT diagram alongside a table, with numbers and reasons for withdrawal and/or exclusion from analysis given.

18.1 Missing data analyses

Data collected from participants at 12 week and 24 weeks follow up will be summarised for the primary outcome separately for each treatment group. The baseline characteristics of those who do/do not provide 12 week and 24 week follow up data will be compared, in order to demonstrate whether or not 12 week and 24 week follow up can be assumed to be missing at random (at least with respect to recorded baseline characteristics). Any important differences between those who do/do not provide follow up data will be noted.

The online data collection system has been set up to ensure that participants are required to complete the primary outcome measure (GHQ-28) before completing any other measures, in

order to maximise follow up for the primary outcome. In addition, participants are unable to submit any questionnaire with any missing fields; therefore, by design, there will be no missing data within questionnaires. Missing data will only arise when participants fail to submit a questionnaire, rather than because of missing data within returned questionnaires.

The only instance where missing outcome data may arise is in the case of primary outcome (GHQ-28) data returned by post. In such instances, the participant's GHQ-28 data will be included in analysis under the following conditions: if >50% of questions from each subscale of the GHQ-28 are complete, each of the four subscale scores will be calculated by assuming that the missing answers to questions within that subscale take the mean value of that participant's answers for all completed questions within that subscale.

Note that, when completion dates are missing for postal GHQ-28 questionnaires, the midpoint between dates of sending and receiving questionnaires will be used to estimate the date of completion. If not recorded, the date of receipt will be imputed using the mean number of days between sending and receiving questionnaires for all postal GHQ-28 questionnaires with accurately recorded sending and receiving dates.

For this reason, missing data imputation will not be carried out. Instead a joint modelling approach (using baseline, 12 week and 24 outcome data) will be used to assess the impact of missing data at 24 weeks on the conclusions drawn from analysis on each primary and secondary efficacy outcome; i.e. this analysis will assess whether there is any difference in outcome (here the longitudinal outcome rather than the outcome at 24 weeks alone) between the randomised arms adjusted for missingness, by inherently allowing for the correlation between patterns in missingness and outcome.

18.2 Joint modelling analysis

Joint modelling of the longitudinal outcome data and time to dropout will be carried out using the "stjm" command in Stata, in order to demonstrate any association between these two processes. A longitudinal trajectory plot will be produced using the subsidiary command "stjmgraph"; this graph will highlight any potential dependence between the longitudinal profiles and dropout.

19. Causal analysis

In order to investigate the relationship between website use and outcome, data will be recorded on baseline covariates (correlated with both website use and outcome) and

relevant website use (from participants in both randomised arms). Appropriate causal methods will be implemented to estimate efficacy of actual website use on the primary outcome (GHQ-28 at 24 weeks), supplementing ITT estimates of effectiveness.

19.1 Instrumental variable regression

In particular, **instrumental variable regression** will be employed to estimate the impact of intervention use on outcome using the ivreg command. Intervention use (i.e. of the REACT intervention site) will be summarised as a single continuous covariate derived from web usage data. Given the extensive password protection of the REACT intervention site, it will be assumed that participants assigned to the control arm did not access the REACT intervention site over the 24 weeks of follow up (i.e. null intervention use). Therefore, web usage will be recorded for REACT participants only, and will be defined in terms of each participants' total number of webpage downloads from the REACT intervention site over 24 weeks of follow up for REACT participants (summarised using either mean and SD, or median and IQR, depending on distribution). Exploratory analyses using other continuous measures of website use (participants' total number of logins over 24 weeks of follow up and total time spent logged on to REACT intervention site) will also be presented. The suitability of using randomisation as the instrument in this regression will be assessed using tests of exogeneity, redundancy and under/weak identification, which are reported as part of the execution of the ivreg commands.

An additional sensitivity analysis using IV regression will include all outcome data (i.e. including data received beyond 30 weeks) to assess whether prolonged exposure to the intervention further impacts on outcome. Additional exploratory analyses will be conducted as to the impact on this IV model of the baseline variables that were included in the final regression model (along with baseline GHQ-28 score) for the primary outcome (see section 21.1).

An exploratory causal analysis will be carried out to investigate the hypothesis that "lurking" (reading but not posting) on social forums lead to worse outcomes than posting on (or not logging on at all to) social forums. In the absence of a second instrument to facilitate this three-way comparison (lurking vs non-use vs use), observational methods (adjusting for baseline covariates that are likely to confound the relationship between lurking/use/non-use of social forums and outcome) will be used to assess the causal impact of lurking (vs non-use vs use) on GHQ-28 at 24 weeks. The validity of this analysis will be assessed by repeating this model for the binary comparison user vs non-user (with lurkers included as

users, as the null category must reflect pure non-use to adhere to the "exclusion restriction") adjusting for the same baseline covariates, and comparing the group effect from this model with that obtained using IV regression for the users vs non-users comparison (with randomisation as the instrument). The validity of ordinary linear regression for this binary comparison will also be evident from the test of exogeniety reported as part of the ivreg command. Note that lurkers are defined as those who log on to the forum at least once over the 24 weeks but do not post any comments, users are defined as those who log on and post at least once on the forum over 24 weeks, and non-users are defined as those who never log on to the forum over 24 weeks.

20. Mediation analysis

The hypothesised mechanism of change for the REACT intervention is that participants' primary outcome (distress levels as measured by GHQ-28) is mediated by their understanding of the service user's disorder, their insight into the service user's experiences and their perceived ability to cope. These potential mediators have been measured by the Brief IPQ (including an additional single question added to the BIPQ which assesses perceived coping) and COPE.

20.1 Process measure: Brief Illness Perception Questionnaire

Brief illness perception questionnaire (IPQ) measured at 12 and 24 weeks follow up

20.1.1 Derivation

The brief IPQ consists of 16 questions (rated by the carer between 0 and 10) related to perceived illness by the service user or the carer, and can therefore be summarised into two total scores, one relating to the service user and the other relating to the carer. The answer to 7 questions must be inverted (according to the formula where *x* becomes 10-*x)*, in order that a higher value indicates more severe perception of illness consistently across all 16 questions.

20.1.2 Service user score

The service user score is derived by summing the ratings (between 0 and 10) for each of the following questions (using inverted scores as indicated), resulting in a total score between 0 and 80 (with a higher score indicating more severe perception of illness):

- Scores from questions 1, 3, 8, 12 and 13
- Inverted scores (10-*x*) from questions 4, 7 and 10

20.1.3 Carer score

The carer score is derived by summing the ratings (between 0 and 10) for each of the following questions (using inverted scores as indicated), resulting in a total score between 0 and 70 (with a higher score indicating more severe perception of illness):

- Scores from questions 2, 9, 14 and 16
- Inverted scores (10-*x*) from questions 5, 11 and 15

20.1.4 Additional item on coping

The additional item added to the brief IPQ (number 6) will be presented separately:

- Inverted scores (10-*x*) from question 6

20.1.5 Analysis

Data collected from participants at 12 week and 24 weeks follow up will be summarised for each treatment group. The baseline characteristics of those who do/do not provide 12 week and 24 week follow up data will be compared, in order to demonstrate whether or not 12 week and 24 week follow up can be assumed to be missing at random (at least with respect to recorded baseline characteristics). Any important differences between those who do/do not provide follow up data will be noted.

The two BIPQ scores (service user score and carer score), as well as the additional item on coping, at both 12 and 24 weeks will be summarised for each randomised group using means and standard deviations (if normally distributed) or medians and IQRs (if not normally distributed), and compared between randomised groups using ANCOVA, adjusting for the corresponding baseline subscale score (if normally distributed), or Mann Whitney U tests (if not normally distributed). Multivariate ANCOVA (“manova” in Stata) will be used to assess the impact of intervention group on each of these subscales, while taking into account correlation between the subscales. Four multivariate statistics (commonly reported in MANOVA) will be presented: Wilks’ lambda, Pillai’s trace, Lawley–Hotelling trace and Roy’s largest root. If the subscales are not normally distributed, an appropriate transformation (e.g. log transformation) will be applied, and multivariate ANVOCA will be applied to the transformed data, adjusting for the corresponding baseline score.

The “stjm” code in Stata will be used to fit multivariate joint models, in order to assess the impact of missing outcome data on each of the two scores and the additional item on coping, allowing for correlation between each of these scores.

20.2 Process measure: COPE

Brief COPE measured at 12 and 24 weeks follow up

20.2.1 Derivation

The brief COPE consists of 28 questions (rated by the carer between 1 and 4) related to extent of coping methods used by the service user to cope with the stress related to caring for the service user.

The following values are assigned to the answers from each question:

- A value of 1 is assigned to the answer “I haven’t been doing this at all”
- A value of 2 is assigned to the answer “I’ve been doing this a little bit”
- A value of 3 is assigned to the answer “I’ve been doing this a medium amount”
- A value of 4 is assigned to the answer “I’ve been doing this a lot”

There is no overall score for the brief COPE; instead the following subscales are computed as follows:

- Self-distraction: questions 1 and 19
- Active coping: questions 2 and 7
- Denial: questions 3 and 8
- Substance use: questions 4 and 11
- Use of emotional support: questions 5 and 15
- Use of instrumental support: questions 10 and 23
- Behavioural disengagement: questions 6 and 16
- Venting: questions 9 and 21
- Positive reframing: questions 12 and 17
- Planning: questions 14 and 25
- Humour: questions 18 and 28
- Acceptance: questions 20 and 24
- Religion: questions 22 and 27
- Self-blame: questions 13 and 26

20.2.2 Analysis

Each of the above subscales will be summarised for each intervention group at 12 and 24 weeks separately using medians and IQR, and compared between randomised groups using Mann Whitney U tests. Multivariate ANCOVA (“manova” in Stata) will be used to assess the impact of intervention group on each of these subscales, while taking into account

correlation between the subscales. Four multivariate statistics (commonly reported in MANOVA) will be presented: Wilks' lambda, Pillai's trace, Lawley–Hotelling trace and Roy's largest root. If the subscales are not normally distributed, an appropriate transformation (e.g. log transformation) will be applied, and multivariate ANVOCA will be applied to the transformed data, adjusting for the corresponding baseline score.

The "stjm" code in Stata will be used to fit multivariate joint models, in order to assess the impact of missing outcome data on each of the two scores and the additional item on coping, allowing for correlation between each of these scores.

20.2.3 Mediation analysis

In order to test whether these potential mediators actually predict change in outcome, mediation analysis will be carried out using instrumental variable regression (ivreg in Stata, as for causal analysis above), adjusting for each of the baseline variables that were included in the final regression model (along with baseline GHQ-28 score) for the primary outcome (see section 21.1). Each of the potential mediators will be assessed individually in this exploratory analysis:

- Overall BIPQ score
- Additional IPQ coping question
- Brief COPE summary scores (each individual score will be assessed individually)
    - Self-distraction
    - Active coping
    - Denial
    - Substance abuse
    - Use of emotional support
    - Use of instrumental support
    - Behavioural disengagement
    - Venting
    - Positive reframing
    - Planning
    - Humour
    - Acceptance
    - Religion
    - Self-blame

Due to the number of variables being considered, this mediation analysis will be entirely exploratory in order to generate hypotheses for confirmation in future studies.

21. Additional analyses

21.1 Prognostic variable regression model of primary outcome (total GHQ-28)

Exploratory analyses will be conducted to determine the impact of the following baseline factors (in addition to baseline GHQ-28 score) on the intervention effect on total GHQ-28 score at 24 weeks, by including the following baseline factors in the linear regression model along with baseline GHQ-28 score and randomised intervention:

- age (as older relatives are likely to have been caring for longer, and may be more likely to have their own physical health needs)
- gender (as women are likely to be impacted by service user, as they tend to have more input and take more responsibility in caring roles within society)
- ethinicity (as minority ethnic groups tend to be subject to barriers to receiving support)
- marital status (as a relationship may provide support for carer/relative)
- living arrangements (as above)
- dependents (may add to the stress of the carer/relative)
- highest education (as the intervention may be overly complex for those without sufficient reading/cognitive ability)
- employment (as work commitment may impede use of the intervention)
- home internet access (as this will aid ease of use of the intervention)
- caring role (as the impact of intervention is likely to be related to the burden of care)

Results from each individual model (i.e. the regression coefficient, 95% CI and p-value associated with each baseline factor when added to the model containing baseline GHQ-26 and randomised intervention) will be presented for each of these baseline factors separately. A stepwise selection procedure (with entry/exit criterion based on p-values of 0.05/0.1 respectively) will be used to determine which baseline variables are included in the final multivariable regression model (along with baseline GHQ-28 and randomised intervention).

This exploratory analysis will also include a formal test of a treatment–covariate interaction to assess the effect of relationship with service user (i.e. according to whether the relationship between service user and carer/relative is parental or not) by including the "intervention-service user (non/parental) relationship" interaction term in the linear regression model (along with baseline GHQ-28, randomised intervention and main effect of service user

(non/parental) relationship). This hypothesis is based on results from previous research that suggests a differential treatment effect depending on whether the relationship between service user and carer/relative was non/parental. However, in this previous research, the participants were predominantly mothers – thus it wil be of particular interest to examine the proportion of parental relationships that are mother/father. This will be reported, and if fathers make up more than 5% of the parents, the impact of a treatment-mother interaction will be examined as a secondary analysis.

21.2 Retention strategies

Our analysis will also include a test of the relative effectiveness and costs associated with a low value reward (£10) versus a higher value (£20) and a conditional versus unconditional reward. The sample size required in REACT RCT for sufficient power is 666. This was calculated expecting a 70% retention rate, for a conditional reward of £10. All available evidence suggests that increasing the reward, and making it unconditional, will both increase the retention rate. This sample will give us 84% power at the 5% significance level to determine a 10% absolute difference (from 70% to 80%) in retention between conditional or unconditional reward and lower value vs. higher value reward.

The retention rates (defined as the proportion of participants who provide primary outcome data, GHQ-28, at 24 weeks) will be assessed

1) according to the value of the reward (£10 versus £20)
2) according to the conditional or unconditional nature of the reward

In particular, the number (proportion) of participants providing 24 week follow up data will be presented and compared using the chi-square test (or the Fisher's Exact test, if any expected counts are <5). The independent impact of intervention group on retention rates will be explored by including intervention group along with value of the reward (or un/conditional nature of the reward) as an explanatory variable in logistic regression.

21.3 Recruitment strategies

Finally, we will compare the effect of tailored recruitment strategies (e.g. media advertising; Facebook; chat rooms) in the REACT RCT. Using data collected from participants about where they found about the REACT trial, we will compare the proportion of participants recruited to REACT via each of the recruitment strategies. We will also examine differences in demographic characteristics of participants entering the RCT via each of the recruitment strategies (including age, sex, ethnic group, education level and deprivation level).

21.4 Participants' experiences of the intervention

Participants in the REACT intervention group are asked the following questions (based on previous published work investigating perceived safety and satisfaction with online interventions) at 12 and 24 weeks post randomisation:

- To assess whether relatives feel supported by the REACT supporters: 'I always feel supported by the REACT Supporters (strongly disagree, disagree, agree, strongly agree)
- To assess whether relatives feel supported by the REACT Group: 'I always feel supported by the REACT GROUP (strongly disagree, disagree, agree, strongly agree)
- To assess whether relatives feel the site is safe: 'I always feel the REACT site was a safe and confidential environment (strongly disagree, disagree, agree, strongly agree)

The number (proportion) of participants in the REACT intervention group who answered each of these options will be presented for 12 weeks and 24 weeks separately.

21.5 Appropriate use of the site

The following data collected automatically via the site will be summarised for the REACT intervention group only:

- Number of times a relative flagged content as requiring attention
- Number of times the REACT Supporter hid a comment from the site
- Number of participants' accounts suspended

21.6 Clustering

As we will not individually identify service users in this study, it is theoretically possible that multiple family members could sign up to take part in the study. It is also theoretically possible that relatives from the same support group sign up to the study, which could also lead to a clustering effect if factors associated with the group impact on distress levels.

We will aim to minimise this by explaining in detail to relatives (and potential referrers) why it is very important that only one relative per service user takes part in the study. This will include a lay description of the concepts of clustering and contamination on the registration site and recruitment materials. Relatives are required to indicate at registration that to the best of their knowledge, they are the only relative of the person they care for who is taking part. Thus we cannot eliminate, but will aim to minimise, any risk of clustering. However, evidence to date regarding clustering effects suggests that even if this does occur, it is unlikely to have a major impact on the power of the study.

To illustrate, the design effect is 1+p(m-1), where p is the ICC and m is the average cluster size. If we assume an ICC of 0.05 (based on Baldwin et al 2011 doi: 10.1080/16506073.2010.520731 as the only relevant estimate), and an average cluster size of 2 (a worst case scenario meaning all the service users have two relatives included), the design effect would be 1.05. Our target sample size after dropouts is 466, so the effective sample size would be 466/1.05=444 and the power only slightly reduced to 88%.

As we will not be collecting information from participants that could be used to identify members of the same family or support group, it will not be possible to directly estimate the ICC for this analysis. Therefore, a sensitivity approach will be applied, either by multiplying the resulting standard error by sqrt(1.05) or by taking a good approximation to the significance level of 4.5% rather than 5%.

22. Safety Evaluations

22.1 Data sets analysed

Adverse events will be assessed only in terms of the number of participants for whom it was necessary to trigger the risk protocol (for both low risk and high risk events). These analyses will be carried out according to randomised group.

22.2 Presentation of the data

The risk protocol may be triggered either by the REACT Supporters (who identify a clinical or safeguarding risk via communication with participants by email or forum, and then escalate this to research team) or by the trial manager (who may identify a risk during follow up phone calls). These risk events will be flagged as low risk (non-immediate/severe risk) or high risk (immediate risk of severe harm).

Low risk events also include red flag answers given in response to the following questions from baseline/follow up questionnaires:

- GHQ-28 D3: Felt that life isn't worth living? Answered "Much more than usual"
- GHQ-28 D4: Thought of the possibility that you might make away with yourself? Answered "Definitely has"
- GHQ-28 D6: Found yourself wishing you were dead and away from it all? Answered "Much more than usual"
- GHQ-28 D7: Found that the idea of taking your own life kept coming into your mind? Answered "Definitely has"

- CWS Q29: During the past 4 weeks, how concerned were you about the person you care for/support being aggressive or threatening towards you? (e.g. verbal threats, sexual aggression, physical intimidation)
- CWS Q30: During the past 4 weeks, how concerned were you about the person you care for/support harming themselves? Answered “A lot”

Thus low risk events will be categorised as those (i) identified by REACT supporters (ii) identified by TM (iii) resulting from red flag responses to GHQ-28/CWS questions.

The number of participants for whom it was necessary to trigger the risk protocol will be presented overall and for each randomised group (separately for low risk and high risk events). These descriptive analyses will also include reference to the information source which informed the triggering of the risk protocol (for example, REACT online discussion forum, assessment questionnaire or TM follow up telephone call) in order to assess the potential for detection bias (when the likelihood of the detection of risk is associated with intervention arm). No formal statistical testing will be undertaken.

The number of red flag answers will be reported separately for each of the six questions listed above for baseline, 12 week and 24 week follow up, split by intervention group. In addition, the total number of participants in each intervention group with at least one red flag item will be presented, along with the total number of participants in each intervention group with multiple items.

23. References

1. International Conference on Harmonisation. Topic E3 Structure and content of clinical study reports.(CPMP/ICH/137/95). 1995
2. International Conference on Harmonisation. Topic E9 Statistical Principles for Clinical Trials (CPMP/ICH/363/96). 1998.
3. Guidance on Statistical Analysis Plans.
4. Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c332.