Protocol for RCT with mothers during the antenatal period

Provide the evidence as to whether an intervention to support mothers who are suffering from anxiety or depression during the antenatal (AN) period will make a significant difference to mothers receiving systemic therapy sessions in comparison to mothers in the control group

A **randomised controlled trial** will be carried out to examine:

whether an intervention offered to the Intervention Group to support mothers who are suffering from anxiety or depression during the antenatal period (normally after the 12th week of pregnancy) will make a significant difference to these parents and their babies in comparison to those in the control group

To describe this randomized controlled study, the 2018 Consolidated Standards of Reporting Trials – Social and Psychological Interventions (CONSORT-PSI 2018) Montgomery et al., 2018 will be followed.

For this purpose, the trial design, the participants, randomization, the intervention, the control group, measures, and analytic methods will be described.

Trial Design

A two-group design is appropriate:

Those mothers who score above the cut-off point on the EPDS and the GAD 7 with depression or anxiety will receive psychiatric assessment plus medication (if needed) plus a monthly phone call by a midwife, plus 12 sessions of systemic therapy in the **Intervention Group**

And a psychiatric assessment plus medication (if needed) plus a monthly phone call by a midwife in the **Control Group**

The outcomes

AN Mothers who score above the cut-off point on the EPDS and the GAD 7

with anxiety and depression in the Intervention Group will have a better outcome reflected through:

(i) a lower and statistically significant score on the EPDS and GAD7 when compared to those in the control group
(ii) a higher statistically significant score on the Dyadic Adjustment Scale Revised (R-DAS) when compared to those in the Control Group

The allocation ratio is intended to be as follows:
60 mothers; (30 ANmothers will be in the intervention group, whereas **30**ANmothers will be in the control group)

During the screening process, AN mothers who score above the cutoff point in the EPDS and the GAD7 will be put in one cluster. The randomisation into the intervention and the control group will take place through the process described below.

Randomisation

Participants will be randomly assigned using a computer-generated sequence to one of the two study arm groups, that is the intervention group or the control group. These randomly generated treatment allocations will be placed in sequentially numbered opaque, sealed envelopes. Therefore, once a participant consents to enter the study, an envelope will be opened by the clinician and the particiatiping mother will be offered the allocated treatment regime. The statistician responsible for generating the randomisation sequence will not be responsible for recruiting participants and will not be administering it. Altogether these processes will ensure that each participant has an equal chance of being assigned to any of the two groups and that the investigator cannot predict in which group each participant will be assigned to (Cochrane, 2013).

Mothers who withdraw from treatment after the randomization would have taken place would still be included in the intervention and control groups respectively according to how they would have been assigned to mitigate intervention effect bias (McCoy, 2017). The Intention To Treat Principle (ITT) will be applied, to reduce the risk of intervention bias.

Participants

Inclusion Criteria

1. Batches of 100 and later 150 mothers belonging to the 9 firms within the Gynae Outpatients Antenatal Clinic will be universally screened for mental health difficulties usually during the 12th week of pregnancy before the 19th week. The rationale behind the latter time frame is meant to give sufficient time for the intervention to take place before delivery.
2. Same-sex couples will be included.
3. A mother who scores as depressed on the EPDS or anxious on the GAD7 will be eligible to proceed for further psychiatric assessment through the MINI
4. AN Mothers who during their pregnancy are self referred or are referred by professionals to the Perinatal Mental Health Service (PMHS)

Exclusion Criteria

1. AN Mothers suffering from severe mental health problems such as Puerperal Psychosis, Schizophrenia, Bipolar Disorder, and Borderline personality Disorder
2. Mothers with drug addiction problems
3. Mothers living in a context of domestic violence
4. Mothers whose babies are at risk of being taken away from them by Child Protection Services
5. Lone Mothers

All the above-excluded mothers will be offered treatment as usual.

Recruitment of Participants

The screening process will take place as explained hereunder

A. Screening during the first prenatal booking visit

Screening will take place in two stages.

In the first instance screening would take place at two points during the perinatal period

(a) during the first prenatal booking visit at Mater Dei Hospital (normally taking place after the 12th week of pregnancy)
(b) In the first instance, over and above the standard check-up of the expecting mother, screening would take place by asking A BATCH of 100 or 150 AN mothers at a time who are assigned to the 9 firms offering a service to fill in a screening sheet including demographic data, the Whooley questions (Whooley et al 1997) which we refer to as Phase1 Assessment, to assess current mental health, and a question related to their relationship with their partner (See Screening Sheets 1 to 4 in the Appendix).
c. For this purpose, those AN mothers who would be attending for the first prenatal booking visit, would be sent an Information sheet and Consent form explaining the project by post. Should they agree to take part in the project, they will be invited to fill in an enclosed screening sheet either online or by hand. Those AN mothers who would need help to fill in the screening sheets, will have at their disposal a phone in service, available for an hour a day during weekdays, where they could phone for more information and support.
d. Moreover, on the day of the booking visit, mothers who might have forgotten to fill in the screening sheet but would be willing to fill it up, would be asked to do so by the midwife on duty at the Perinatal Mental Health Service. Support to fill in the screening sheet would be offered if needed. The screening sheets would be collected by the midwife or the recruiters within the Perinatal Mental Health Service for further follow up as shall be described further down.

B. Screening of referrals received by the Perinatal Mental Health Service (PMHS)
From September 2022 mothers who are referred directly by doctors and midwives to the PMHS followed the same assessment procedure as above in terms of Phase1Assessment

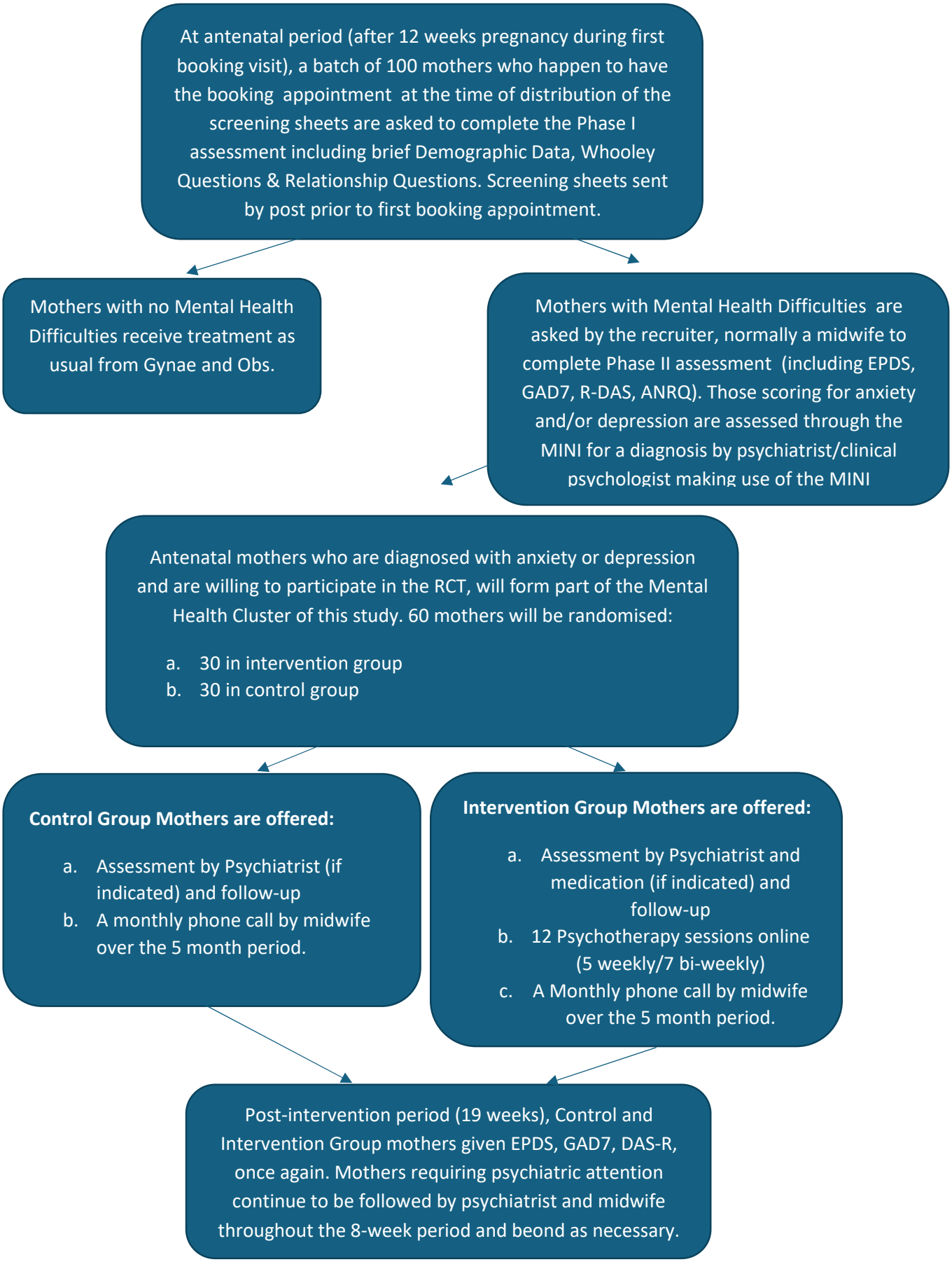
Screening would take place in a stepwise approach:

In the first instance, over and above the standard check-up of the expecting mother, screening would take place by asking her to fill in a screening sheet including demographic data, the Whooley questions to assess current mental health, and a question related to her relationship with her partner (See Screening Sheets 1 to 4 in the Appendix).

Mothers deemed needing more support following the traditional check-up of the mother at booking

A Flowchart drawn overleaf explains the steps

TRIAL FLOW CHART – (Mental Health Cluster) ANTENATAL MOTHERS



As per **TIDier checklist** the intervention for first cluster on Mental Health Difficulties will be as follows:

1. Name of intervention: 12 psychotherapy sessions over 19 weeks (the first five sessions on a weekly basis in the first five weeks, and the remaining seven sessions taking place fortnightly. In addition to assessment of a psychiatrist and medication if indicated plus 5 monthly phone calls by midwife.

2. Why this intervention: The reason for choosing the psychotherapeutic sessions is based on literature that suggests that psychotherapy whether it is interpersonal, cognitive behavioural or non directive is helpful. Pearson (2008) reports that psychotherapy especially interpersonal therapy has been found to be effective in the case of antenatal depression for mothers., whereas Misri et al (2015) also recommend psychotherapy and mindfulness for those mothers suffering from anxiety. Considering the current training of family therapists in Malta and their subsequent employment within the health service, we also wanted to find out whether family therapists trained in systemic psychotherapy would be able to support antenatal mothers effectively in the case of antenatal depression or anxiety.

3. What does intervention consist of: a total of 12 sessions of psychotherapy lasting a maximum of one hour each will be given to those parents indicated as showing mental health difficulties.

4. The psychotherapy may take different formats such as individual sessions for those presenting with depression and/or anxiety, couple therapy for those with anxiety and or depression and a strained couple relationship.

5. Procedure used in intervention: Participants who have been randomised into the intervention group will be offered a set of 12 appointments for sessions which would need to take place over a period of 19 weeks with their designated therapist.

6. Intervention provider: the 12 sessions of psychotherapy will be offered by a qualified systemic psychotherapist/s. This professional will set the appointments with the respective participants.

7. How is the intervention delivered: the 12 sessions of psychotherapy will be offered remotely using platforms such as zoom, due to circumstances that do not allow for face-to-face contact. These circumstances are primarily related to Covid-19 safety measures, and/or challenges for parents who may find it particularly difficult to be physically present for their appointments

8. The setting where the remote sessions will take place would be the therapists' and participants' respective households/office. What is important is that privacy and confidentiality can be maintained at all times.

9. When and how much intervention: The sessions will start to be offered following randomisation into the intervention group. 12 sessions of one hour each with each participant.

The first 5 sessions will be offered on a weekly basis, the remaining seven will be offerred forthnightly

10. Tailoring of intervention: Each respective therapist offering the sessions to participants will follow the same time-frame, number of sessions and duration. The aim of each therapist would be to focus on supporting the participants with their mental health difficulties.

11. Modifications to intervention: There were no modifications

12. The Control group will receive psychiatric medication (if necessary) and receive a monthly phone call from the midwife

13. Adhering to intervention: The psychotherapist will be able to report to the research manager, for the sake of accountabilty and verification of sessions, and to discuss if there are any difficulties in the process. These meetings would support the psychotherapists in their work with participants, whilst also helping them to adhere to the provision of service.

Measures

All Pre-Measures will take place before the interventions for all participants prior to randomisation into intervention and control groups respectively.

Post measures will take place after the intervention which is of 19 weeks duration.

We will evaluate such an outcome with the following pre-post measures.

Pre Measures :

(i)EDPS for prenatal and postnatal depression of the parents

The Edinburgh Postnatal Depression Scale (EPDS) is one of the most widely used screening instruments for assessing symptoms of perinatal depression and anxiety both in a clinical setting and also for research purposes (Gibson et al, 2009; Kozinsky and Dudas, 2015).

It is a10-item scale written in the past tense with four possible responses graded according to the severity and duration of the symptom during the previous seven days. The scoring method is simple: 0 for absence of the symptom; 3 for maximum severity and duration; 1 and 2 which were intermediate. The total score ranged from 0 to 30.

This self-reporting instrument was originally developed in the U.K. Cox et al, (1987) specifically for childbearing women. The popularity of this brief instrument reflects the original British validation study

in which nine out of ten women who were diagnosed by a psychiatrist as being depressed after giving birth were correctly identified in a blinded comparison with scores above a cut-off on the EPDS. The psychometric properties of the EPDS in primary health care were: 86 % sensitivity (correctly identifying true cases), 78 % specificity (correctly identifying people without the condition) and 73 % positive predictive value (proportion of respondents scoring positive in the test who had a mental disorder diagnosed by clinical interview) (Cox et al, 1987).

The EPDS was acceptable to women, took only 5 minutes to complete and was rapidly scored. It screens only for depression and those who score below the cut-off should not be assumed to have no psychiatric disorder. A higher response rate (95%) is obtained when the EPDS is sent by post and careful follow-up of non-responders is carried out (Cox et al, 1993).

The Maltese translated EPDS had good face validity and had satisfactory sensitivity and specificity for identifying depression both antenatally and postnatally. Cut-off scores of 13/14 antenatally and 11/12 postnatally are recommended when using the Maltese EPDS (Felice et al, 2005). A recent thematic review and meta-analysis of individual participant data on the accuracy of the EPDS for screening to detect major depression among pregnant and postpartum women (Levis et al 2020) found that EPDS cutoff of ≥ 11 maximized combined sensitivity and specificity; a cutoff of ≥ 13 was less sensitive but more specific. To identify women in pregnancy and postpartum with higher symptom levels, a cutoff of 13 or greater could be used. Lower cutoffs could be used if the intention is to avoid false negatives and identify most patients who meet diagnostic criteria.

(ii) GAD 7

The GAD-7 (Spitzer et al.,2006) consists of 7 items and is a self-rated assessment developed to screen for GAD in primary care populations.

The GAD-7 score is calculated by assigning scores of 0, 1, 2, and 3, to the response categories of 'not at all', 'several days', 'more than half the days', and 'nearly every day', respectively, and adding together the scores for the seven questions.

Scores of 5, 10, and 15 are taken as the cut-off points for mild, moderate and severe anxiety, respectively. When used as a screening tool, further evaluation is recommended when the score is 10 or greater.

Using the threshold score of 10, the GAD-7 has a sensitivity of 89% and a specificity of 82% for GAD. It is moderately good at screening three other common anxiety disorders - panic disorder (sensitivity 74%, specificity 81%), social anxiety disorder (sensitivity 72%, specificity 80%) and post-traumatic stress disorder (sensitivity 66%, specificity 81%).

Test re-test

A test-retest was carried out using a randomly selected sample of 23 ante-natal and post-natal mothers to assess its reliability. It was administered twice allowing a one-week period between the first and second administration, followed by an item analysis of the two sets of responses. The English version was first administered, followed by the Maltese version. Internal consistency was first checked for GAD-7 for both English and Maltese versions using Cronbach's Alpha. Results indicated an excellent internal consistency for the English version questionnaire between items with a Cronbach’s alpha value of 0.893. The Cronbach’s alpha value for the Maltese version increased slightly to 0.897. Both versions indicated an excellent internal consistency between the items.

Cohen's Weighted kappa was then used as a measure of agreement for ordinal scales between the English and Maltese versions. The percentage agreement of GAD-7 ranged from 39.1% to 73.9% and all the p-values resulted to be greater than 0.05 level of significance, indicating sufficient test-retest reliability (Table 1).

Question	Percentage of Agreement	Cohen's Weighted Kappa	P-value	Result
Q1. Feeling nervous, anxious, or on edge?	69.6%	0.642	<0.001	Statistically significant
Q2. Not being able to stop or control worrying	65.2%	0.485	<0.001	Statistically significant
Q3. Worrying too much about different things	39.1%	0.297	0.031	Statistically significant
Q4. Trouble relaxing	73.9%	0.721	<0.001	Statistically significant
Q5. Being so restless that it's hard to sit still	73.9%	0.713	<0.001	Statistically significant
Q6. Becoming easily annoyed or irritable	56.5%	0.536	<0.001	Statistically significant
Q7. Feeling afraid as if something awful might happen	60.9%	0.521	<0.001	Statistically significant

Table 1: Test-retest results of GAD-7

Finally, the GAD-7 total score for the seven items was calculated for both English and Maltese versions. An intraclass correlation was used to check for the correlation (agreement) between the Maltese and English version. The estimated reliability between the Maltese and English versions was 0.819, with 95% Confidence Interval (0.625, 0.919), which indicates an excellent reliability. Moreover, the F-statistic value is 10.071 with degrees of freedom (df1 and df2) equal to 22 and 23, respectively. One can note that the p-value is <0.001. Since the resulting p-value is less than 0.05, this indicates that there is a statistically significant agreement between the Maltese and English versions for GAD-7 total score - *ICC = 0.819 (95% CI, 0.625 to 0.919), F(22,23) = 10.071, p < 0.00.*

(iii) DAS-R Dyadic Adjustment Scale Revised For parents who are recruited during the Prenatal and the Postnatal period, the Dyadic Adjustment Scale-Revised will be used. Why use it? As a predictor of good coparenting. The Scale is a 14 item self-report tool, to assess couple satisfaction and to evaluate how each partner within the couple perceives his or her relationship.

The DAS-R is widely used and considers married and cohabiting couples. It takes about six or seven minutes to complete and there are several validations of it around the world . Several studies across the globe established the reliability, validity, and stability of the instrument (Hollist et al 2012). This instrument is not validated on a Maltese population.

Test re-test

A test-retest technique was carried out using a randomly selected sample of 36 ante-natal and post-natal mothers to assess its reliability. Out of these 36 respondents, only 15 (41.7%) were valid for the analysis. DAS-R was administered twice allowing a one-week period between the first and second administration, followed by an item analysis of the two sets of responses. The English version was first administered, followed by the Maltese version.

Internal consistency was first checked for DAS-R for both English and Maltese versions using Cronbach's Alpha. Results indicated a good internal consistency for the English version questionnaire between items with a Cronbach's alpha value of 0.711. The Cronbach's alpha value for the Maltese version increased slightly to 0.781. Both versions indicated a good internal consistency between the items.

Cohen's Weighted kappa was then used as a measure of inter-rater agreement for ordinal scales. The percentage agreement of DAS-R ranged from 66.7% to 100.0%, with all p-values for each question resulted to be less than 0.05 level of significance, indicating sufficient test-retest reliability between the English and Maltese versions (Table 2).

Finally, the DAS-R total score for the fourteen items was calculated for both English and Maltese versions. An intraclass correlation was used to check for the correlation (agreement) between the Maltese and English version. The estimated reliability between the Maltese and English versions was 0.903, with 95% Confidence Interval (0.743, 0.966), which indicates an excellent reliability. Moreover, the F-statistic value is 19.618 with degrees of freedom (df1 and df2) equal to 14 and 15, respectively. One can note that the p-value is <0.001. Since the resulting p-value is less than 0.05, this indicates that there is a statistically significant agreement between the Maltese and English versions for DAS-R total score - *ICC = 0.903 (95% CI, 0.743 to 0.966), F(14,15) = 19.618, p < 0.00.*

Question	Percentage of Agreement	Cohen's Weighted Kappa	P-value	Result
Q1. Religious matters	80.0%	0.727	<0.001	Statistically significant
Q2. Demonstrations of affection	80.0%	0.743	<0.001	Statistically significant
Q3. Making major decisions	80.0%	0.563	0.011	Statistically significant
Q4. Sex relations	86.7%	0.659	<0.001	Statistically significant
Q5. Conventionality (correct or proper behaviour)	73.3%	0.670	<0.001	Statistically significant
Q6. Career decisions	80.0%	0.751	<0.001	Statistically significant
Q7. How often do you discuss, or have you considered divorce, separation or terminating your relationship?	80.0%	0.717	<0.001	Statistically significant
Q8. How often do you and your partner quarrel?	73.3%	0.587	<0.001	Statistically significant
Q9. Do you ever regret that you married (or lived together)?	100.0%	1.000	<0.001	Statistically significant
Q10. How often do you and your mate "get on each other's nerves"?	73.3 %	0.694	<0.001	Statistically significant
Q11. Do you and your mate engage in outside interests together?	66.7%	0.400	0.012	Statistically significant
Q12. Have a stimulating exchange of ideas?	73.3%	0.773	<0.001	Statistically significant
Q13. Work together on a project	66.7%	0.730	<0.001	Statistically significant
Q14. Calmly discuss everything	73.3%	0.689	<0.001	Statistically significant

Table 2: Test-retest results of DAS-R

(iv) *The Antenatal Risk Questionnaire (ANRQ): Acceptability and use for psychosocial risk assessment in the maternity setting*

ROC curve analysis for the ANRQ yielded an acceptable area under the curve of 0.69. The most 'clinically' useful cut off on the ANRQ was a score of 23 or more, yielding a sensitivity of 0.62 and specificity of 0.64 with positive predictive value of 0.3. The odds that a woman scoring 23 or more on the ANRQ is also a case was 6.3 times greater than for a woman scoring less than 23. Acceptability of the ANRQ was high among both women and midwives.

The ANRQ is a highly acceptable self-report psychosocial assessment tool that aids in the prediction of women who go on to develop postnatal depression. In combination with a symptom based screening measure (e.g., the Edinburgh Postnatal Depression Scale) and routine questions relating to drug and alcohol use and domestic violence, the ANRQ becomes most useful as a key element of a "screening intervention" aimed at the early identification of mental health risk and morbidity across the perinatal period. Evaluation of this model in terms of clinical outcomes remains to be undertaken.

Cut off point:

When using an ANRQ cut-off point of 31 or more (equal costs of errors), we obtained a sensitivity of 0.35, specificity of 0.89 and a positive predictive value of 0.43. A score of 23 or more gave approximately equal sensitivity (0.62) and specificity (0.64) and a positive predictive value of 0.30.

Test re-test

A test-retest technique was carried out using a randomly selected sample of 21 ante-natal mothers to assess the ANRQ reliability. Out of these 21 respondents, 16 (76.2%) had valid responses for the analysis. ANRQ was administered twice allowing a one-week period between the first and second administration, followed by an item analysis of the two sets of responses. The English version was first administered, followed by the Maltese version. Cohen's kappa was used to measure inter-rater agreement for the binary (Yes-No) questions and its extension, Cohen's weighted kappa was used as a measure of inter-rater agreement for ordinal scales.

The percentage agreement of ANRQ ranged from 57.1% to 100.0%, with all p-values for each question resulted to be less than 0.05 level of significance, indicating sufficient test-retest reliability (Table 3).

Finally, the ANRQ total score was calculated for both English and Maltese versions. An intraclass correlation was used to check for the correlation (agreement) between the Maltese and English version. The estimated reliability between the Maltese and English versions was 0.877, with 95% Confidence Interval (0.691, 0.955), which indicates an excellent reliability. Moreover, the F-statistic value is 15.262 with degrees of freedom (df1 and df2) equal to 15 and 16, respectively. One can note that the p-value is <0.001. Since the resulting p-value is less than 0.05, this indicates that there is a statistically significant agreement between the Maltese and English versions for ANRQ total score - *ICC = 0.877 (95% CI, 0.691 to 0.955), F(15,16) = 15.262, p < 0.00.*

Question	Percentage of Agreement	Kappa Value	P-value	Result
Q1. When you were growing up, did you feel your mother was emotionally supportive of you?	100.0%	1.000	<0.001	Statistically significant
Q2a. Have you ever had 2 weeks or more when you felt particularly worried, miserable, or depressed?	93.8%	0.818	<0.001	Statistically significant
Q2b. Do you have any other history of mental health problems?	100.0%	1.000	<0.001	Statistically significant
Q2c. Seriously interfere with your work and your relationships with friends and family?	69.2%	0.527	<0.001	Statistically significant
Q2d. Lead you to seek professional help?	100.0%	1.000	<0.001	Statistically significant
Q2e. Did you take tablets/herbal medicine?	100.0%	1.000	<0.001	Statistically significant
Q3. Is your relationship with your partner an emotionally supportive one?	93.8%	0.923	<0.001	Statistically significant
Q4a. Have you had any stresses, major changes or losses in the last 12 months?	87.5%	0.750	0.002	Statistically significant
Q4b. How distressed were you by these stresses, changes or losses?	57.1%	0.720	0.004	Statistically significant
Q5. Would you generally consider yourself a worrier?	62.5%	0.460	0.010	Statistically significant
Q6. In general, do you become upset if you do not have order in your life (e.g. regular time table, a tidy house)?	81.3%	0.726	<0.001	Statistically significant
Q7. Do you feel you have people you can depend on for support with your baby?	75.0%	0.701	<0.001	Statistically significant
Q8. Were you emotionally abused when you were growing up?	100.0%	1.000	<0.001	Statistically significant
Q9. Have you ever been abused sexually or physically?	100.0%	1.000	<0.001	Statistically significant

Table 3: Test-retest results of ANRQ

Ethics

The Ethics Application for the study was approved by FREC and UREC. UREC-DP2011001SWB - 6543_05102020_Angela Abela

Analytic Methods

The analysis of the data will be conducted by the Statistical Package for the Social Sciences, IBM SPSS 28 and R Studio.

RCT 1 the antenatal period: Provide the evidence as to whether an intervention to support mothers who are suffering from anxiety or depression during the antenatal period will make a significant difference to these mothers viz a viz the parents in the control group

Baseline characteristics will be compared between the two groups (Intervention/Control) to observe whether the two groups showed any demographic differences. To assess baseline participant differences, binominal logistic regression will be used to test for demographic differences across the two study groups – the intervention group and the control group. Any baseline demographic factor that differed between the two groups at $p < 0.10$ will be included as a potentially confounding variable in follow-up models testing for the effect of the intervention. Effect sizes will be given in terms of odds ratios for analyses using binomial logistic regression, , with 1 meaning no difference, values under 1 meaning less likely, and values over 1 meaning more likely to occur in the intervention compared to control group. However, this approach will be feasible if there is enough detail to conduct such analysis for certain demographic variables. If such problem arises, alternatively, a Chi-Square test of association will be carried out by selected demographic variables to see whether the two groups showed any demographic differences. The null hypothesis specifies that there is no association between the two variables and will be accepted if the p-value exceeds the 0.05 level of significance. The effect size for the chi-square test was measured using Cramer's V.

Pre-and post-means, medians, and standard deviations for all the measures, namely the GAD-7, EPDS and R-DAS will be compared and analysed between the respective groups, namely for the intervention and control group participants. Descriptive measures will also be presented visually using boxplots comparing pre-and post- test scores for each group. The paired samples t-test can be used to compare the mean scores of the pre-and post-sessions of both groups. If the differences of the pre-post measures deviate from the normality assumption, the Wilcoxon Signed-Ranks test should then be used to compare the scores between the pre-post periods for both intervention and control groups. The change of the pre- and post-test scores can be categorized into three groups as no change, lower

post-test score and higher post-test score and compared through the Chi-Square test between intervention and control to assess whether the increase or decrease in the post-test score is differing among the two groups.

Next, to assess the effect of the study condition, there are three adequate ways to analyse a numerical outcome (GAD-7, EPDS and R-DAS) where the common point of these approaches is that the pre-scores must be considered (the best approach will be chosen accordingly):

1. Independent samples-test on change of scores (i.e., the post-minus-pre differences), assuming the normality assumption is satisfied. The overall test is statistically significant if it yields a sufficiently small p-value. Following the hypothesis testing, the Cohen's d effect size statistic will be calculated for intervention effects. If a nonparametric approach is chosen (assuming normality assumption is deviated), the Mann-Whitney test should be performed on change scores. The effect size for the Mann-Whitney U test is calculated by dividing the standardised test statistic z and square root of the number of pairs n.

2. Two-way mixed-model ANOVA with group as between-subjects factor (intervention and control groups) and time-point as within-subject factor (i.e., repeated measures; with two levels, baseline, and post-intervention). Note that the p-value for interaction in this design will be equivalent to the p-value from the independent samples t-test on change scores mentioned in (1). This will also answer the essential question of the study: do the changes from baseline differ between the groups?

3. Analysis of covariance (ANCOVA) of post-intervention scores with baseline score as the covariate. This approach differs subtly from the two mentioned above. Analysis of covariance treats the pre-test value as a covariate that can be a source of variation that may influence post-test scores, and accordingly the post-test score is regressed on both the pre-test score and the grouping variable. ANCOVA adjusts the pre-test scores, increasing the power to determine whether or not there has been a treatment effect. This approach tests the null hypothesis of no difference between the control and treatment post.

Should the results indicate that there has been significant improvement at post-testing, one should also evaluate for any differences in the change of scores between the two groups through stepwise linear regression using a bootstrapping procedure with 5000 replications recommended for analyses with smaller sample sizes (Preacher and Hayes, 2008). Groups alone will be included in Model 1, and any confounds that differed across the two groups at baseline will be included in Model 2, allowing for the comparison of the effect size as a function of study condition in unadjusted and adjusted models. Results will be determined to be statistically significant if p values were <0.05.

Furthermore, it would be of interest to see whether there is an association between a high score in ANRQ and the incidence of depression and anxiety. Thus, the idea is to check for any possible correlation between ANRQ and EPDS and ANRQ and GAD-7 for both pre- and post-testing, through the Pearson correlation or alternatively the Spearman correlation, if the assumption of normality is not satisfied.

The assumption of normality to perform all the above tests will be evaluated by the Shapiro-Francia test. If the assumption of normality has been violated, the p-value will be less than 0.05. If the assumption of normality has not been violated, the p-value will be greater than 0.05. This is because the Shapiro-Wilk test is evaluating the null hypothesis that the distribution of the data is equal to a normal distribution. Rejecting the null hypothesis means that the distribution of the data is not equal to a normal distribution.