Building resilience through socio-emotional training programme (ReSET): Statistical analysis plan

Peter Martin and the ReSET team

Version history log

Version	Date	Details of Change
01	29 August 2024	-

1. Study Summary

For full details see the trial protocol (Viding et al., 2024).

Title	Building Resilience through Socio-Emotional Training (ReSET) programme: a cluster randomized controlled trial of a new transdiagnostic preventative intervention for adolescents
Short title	Resilience through Socio-Emotional Training (ReSET)
Chief Investigators	Prof. Pasco Fearon & Prof. Essi Viding
Statistician	Dr Peter Martin
Design	A cluster-randomised controlled trial with young people in school years 7-9 who are at risk of mental health problems. Within each participating school, school years will be randomised to receive either the ReSET preventative intervention or not (passive control).
Primary objective	To evaluate the effect of the ReSET preventative intervention on two primary outcomes: mental wellbeing as measured by the Edinburgh-Warwick Mental Wellbeing Scale (short form); and psychopathology symptoms as measured by the Strengths and Difficulties Questionnaire.
Population	Young people attending secondary school, aged 12 – 14 years. Target sample size: 540.
Study Type	Interventional randomised controlled trial
IRAS Number	322531

2. Introduction

2.1. Purpose and scope of the statistical analysis plan

This Statistical Analysis Plan was written by Peter Martin and describes the main statistical analyses to be applied to the data from the ReSET trial.

2.2. Timing of Analysis

The analyses described within this analysis plan will begin to be performed after all the data from the primary endpoint (post-intervention) have been entered, checked and locked and this analysis plan has been finalised. Further analyses will be performed after all the data from the 1-year-follow up have been entered, checked and locked.

2.3. Data checking

Before analysis and database lock, basic checks will be performed on the quality of the data, focusing on identifying:

- Missing data
- Data outside expected range
- Other inconsistencies between variables, e.g. in the dates the questionnaires were completed

If any inconsistencies are found, the corresponding values will be double checked with the researchers and corrected if necessary in the source data. This checking process and subsequent changes will be documented.

3. Description of the trial

3.1 Intervention

The intervention combines a group intervention (Interpersonal Therapy – Adolescent Skills Training, IPT-AST) and computerised cognitive-emotional training. IPT-AST consists of eight group sessions, as well as one pre-group and one mid-group session that is completed individually or with the attendance of a parent or carer. All group sessions are 90 minutes in length, while individual sessions are 60 minutes each. Computerized cognitive-emotional training tasks will be completed by participants as part of these sessions. (See the trial protocol for details.)

3.2 Randomization

The trial uses cluster randomisation at the school year level. In each school, we will have a younger and older cohort. In each school, one cohort is allocated to the hybrid intervention

arm and the other to the non-intervention control arm; this is balanced across participating schools using block randomization. Researchers are blind to participants' group allocation during data collection. For details see the trial protocol.

3.3 Duration of the treatment period and frequency of follow up

In both arms, data will be collected at baseline, after the conclusion of the intervention (after about 10 weeks in the control group) and at twelve months from randomisation. In addition, social network data will be collected alongside the screening data from entire year groups in addition to the pre- and post-assessment timepoints with study participants. The analysis of network data is not part of this statistical analysis plan.

4. Data Collection

4.1 Participant characteristics

The following demographic characteristics will be collected at baseline: date of birth, sex, gender, sexual orientation, pubertal status, and ethnicity.

4.2 Outcome data

Outcome data will be collected at baseline, end of treatment, and twelve months follow-up.

4.2.1 Primary outcomes

The two primary outcomes are psychopathology symptoms and mental wellbeing. Psychopathology symptoms will be assessed using the Total Difficulties Score of the Strengths and Difficulties Questionnaire (SDQ). Mental Wellbeing will be measured by the summary score of the Warwick and Edinburgh Mental Wellbeing Scale (WEMWBS). The primary endpoint will be the immediate post-intervention assessment point. Psychopathology symptoms and mental wellbeing are conceptualized as dual primary outcomes (rather than co-primary outcomes), and the statistical analysis will adjust for multiple hypothesis tests. For details see section 5.5.

4.2.2 Secondary outcomes

The following secondary outcomes will be measured:

- Depression: Patient Health Questionnaire (PHQ-8)
- Generalized Anxiety: Generalized Anxiety Disorder Assessment-7 (GAD-7)
- Sleep phenotypes: bespoke questionnaire (see trial protocol)
- Alcohol use (AUDIT)
- Drug use (DUDIT)

- General psychopathology, internalizing problems, and externalizing problems: these will be measured as latent factors in a bifactor confirmatory factor analysis using items from the SDQ and the Me & My Feelings Questionnaire (MMFQ). This analysis is described in section 5.6.3.
- Peer social networks (see Table 1 for details)

Details of all outcome measures and scoring methods are given in Table 1.

4.2.3 Mediating mechanisms

The following mediating mechanisms will be measured.

- Emotion perception
- Emotion regulation
- Interoceptive accuracy and attention
- Self-perception
- Self-ratings of social relationships (parent and peer attachments, bullying victimisation and loneliness)
- Peer social networks (including different indicators to those listed in Table 1)

The trial protocol describes in detail how these variables are measured. The mediation analyses are not described in this Statistical Analysis Plan.

4.3 Fidelity

Fidelity of the intervention is assessed by an observation form in sessions 1, 2, and 5 of the programme. Observation items measuring fidelity to intervention components are scored 0 (No), 0.5 (partially), and 1 (Yes). The sum of the scores will be calculated for every observed session and converted to a percentage score. The indicator of overall fidelity of the intervention is the average percentage score across groups in programme session 5.

4.4 Acceptability

Acceptability of the intervention will be assessed via interviews with a subset of participants (see Protocol). This is part of the process evaluation and not discussed in this Statistical Analysis Plan. We will also measure adherence as one indicator of acceptability (see section 5.3).

Table 1: Outcome measures and scoring

Outcome	Details and scoring
SDQ Total Difficulties (Goodman, 1997, 2001)	The SDQ Total Difficulties Score is calculated as the sum of the 20 items comprising the subscales of Emotional Problems, Conduct Problems, Peer Problems, and Hyperactivity/Inattention. Some items are reversed such that a higher score indicates more difficulty. Each item is rated on a 3-point scale (Not True, Somewhat True, Certainly True) and scored 0, 1, 2. The SDQ Total Difficulties Score ranges from 0 to 40. Individual missing items can be imputed using individual mean imputation by sub-scale, as long as a maximum of two items are missing in a particular subscale (https://www.sdqinfo.org/py/sdqinfo/c0.py). For example, if two items from the Emotional Problems subscale are missing for a participant, the mean of the remaining three items for that participant will be calculated and used to replace the two missing values. If any subscale has more than two items missing for an individual, the SDQ Total Difficulties Score will be considered missing for that individual at that time point.
Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS) (Tennant et al., 2007)	The WEMWBS is a 14-item scale with items rated on a 5-point scale from "none of the time" (1) to "all of the time" (5). A total score is obtained by summing the 14 item scores. The WEMWBS score ranges from 14 to 70. A higher score indicates higher well-being.
PHQ-8 (Kroenke et al., 2009)	The PHQ-8 is an 8-item measure of Depression. Items are rated on a 4-point scale from 0 ("not at all") to 3 ("nearly every day"). A total score is obtained by summing the eight item scores. The PHQ-8 score ranges from 0 to 24. In line with previous literature (Kroenke et al 2009), we will take a score of ≥10 as indicating depression for analyses of caseness.
GAD-7 (Spitzer et al., 2006)	The GAD-7 is a 7-item measure of Generalized Anxiety. Items are rated on a 4-point scale from 0 ("not at all") to 3 ("nearly every day"). A total score is obtained by summing the eight item scores. The GAD-7 score ranges from 0 to 21. In line with previous literature (Spitzer et al 2006), we will take a score of ≥10 as indicating generalized anxiety disorder for analyses of caseness.
Sleep phenotype	Sleep quality is measured by two items: INSFREQ: During the past month how often have you had difficulty falling asleep, staying asleep or have had a problem with waking too early? (7-point scale, from once a week to seven times a week) INSIMPAIR: If you have reported that you have had difficulties sleeping to what extent has this led to daytime impairment? (5-point scale, from Not at all to Very Much)

Outcome	Details and scoring
	Our secondary outcome is insomnia. Insomnia cases are defined as those that report both INSFREQ ≥ 3 and INSIMPAIR ≥ 3.
Alcohol use disorders identification test (AUDIT)	The AUDIT is a 10-item assessment of alcohol use. Items 1-8 are rated on 5-point response scales (scored 0-4), while items 9 and 10 are rated on 3-point response scales (scored 0, 2, 4). A total score is obtained by summing the ten item scores (https://auditscreen.org/about/scoring-audit/). The AUDIT score ranges from 0 to 40.
Drug use disorders identification test (DUDIT)	The DUDIT is an 11-item assessment of drug use. Items 1-9 are rated on 5-point response scales (scored 0-4), while items 10 and 11 are rated on 3-point response scales (scored 0, 2, 4). A total score is obtained by summing the eleven item scores (https://comorbidityguidelines.org.au/appendix-x-drug-use-disorders-identification-test-dudit/dudit-scoring-and-interpretation). The DUDIT score ranges from 0 to 44.
General psychopathology, internalizing symptoms, externalizing symptoms	These concepts will be measured via a confirmatory factor analysis as described in section 5.6.3.
Social network variables: indegree for best friend, indegree for likeability, outdegree for advice-seeking	Social network questions were administered to the whole year group, and there was no restriction on the number of same- or cross-sex peers that each participant could nominate. Our secondary outcomes from the network data are three count variables: *Indegree for best friend*: This is a count of all the nominations received by a participant in response to the peer nomination question "Who in your year group are your best friends?". *Indegree for likeability*: This is a count of all the nominations received by a participant in response to the peer nomination question "Which students in your year group do you like?". *Outdegree for advice-seeking*: This is a count of all the nominations sent by a participant in response to the peer nomination question "Who in your year group gives good advice to you when you're upset?".

5. Data analysis

Analyses will be carried out based on the intention to treat principle, comparing the groups as randomised regardless of compliance with the intervention. The primary analysis will be performed on observed outcome values (without imputation, except missing item imputation discussed below to enable us to calculate total scores). All statistical hypothesis

tests will be two-sided. Confidence intervals will be symmetric around the point estimate, using the 95 % level of confidence.

5.1 Recruitment and representativeness of recruited patients

A consort diagram will be constructed to describe the flow of subjects through the trial (http://www.consort-statement.org/). The diagram will detail the number of subjects: invited to participate; agreeing to enter the study (with reasons for refusal); receiving the intervention (with reason for not receiving this); followed up and withdrawn (with reasons).

5.2 Baseline characteristics

Baseline characteristics of the young people will be summarised by treatment group to gauge the balance in characteristics between the randomised groups. The results will be presented as means, standard deviations, medians and inter-quartile ranges for numeric variables; and frequencies and percentages for categorical variables. No statistical hypothesis testing will be used.

5.3 Adherence to treatment, attrition and missing data

Adherence. Adherence will be measured as the proportion of treatment sessions attended. The mean, range, and interquartile range of this proportion will be reported.

Attrition. Some loss to follow-up is expected over twelve months. Reasons for missing outcome data will be described and frequency (%) of subjects with missing data, by reason will be provided for each randomised group (and for each outcome).

5.4 Adverse event reporting

Adverse events (AE) and serious adverse events (SAE) will be summarised. A protocol addition will be written to define these.

5.5 Analysis of primary outcomes

Baseline and post-intervention scores on the WEMWBS and SDQ Total Difficulties will be summarized separately for the intervention and control groups using means, standard deviations, and quartiles.

The primary outcome analysis will use a partially clustered mixed effects model with heteroscedastic errors. This model uses a random effect for intervention groups to take into account that participants are clustered in treatment groups in the intervention arm, but not in the control arm. The model will adjust for baseline score of the outcome measure. We will additionally control for clustering within schools via a random effect, as well as for school year (cohort) and the timing of measurement (school term) via a fixed effects.

Controlling for school term accounts for potential seasonal variation in mental health, given that the intervention was implemented at different times in the year for different groups of students.

For each primary outcome separately, we will fit the following model (Candlish et al., 2018; Flight et al., 2016);

$$Y_{ijk} = \left(\beta_0 + u_{jk}T_i + u_k\right) + \beta_B Y_{B,ijk} + \beta_C C_{jk} + \beta_M M_{ijk} + \beta_T T_{ijk} + \epsilon_{ijk}T_{ijk} + r_{ijk}(1 - T_{ijk})$$

where

- Y_{ijk} is the outcome (WEMWBS or SDQ Total Difficulties Score) for student *i* in intervention group *j* and school *k*;
- T_{ijk} is the intervention indicator variable (coded 0 for controls, 1 for students in the intervention);
- u_{jk} is a random intercept for intervention group *j* at school *k*;
- u_k is a random intercept for school *k*;
- $i = 1, \dots, n_{jk}$ identifies individuals;
- $j = 0, 1, \dots, n_k$identifies the intervention groups ($n_k = 3$ in a typical school); *j* = 0 identifies the control participants in each school;
- $k = 1, \dots, n$ identifies the schools;
- $Y_{B,ijk}$ is the baseline score on the outcome variable;
- C_{jk} is the cohort (school year), the level of randomization (coded 0 for the younger cohort, 1 for the older cohort);
- M_{ijk} is the time at which the outcome was measured, coded as school terms (autumn, spring, or summer term);
- β_T is the parameter of interest, the adjusted mean difference in post-intervention outcome scores between trial arms;
- $r_{ijk} \sim N(0, \sigma_r^2)$ is an individual error term for the control participants
- $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$ is an individual error term for the intervention participants

We will use adjusted degrees of freedom (Kenward-Roger) and restricted maximum likelihood procedure (REML) for estimation, as recommended by Candlish et al (2018).

The evidence for a treatment effect will be evaluated via a two-sided t-test on the adjusted mean difference in outcome scores between trial arms (coefficient β_T in the mixed effects model), using a 2.5% level of significance. Analyses will be carried out in the R software (R Core Team, 2019) and Stata version 18 (or later).

The primary endpoint will be post-intervention. Analyses will compare groups defined by intention-to-treat and include all those with available outcome data. Analogous analyses will investigate the evidence for a treatment effect at twelve-month follow-up.

5.6 Analysis of secondary outcomes

The results for the secondary outcomes will be presented as estimates with 95% confidence intervals. P-values will not be reported. Analyses will compare groups defined by intention to treat and include all those with available data. Analyses for post-intervention outcomes will be conducted alongside the primary analysis. Analyses for twelve-month follow-up data will be conducted when data will have become available.

5.6.1 Depression and Anxiety

Depression (PHQ-8) and Generalized Anxiety (GAD-7) will be analysed as continuous outcomes using analogous partial mixed effects models as for the primary analysis. Caseness will be defined according to established cut-off points (PQH-8 $\geq$ 10, GAD-7 $\geq$ 10). Mixed effects Poisson regression will be conducted to estimate the ratio of risks of depression and generalized anxiety, respectively, between the randomized groups. The model will allow for partial clustering, analogous to the linear model for continuous outcomes.

5.6.2 Insomnia (sleep phenotype)

The variables INSFREQ (frequency of insomnia symptoms) and INSIMPAIR (degree of impairment from insomnia symptoms) will be described separately for the intervention and control groups. Mixed effects Poisson regression will be conducted to estimate the ratio of risks of insomnia between the randomized groups. The model will allow for partial clustering.

5.6.3 General psychopathology, internalizing and externalizing problems

We will use confirmatory factor analysis to develop measures of general psychopathology ("p-factor"), internalizing problems, and externalizing problems. Following Patalay et al (2015), we will use a set of 25 observed variables:

- The five items from the SDQ emotional problems sub-scale
- The five items from the SDQ conduct problems sub-scale
- The six items from the Me and My Feelings Questionnaire (MMF, formerly known as "Me and My School") behavioural difficulties subscale
- Nine items from the 10-item MMF emotional difficulties subscale (the item "I am shy" will not be used, following Patalay et al 2015: p. 16).

Confirmatory factor analysis will be estimated in the R, Stata, or the Mplus software. Observed variables will be treated as categorical. The factor analysis will estimate a bifactor model as presented in Patalay et al. (71). This model yields three factors: a general

psychopathology factor (“p-factor”), and two specific factors measuring internalizing problems and externalizing problems, respectively. We will assess the fit of this model to our data and estimate an amended model if necessary for estimation convergence or goodness of fit. This process will be conducted before any comparisons between the treatment and control groups will be made with regard to the resulting factors, and the process will be documented.

We will estimate this model using baseline and post-intervention data, estimating a two-level factor analysis to account for dependency between observations within the same student. Factor scores from the bifactor model will then be saved and used in secondary outcome analysis. This yields three variables, each measured at baseline and post-intervention:

- General Psychopathology
- Internalizing Problems
- Externalizing Problems

These will be analysed separately using partially clustered linear mixed effects models, as for the primary analysis.

If the multilevel confirmatory factor analysis cannot be estimated (for example, due to failure to converge) we will try the following simpler models in turn:

(1) Instead of a multilevel model, we would fit a multi-group factor analysis (with baseline and follow-up time points defined as separate groups) and loadings constrained to be equal across groups
(2) If the model described under (1) also cannot be estimated, we would fit the model described under (1) but treat the observed variables as interval scaled, instead of as ordinal.

5.6.4 Alcohol and drug use

The distributions of the AUDIT and DUDIT measures will be described via means and quartiles separately for the intervention and control groups. As these variables are expected to feature a large proportion of zeroes (indicating no alcohol use and no drug use, respectively) in our data (Sfendla et al., 2022), the secondary outcomes will be binary variables indicating any alcohol use and any drug use, respectively (ie contrasting the non-zeroes with the zeroes). Mixed effects Poisson regression will be conducted to estimate the ratio of risks of alcohol use and drug use, respectively, between the randomized groups. The model will allow for partial clustering.

5.6.5 Social network variables: best friend nominations, likeability nominations, and advice-seeking

The distribution of the social network variables will be described via means and quartiles separately for the intervention and control groups. Mixed effects generalized Poisson regression will be conducted to estimate the ratio of indegree for best friend, indegree for likeability, and outdegree for advice-seeking. A random effect will be fitted to each individual to account for overdispersion.

5.7 Missing items in scales and subscales of primary and secondary outcomes

If individual items in scales or subscales are missing, we will proceed as follows:

- We will follow official guidance for those measures where it exists (see section 4.2.2).
- Where no guidance exists, we will use individual mean imputation if 20% or fewer items are missing for an individual in a questionnaire. For example, in a scale with 10 items, individual mean imputation will be applied to individuals with up to 2 items missing. The average value for the complete items will be calculated for that individual and used to replace the missing values. The scale score will be calculated based on the complete values and these replacements.
- For the confirmatory factor analysis (section 5.6.3), we will use Full Information Maximum Likelihood. No imputation will be performed.

5.8 Sensitivity and other planned analyses

Sensitivity analyses for missing outcome data

Under the assumption that data are Missing At Random (MAR), two approaches will be taken for the primary outcome:

1) We will refit models to obtain estimates adjusted for variables associated with missingness. To identify predictors of missing data, characteristics of participants with and without missing outcome data will be compared using logistic regression models (with missing yes/no as the outcome). The main analysis model will be refitted to adjust for any characteristics found to be associated with missingness and the outcome of interest.
2) We will use multiple imputation methods. The imputation model will include the outcome of interest, socio-demographic baseline data and any other variables possibly related to missingness and the outcome. The imputations will be performed by study arm. We will use the number of imputations that is around the proportion of missingness (e.g. 20 imputed sets for 20% missing data) and combine the result using Rubin's rules.

Supportive analyses

For the primary and secondary outcomes using the same modelling approaches as described previously, the treatment effect will be estimated adjusting for any concerning imbalances in baseline characteristics.

Analysis of repeated measurements of outcome

Once data from the 12-month follow-up are available, we will use a mixed effects model based on all participant outcome data over 12 months to investigate how the primary and secondary outcomes change over time. The model will include time (twelve-months follow up vs end of treatment) as a categorical variable as well as an interaction of time with the intervention indicator variable. Based on this model we will obtain an estimate of the difference between groups at both 6 and 12 months under MAR assumptions.

Per protocol analysis

A per-protocol analysis will be conducted for the primary outcomes, using the same statistical models as specified above. The per-protocol analysis will include all control group participants, but include only those intervention group participants who have attended at least five of their intervention group sessions.

5.9 Model checking

The linear mixed effects models assume that the residuals are normally distributed and homoscedastic. This will be checked using residual plots. Since our outcome measures are scales with fixed minima and maxima, outliers with high leverage are unlikely to occur, but we will check for influential observations also. If substantial departures from normality or homoscedasticity occur, additional sensitivity analyses using suitable transformations of the relevant outcome variables will be considered.

5.10 Convergence issues

If the partially clustered mixed effects models do not converge, we will attempt the following solutions (in order): replace the random effect for school with fixed effects; remove the “timing of measurement” covariate; constrain the model to homogeneous standard errors; estimate a fixed effects model and use bootstrapping to estimate standard errors.

References

Candlish, J., Teare, M. D., Dimairo, M., Flight, L., Mandefield, L., & Walters, S. J. (2018). Appropriate statistical methods for analysing partially nested randomised controlled trials with continuous outcomes: A simulation study. *BMC Medical Research Methodology*, *18*(1). https://doi.org/10.1186/s12874-018-0559-x

Flight, L., Allison, A., Dimairo, M., Lee, E., Mandefield, L., & Walters, S. J. (2016). Recommendations for the analysis of individually randomised controlled trials with clustering in one arm - A case of continuous outcomes. *BMC Medical Research Methodology*, *16*(1). https://doi.org/10.1186/s12874-016-0249-5

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *38*(5), 581–586. http://www.ncbi.nlm.nih.gov/pubmed/9255702

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry*, *40*(11), 1337–1345. https://doi.org/10.1097/00004583-200111000-00015

Kroenke, K., Strine, T. W., Spitzer, R. L., Williams, J. B. W., Berry, J. T., & Mokdad, A. H. (2009). The PHQ-8 as a measure of current depression in the general population. *Journal of Affective Disorders*, *114*(1–3), 163–173. https://doi.org/10.1016/j.jad.2008.06.026

Patalay, P., Fonagy, P., Deighton, J., Belsky, J., Vostanis, P., & Wolpert, M. (2015). A general psychopathology factor in early adolescence. *British Journal of Psychiatry*, *207*(1), 15–22. https://doi.org/10.1192/bjp.bp.114.149591

R Core Team. (2019). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Sfendla, A., Bador, K., Paganelli, M., & Kerekes, N. (2022). Swedish High School Students' Drug and Alcohol Use Habits throughout 2020. *International Journal of Environmental Research and Public Health*, *19*(24). https://doi.org/10.3390/ijerph192416928

Spitzer, R., Kroenke, K., Williams, J., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder. *Archives of Internal Medicine*, *166*, 1092–1097. https://seaep.es/wp-content/uploads/2019/11/2.3.1.2CUESTIONARIO-BREVE-DE-EVALUACIÓN-DEL-TRASTORNO-DE-AN-SIEDAD-GENERALIZADA-GAD-7.docx

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., Parkinson, J., Secker, J., & Stewart-Brown, S. (2007). The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. *Health and Quality of Life Outcomes*, *5*(1), 63. https://doi.org/10.1186/1477-7525-5-63

Viding, E., Lloyd, A., Law, R., Martin, P., Lucas, L., Wu, T. C. H., Steinbeis, N., Midgley, N., Veenstra, R., Smith, J., Ly, L., Bird, G., Murphy, J., Plans, D., Munafo, M., Penton-Voak, I., Deighton, J., Richards, K., Richards, M., & Fearon, P. (2024). Trial protocol for the Building Resilience through Socio-Emotional Training (ReSET) programme: a cluster randomised controlled trial of a new transdiagnostic preventative intervention for adolescents. *Trials*, *25*(1), 1–21. https://doi.org/10.1186/s13063-024-07931-2