Measuring and treating negative symptoms: developments and challenges

Dr Nicola McGuire, MA Hons, MSc, PhD

Submitted in partial fulfilment of the requirements for the degree of

Doctorate in Clinical Psychology

School of Health and Wellbeing

College of Medical, Veterinary and Life Sciences

University of Glasgow

November 2025

The University of Glasgow Plagiarism Statement

The University of Glasgow Plagiarism Statement can be found in the *University Regulations* at

https://www.gla.ac.uk/myglasgow/studentconduct/plagiarism/

This should be read in conjunction with the discipline specific guidance provided in the Doctorate in Clinical Psychology Handbook.

If you are still unsure or unclear about what plagiarism is or need advice on how to avoid it,

SEEK HELP NOW!

You can contact any one of the following for assistance:

Module Coordinator

Research Supervisor

Student Learning Development Service

Remember to also read your Course or Programme Handbook for advice on good academic practice.

Table of contents

List of tables

List of figures

Acknowledgements

First, I would like to thank my supervisor, Professor Hamish McLeod, for his unwavering support, persistence and compassion during this process. There have been many ups and downs over the course of this research and your guidance and motivation have been instrumental to helping me see this project through to completion. I would also like to thank Professor Andrew Gumley, who contributed his views and positivity to an early version of this project. Thank you to Professor Jon Evans, my research advisor, for your patience and reflections in navigating this complex and interesting work.

Thanks to Paul Cannon, for your input to the systematic search component of this thesis. Thank you to Dr Natasha Hill, my second reviewer and friend, for your support and commitment to this project even after training. Thank you to Suzanne Syrett and Stephanie Allen, for your invaluable contributions to the patient-facing documents in the MRP project. Thank you also to Dr Rachel Whyte, your contributions to data collection made the completion of this project feasible and I am so grateful we got to work together.

A huge thank you to my study participants who gave their time, energy and commitment to participate in this research. It was a privilege to meet you all and I hope that I have been able to reflect the extent of your contributions accurately. Thank you to my field supervisor, Dr Ian-Mark Kevan, for your continued support and site knowledge in an adaptable context. Thank you also to Dr Danielle Graham for your willingness to participate in this project. Thank you to the NHS nursing staff who facilitated data collection and contributed to key study measures. Thank you to Ilanit Hassan-Ohayon and Paul Lysaker, who unfortunately did not get to see the outcome of this project, for providing training relevant to the MRP and whose therapeutic philosophy significantly influenced this project.

Finally, I couldn't have done this (again) without the help of my amazing and resilient little family: Patrick, mum, dad, and Olivia. We have faced trials and tribulations to get to this point; and I have finished it if it wasn't for the hours of love, childcare, and attendance to my basic human needs that you have given me. I promise not to put you through this again!

Chapter One

Measurement Equivalence of the Clinical Assessment Interview for Negative Symptoms (CAINS) Items Across Cultures: A Systematic Review and Meta-analysis

Prepared in accordance with the author requirements for Journal of Psychopathology and Behavioral Assessment (https://link.springer.com/journal/10862/submission-guidelines?utm_source=slink&utm_medium=journal_finder)

Nicola McGuire[1]*

School of Health and Wellbeing, University of Glasgow, Scotland
[1]email correspondence to: n.mcguire.1@research.gla.ac.uk
School of Health and Wellbeing,
College of Medical, Veterinary and Life Sciences
University of Glasgow
Clarice Pears Building
90 Byres Road
Glasgow, Scotland
G12 8TB
ORCiD: 0000-0001-5332-7259

Statements and declarations

The primary author was responsible for project administration, study conception and design, development of study-specific resources, data curation and analysis, visualization and manuscript drafting, review and editing. NH and HM contributed to the execution of the research methodology through second rater and arbitrator responsibilities respectively. HM also provided supervision for the project and review of the manuscript.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The primary author is funded in her position, which

is mandated to include developing research outputs, by NHS Education for Scotland (NES) in fulfilment of doctoral training for the Doctorate in Clinical Psychology programme. NES played no role in the research design, conduct or write up processes.

The authors have no relevant financial or non-financial interests to disclose.

As the manuscript only uses aggregate reports of participant data from secondary sources, ethical approval is not required, as confirmed by NHS Health Research Authority (HRA) online decision tool.

Data underlying the generation of this manuscript (i.e. consensus decision-making and risk of bias assessment data) and the software code used in statistical analyses are available in the Open Science Framework at the following URL: https://doi.org/10.17605/OSF.IO/M9B4S. Data extraction and search results are not publicly available due to density of information and their replicability from reported methods, but are available upon request.

Abstract

Negative symptoms, encapsulating deficits in motivation, pleasure and expressivity, are multifaceted sources of symptom burden requiring effective operationalisation to develop meaningful treatment targets. This review aimed to summarise and explore relationships between the characteristics, psychometric properties, and factor structure of the Clinical Assessment Interview for Negative Symptoms (CAINS) in studies validating the scale cross culturally. Studies were identified via PsycINFO, MEDLINE, EMBASE, the Cochrane Library, grey literature and hand-searching techniques. Data were proportionately dual screened and rated for risk of bias using the Joanna Briggs Institute Critical Appraisal Checklist for Analytical Cross-Sectional studies. Complete-case cross-sectional analyses comprised the majority of the 24 included studies, with a relatively homogenous sample from outpatient psychiatric settings. Meta-analytic methods suggested studies converged on two- and five-factor structure models mapping onto the CAINS interview structure and expert-consensus domains for negative symptoms, with acceptable to excellent fit psychometric properties. No immediate cultural trends were identified although European and Asian samples were sources of variability. The robust results may be reassuring for researchers utilising this measure in theoretically-informed research, however confidence in the evidence would be improved by clearer reporting methods, and analyses with a wider range of clinical samples and indices of cultural differences.

Keywords

Cross-cultural comparison; psychometrics; psychotic disorders; schizophrenia; anhedonia; apathy

1.1. Introduction

Negative symptoms including reduced engagement in goal-directed activity, social engagement and experiences of pleasure, and decreased use of expressive gestures and vocalisation (Galderisi et al., 2018) are a significant component of schizophrenia spectrum diagnoses (Bleuler, 1911). They contribute to high symptom burden and impairments in functioning more than other psychosis symptoms (Abplanalp et al., 2022). Negative symptoms also affect people with other mental health difficulties such as depression (Guessoum et al., 2020). Negative symptoms which are persistent, and unlikely to change due to alleviation of other difficulties (e.g. positive psychosis symptoms, medication side-effects), have less directly targeted treatment options available in clinical practice (Giordano et al., 2022).

Strauss and Cohen (2017) have argued that difficulties in operationalisation have hampered mechanistic understanding of negative symptoms necessary for treatment. The National Institute of Mental Health (NIMH) consensus statement (Kirkpatrick et al., 2006) outlines the five domains thought to best capture negative symptoms (which are often grouped into two categories; summarised in Table 1.1). The best-fitting negative symptoms factor structure, and relative importance of contributory items in terms of functional outcomes and propensity for treatment is a contested subject (Giordano et al., 2023). The European Psychiatric Association (EPA) proposed that “gold standard negative symptom measures” should consist of items measuring each of these domains. Prominent first-generation negative symptom measures (the Positive and Negative Syndrome Scale (Kay et al., 1987), and the Scale for the Assessment of Negative Symptoms, (Andreasen, 1989)) were downgraded due to failure to quantify subjective experience in some domains and inclusion of items considered extraneous to the negative symptoms construct (Galderisi et al., 2021).

Table 1.1: Description of negative symptom domains

Category	Domain	Type of reduced functioning
Experiential deficits	Avolition	Engagement in goal directed activities due to diminished motivation
	Anhedonia	Experiences of pleasure, during and when anticipating events
	Asociality	Engagement in social activity due to reduced motivation for social relationships
Expressive deficits	Blunted Affect	Use of facial and vocal expression and gestures
	Alogia	Quantity of speech and spontaneous elaboration

The EPA proposed that, the Brief Negative Symptoms Scale (BNSS; Kirkpatrick et al., 2011) and the Clinical Assessment Interview for Negative Symptoms (Kring et al., 2013) best represent these domains, have been extensively validated, adapted for use cross-culturally, and show promising psychometric properties (Wehr et al., 2024; Weigel et al., 2023). However, research demonstrates that aspects of environmental milieu (e.g. mass-media influences, laws and politics) interact with experience of negative symptoms (Strauss, 2024). It is therefore possible that negative symptoms may vary cross-culturally. A review of the BNSS has explored the factor structure and psychometric properties reported cross-culturally (Tatsumi et al., 2020), but no such review exists for the CAINS despite additional versions of the CAINS being published in several regions and cultures since the last known systematic review (Wehr et al., 2024).

This review will first, systematically identify and summarise the characteristics of studies exploring the factor structure of the CAINS, including loading of individual items, overall fit statistics and psychometric properties, and risk of bias will be assessed. This will help characterise how the CAINS has been employed globally and its measurement invariance

cross-culturally. Second, pooled estimates of psychometric properties of the CAINS will be derived and convergence of factor structures on any common models will be identified through meta-analyses. This will support determining the cross-cultural applicability of the CAINS as well its reliability in assessing negative symptoms.

1.1.1 Review questions

1 In research designed to examine the putative factor structure of the CAINS what are the sample, methodological and analytical characteristics of each study?

2 In studies following the original CAINS validation study, what is the convergence of factor structures?

- **2.1** What is the variance in internal consistency for factor structures of the CAINS across studies?
- **2.2** How do the item loadings in these studies compare to the original CAINS validation study?

3 What are the sources of risk of bias in the factor structure of the CAINS seen in studies following the original validation study?

- **3.1** What is the variance in reliability indices (i.e. test-retest, inter-rater) in these studies and how do these compare to the original CAINS validation study?
- **3.2** What associations exist between reliability indices and the cultural characteristics of each sample and factor loadings within individual studies?

While it was expected that variance in findings would be observed across a range of cultures, there was no theoretical basis for predicting the direction of variance in risk of bias, or psychometric properties across studies.

1.2. Methods

1.2.1 Protocol and registration

Methods were pre-registered using the Non-Interventional Reproducible and Open Systematic Reviews (NIRO-SR) template (https://osf.io/m9b4s/). The protocol and final report follow Cochrane Handbook for Systematic Reviews (Higgins et al., 2022), NIRO-SR (Topor et al., 2023) and Preferred Reporting Items for Systematic review and Meta-Analysis (PRISMA; Page et al., 2021) guidelines (see Appendix 1.1 for PRISMA checklist).

1.2.2 Screening and eligibility criteria

Studies were screened against the following criteria:

Population:

Participants aged 16 or older who experienced negative symptoms categorised by a reliable assessment.

Context:

Statistical analyses exploring the CAINS factor structure and reporting study methodology and findings.

Outcomes:

Reporting the CAINS factor structure and loading of individual items was mandatory. The CAINS is a 13-item interview and rating scale for negative symptoms (possible score range 0-40). Each item (summarised in Appendix 1.2) is rated on a Likert scale between 0-4 representing no- to severe- deficits.

Secondary outcomes included:

Inter-rater reliability
Test-retest reliability
Internal Consistency (Cronbach's alpha)
Factor model fit indices (e.g. RMSEA, X^2, CFI, AIC)

Study type:

Only quantitative studies exploring validation of the CAINS factor were included. Case studies and qualitative research were excluded.

Only English language publications were included.

1.2.3 Search strategy

The search strategy was developed following a scoping search and initial review of the literature via Google Scholar, and was approved by a University of Glasgow librarian (P.C.). In an aim to ensure sufficient specificity in the search strategy without

compromising sensitivity, the search focused solely on studies using the term “CAINS” or “Clinical Assessment Interview for Negative Symptoms” in their title or abstract. Bibliographic databases included PsycINFO (via EBSCOhost); MEDLINE and EMBASE (both via Ovid); and the Cochrane Database of Systematic Reviews and Cochrane Central Register of Controlled Trials (via Wiley online library). Grey literature including Google Scholar, EThOS and ProQuest Dissertations & Theses Global were also searched. Forward and backward citation searching was conducted, including a sensitivity check of the CAINS original article to identify any missing studies. CAINS authors were contacted to identify any special issues, conference proceedings or unpublished research and did not highlight any additional sources. The publication period was restricted to 2013 when the CAINS final development and validation study was published. The search was updated prior to final analysis (18/09/2023 and 26/05/2025 respectively).

1.2.4 Study selection

Endnote was used for de-duplication prior to screening, whereafter records were extracted to Excel to record screening decisions and reasoning for each reviewer (accessible online: file entitled Consensus Checking Workbook - https://doi.org/10.17605/OSF.IO/M9B4S). The primary reviewer completed 100% of stage one (simultaneous title and abstract screening) and stage two (full-text review) screening. The second reviewer screened a randomly selected 10% of records per stage, using the systematic review protocol for orientation. Calibration rates were established following an initial pilot of 25 records in stage one, and five records in stage two, blind to the other rater’s responses. Initial levels of agreement for stage one were moderate (κ = 0.448), and following consensus discussion increased to almost perfect agreement in the subsequent round (κ = 0.834). The overall agreement for stage one was substantial (κ = 0.654). There were no disagreements during stage two screening.

1.2.5 Data extraction

Data extraction was completed by the primary reviewer with use of a pre-specified excel template (accessible online: file entitled Data Extraction Template - https://doi.org/10.17605/OSF.IO/M9B4S). Items included relevant methodological information (i.e. sampling, administration of measures, and statistical analyses), and

demographic and psychometric data. Authors were contacted a maximum of three times to attempt to obtain missing data.

1.2.6 Data management and ethical considerations

This study is classified as research under NHS Health Research Authority (HRA) guidelines due to the intention to answer research questions and find new, generalisable findings. However, as only aggregate reports of participant data from secondary sources are used, this review is not subject to governance processes required for protection of individual anonymity, nor is ethical approval required. The categorisation of the review as research, and data management and ethical considerations were confirmed using the HRA online decision tools (https://www.hra-decisiontools.org.uk/). The meta-data resulting from the systematic review methods were stored on a secure University of Glasgow server as per the University of Glasgow Records Management Policy (https://www.gla.ac.uk/media/Media_911649_smxx.pdf).

1.2.7. Analysis

Review syntheses were conducted by the primary researcher. However, the second and third reviewers (responsible for risk of bias assessment and discrepancy arbitration respectively) did review the manuscript to comment on any indicators of error or factors which might influence the findings.

Narrative synthesis

Informed by a summary of narrative synthesis methods (Popay et al., 2006) to address all research questions, participant demographics, mean levels of negative symptoms, and methods and convergence of factor analyses results were synthesized descriptively, using tables and figures where appropriate. A critical review of findings in comparison to qualitative characteristics of individual studies, similarities and differences between studies, and comparisons to the original CAINS study were derived. The robustness of any inferences made was discussed.

Risk of bias assessment

Risk of bias assessment (accessible online: file entitled Risk of Bias reports - https://doi.org/10.17605/OSF.IO/M9B4S) addressed question three, using the Joanna Briggs Institute Critical Appraisal Checklist for Analytical Cross-Sectional studies (Joanna Briggs Institute, 2017). This tool investigates clarity of inclusion criteria and sampling information, validity and reliability of data gathered, influence of any confounders, and appropriateness of statistical analysis used. Considerations based on COSMIN items (Mokkink et al., 2016) supplemented the tool criteria to help differentiate ratings (e.g. "interval/conditions between repeated measures" was used to rate item three: was the exposure measured in a valid and reliable way). The overall appraisal was planned to justify any sensitivity meta-analyses. One hundred percent of the included studies were assessed by the primary reviewer and 20% of the included studies were selected using a randomisation generator in Excel and assessed independently by the second reviewer (NH). Raters had 83.3% agreement across all rounds of assessment and were able to resolve discrepancies without inclusion of an arbitrator.

Quantitative Synthesis

The following meta-analytic methods were used for each research question:

1 Pooled estimates, weighted by sample size were computed for the demographic and symptom specific characteristics of the included samples. These were estimated in random effects meta-analyses given the assumption of likely significant variance in the between study estimates across datasets.

2.1 A pooled estimate of Cronbach's alpha (Cronbach, 1951) was computed in a random-effects meta-analysis weighted by sample size, as this was anticipated to be the most commonly available measure of internal consistency. Several meta-analytic methods have been proposed to pool Cronbach's alpha. Analyses parameters selected were based on findings by López-López et al. (2013). For ease of interpretation, the untransformed alpha coefficients were selected as the primary method of computation, with Hakstian and Whalen (1976) and (Bonett, 2002) transformations conducted as sensitivity analyses to ensure no significant differences accounted for by computation method. The Knapp and Hartung (2003)

correction was applied to test significance of regression coefficients produced by the meta-analytic model.

3.1 A pooled estimate of inter-rater reliability calculated by the Intra-Class Correlation (ICC) coefficient (Shrout & Fleiss, 1979) was calculated using methods detailed in (Noble et al., 2019). Planned pooled estimate of test-retest data was not calculated due to insufficient data available.

Restricted Maximum Likelihood was used to compute between study variance for all meta-analyses, although as sensitivity analyses an alternative method (Dersimmonan-Laird estimation) was also computed, with negligible differences. For each meta-analysis, funnel plot asymmetry was graphed and assessed using Egger's regression test (Egger et al., 1997). The magnitude of bias was also statistically assessed by comparing a bias-corrected estimate of the results using the Duval and Tweedie Trim and Fill method (Duval & Tweedie, 2000). I^2 was used to assess the proportion of sampling error as opposed to true variation in the outcome as described by (Borenstein, 2020) and a prediction interval was computed to give a sense of heterogeneity. These analyses were conducted in R version 4.5.0 using the metafor package (Viechtbauer, 2010) and the code is in Appendix 1.3.

Additionally, due to the sufficiency of data a post-hoc decision was made to compute a meta-analytic exploratory factor analysis using the co-occurrence matrices from each study. Following rationale provided in Escaffi-Schwarz et al. (2025), given that there were more than four possible factors reported across CAINS validation studies, the methods reported by Shafer (2005) were utilised using the coefa package in R to perform the analysis. This allowed exploration of whether the factor structures reported across studies converged on a common model.

1.3. Results

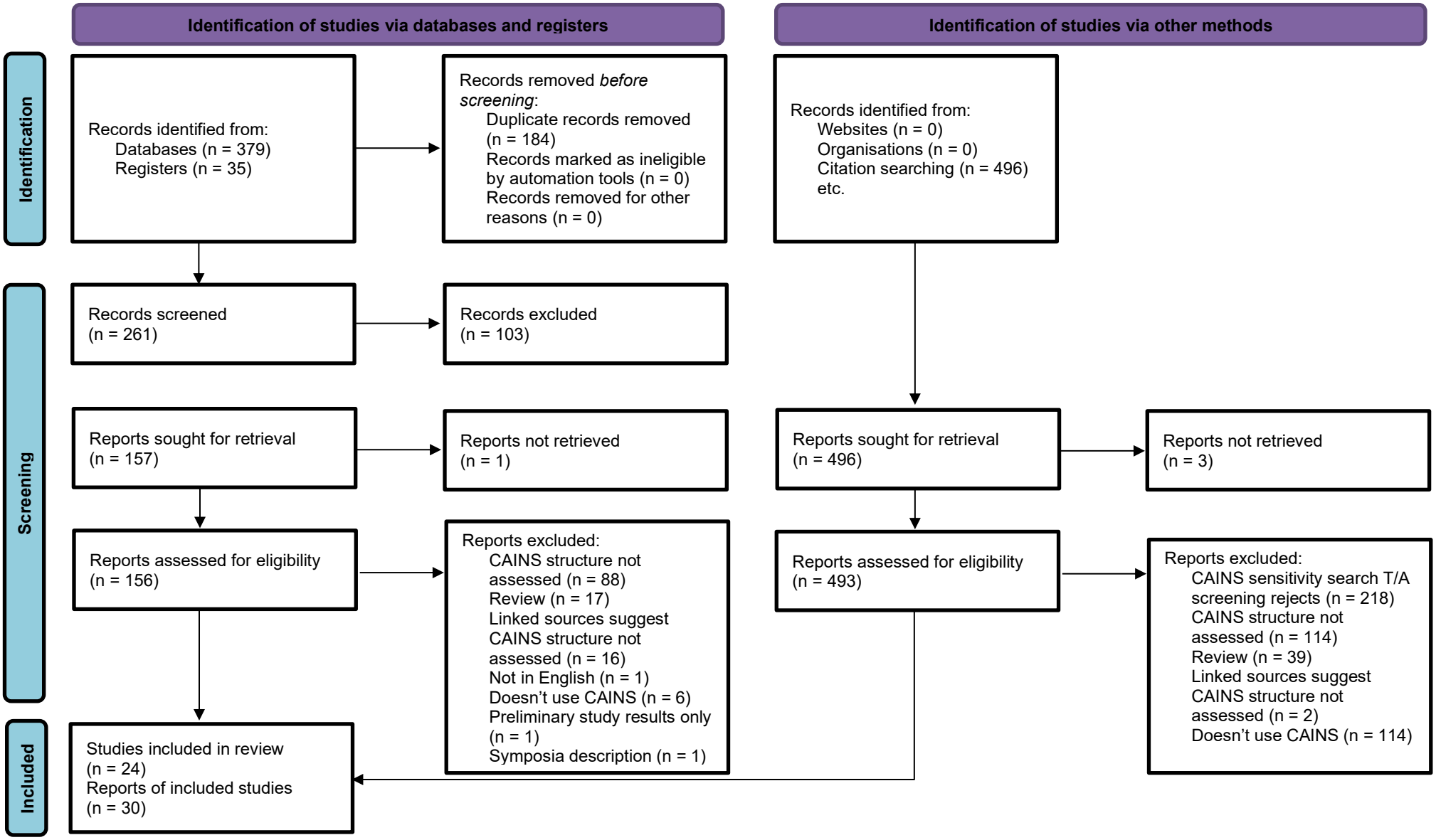
1.3.1 Search results

Figure 1.1 illustrates the search results and screening processes that revealed 24 studies discussed in this section. Two further CAINS validations were identified (one via database searching and one via a grey literature search) which are not included as there were no

reports representing their findings in English (Stevović Injac et al., 2019; Федотов et al., 2024). The original CAINS authors identified no further studies when contacted.

Reports of the same studies were identified based upon matching information across records, and often reflected the publication of works in a thesis or conference abstract prior to a full academic journal publication. As these revealed no necessary additional data and showed no inconsistencies in reporting, they are not discussed further (they are listed in Appendix 1.4). Two (Ahmed et al., 2022; Russo et al., 2021) of the 24 included reports were confirmed to have partially overlapping data with previous publications from the same research groups, therefore for relevant analyses a sensitivity check was planned with these removed.

Fig. 1.1 PRISMA (2020) Flow diagram of identified reports



1.3.2 Study characteristics

Most of the studies included in the review were cross-sectional, complete case analyses. Table 1.2 summarises sample and analytical characteristics of each report as well as CAINS adaptations and the identified factor structure. Most studies had explicit aims either of validating the original (ID: 3, 14, 18, 19), or a translated version of the CAINS (ID: 2, 4, 5, 6, 8, 9, 10, 11, 15, 16, 21, 22, 23, 24), with one report using combined data from several CAINS translations (Russo et al., 2021). Li et al. (2025) aimed to investigate the latent structure of the CAINS and another negative symptom measure in bipolar and major depressive disorder samples to assess transdiagnostic applicability. Li et al. (2022) and Ahmed et al. (2022) reported primary aims to investigate the latent structure of negative symptoms with the CAINS included amongst other covariates. Although Ahmed et al. (2022) reports the best fitting CAINS factor structure, only data related to models with additional latent variables is given therefore it is not discussed further here. Thomson (2019) had primary aims related to exploring relationships between attachment, mentalisation and negative symptoms.

Table 1.2: Included study characteristics

Study	ID	Location	Final sample size	CAINS application	Factor analysis	Factor Structure
Ahmed et al. (2022)	1	USA	178	Assume original CAINS	CFA of one-, two-, five- and hierarchical- factor solutions	Five- and hierarchical-factor
Bengtsson et al. (2022)	2	Sweden	34	Swedish translation and back-translation	Two-factor CFA	Two-factor
Blanchard et al. (2017)	3	USA	501	Original CAINS	EFA: principal axis extraction Promax rotation	Two-factor
Chan et al. (2015)	4	China	68	Mandarin and Cantonese translation (Chinese research assistant) Independent backward translation Follows Beaton et al. (2000) guidelines Revised through discussion Original authors' training materials	PCA	Two-factor
Cuesta et al. (2021)	5	Spain	98	Assume Spanish CAINS (Valiente-Gómez et al., 2015)	EFA: principal axis extraction Oblimin rotation	Two-factor
Engel et al. (2014)	6	Germany	53	German translation following Beaton et al. (2000) guidelines (including manual) Independent backward translation (bilingual research assistant) Compared with original and revised Discussed by staff panel Original training materials used in training	EFA: principal axis extraction Promax rotation	Two-factor

Hosáková et al. (2017)	7	Czech Republic	100	Czech version	CFA: WLSMV estimation	Two-factor
Jang et al. (2017)	8	South Korea	180	Two Korean translations (including manual) harmonised by clinically experienced colleagues Independent back-translation Compared to English version Use of training materials (unclear whether original)	EFA: principal factoring Promax rotation	Two-factor
Jung et al. (2016)	9	South Korea	119	Two Korean translations following Koller et al., (2007) guidelines (Korean speakers) Consensus through discussion Independent back translation (English speakers) Compared with original Piloted (10 people)	CFA Structural Equation Modelling to assess model fit	Two-factor
Kulenović et al. (2022)	10	Bosnia and Herzegovina	81	Language not specified Translation and independent back-translation Original author manual	PCA: Promax rotation Two-factor (original scale) Four-factor (categorisation of poor-fitting items in two-factor solution)	Two- and four-factor

Laraki et al. (2022)	11	France	84	Two independent French translations (including manual): independent expert to review/resolve discrepancies Back-translation (bilingual expert) Revised through discussion Expert panel comparison: back translation and English version (conceptual equivalence) Harmonized version piloted (10 participants - understandable by target population) Finalised, proofread and corrected remaining errors	EFA: principal axis factoring Oblimin rotation Scree plot and parallel analysis indicated a fixed two-factor criterion	Two-factor
Li et al. (2022)	12	China	305	Chinese CAINS (Chan et al., 2015)	CFA: one-, two-, five-, and hierarchical- factor solutions Use of model modification indices for residual correlations to improve model fit	Two-factor
Li et al. (2025)	13	China	331	Chinese version	CFA: one-, two-, five-, and hierarchical- factor solutions (for bipolar-, major depressive- disorder, and combined samples). Latent variables were treated as categorical WLSMV estimation Model modification indices larger than 10 were utilised to improve model fit (excepting the one-factor model).	Five-factor

Rekhi et al. (2019)	14	Singapore	274	Original CAINS	CFA: unacceptable fit statistics EFA (random half of sample): WLSMV estimation oblique rotation Validation CFA (remaining sample).	Four-factor
Richter et al. (2019)	15	Germany	105	German CAINS (Engel et al., 2014)	CFA: one and two-factor model (with modification indices for shared covariance in the MAP subscale) Sample size calculation (based on published guidelines) ML estimation Bollen-Stine bootstrapping: account for multivariate non-normality.	Two-factor
Ristić et al. (2020)	16	Serbia	67	Serbian translation by research assistant (including manual) Back-translated by independent translator	PCA: Promax rotation Parallel analysis to evaluate	Three-factor
Russo et al. (2021)	17	UK and South-Eastern Europe	257	Original and translated CAINS	EFA: principal axis factoring Promax oblique rotation (calibration sample) CFA: ML estimator of EFA one- two- and five- factor solutions. Sensitivity analysis: removed participants categorised with schizoaffective disorder	Data-driven five-factor model
Stoilkovska et al. (2021)	18	Northern Macedonia	82	Participant's answers (assume non-English) evaluated against original CAINS manual	PCA: varimax rotation (fixed two-factor criterion for parsimony)	Two-factor

Strauss et al. (2018)	19	USA	400	Original CAINS	CFA: one-, two-, five-, and hierarchical- factor solutions	Five-factor hierarchical model
Thomson (2019)	20	UK	66	Original CAINS	CFA: data log-transformed for positive skew	Two-factor
Uka et al. (2020)	21	Albania	106	Albanian forward and back translation according to Brislin (1970) guidelines Back-translated version compared to original to resolve discrepancies	PCA	Four-factor
Valiente-Gómez et al. (2015)	22	Spain	100	Two Spanish translations, including manual (researchers fluent in English) Consensus discussed by panel of clinically experienced colleagues Back-translated and compared to original Training materials used, assume original	EFA: principal axis extraction Promax rotation	Two-factor
Vayisoğlu et al. (2024)	23	Turkey	79	Turkish version	CFA and EFA: principal components method Varimax rotation.	Two-factor
Xie et al. (2018)	24	China	185	Chinese CAINS (Chan et al., 2015)	CFA: unitary, original and Chinese factor structure	Two-factor

CFA – Confirmatory Factor Analysis; **EFA** – Exploratory Factor Analysis; **MAP** – Motivation and Pleasure subscale of the CAINS; **ML** – Maximum Likelihood; **PCA** – Principal Components Analysis; **WLSMV** – diagonally weighted least squares

1.3.3 Demographics summary

Demographics appeared largely complete, although only five reports described the level of data completeness explicitly. Some data of interest in the review were systematically missing from most studies (i.e. education, marital and illness trajectory descriptors) and provided in eight cases via author request (Appendix 1.5 details data provided). Some sample descriptives were combined for the purposes of this review where the population had only been reported on as subsets originally (Li et al., 2025; Russo et al., 2021).

Schizophrenia was the most common diagnosis across all studies, with all except two studies (Cuesta et al., 2021; Hosáková et al., 2017) reporting verification against a standardised nosology (most commonly DSM-IV or ICD-10) or diagnostic interview. Diagnostic exceptions included Thomson (2019) who looked at first episode psychosis and at risk mental state, Cuesta et al. (2021) who additionally included participants meeting criteria for mood disorder with psychotic symptoms, and one study investigating the applicability of the CAINS in bipolar and major depressive disorder samples (Li et al., 2025). Only three studies reported collecting inpatient data (Engel et al., 2014; Jung et al., 2016; Richter et al., 2019). Less than half of reports specified the recruitment setting: these included community health and social care services, assertive treatment teams, university psychiatric departments, and more generic medical or mental health facilities.

The age ranges indicated a minimum cut-off of between 16-21, and a maximum of age 59 and no upper limit. Several studies had inclusion criteria stipulating that recruited individuals must have proficient skills in the language the CAINS was being delivered in (ID: 2, 3, 5, 14, 15, 17, 20, 24, and 11 (UK Sample only)). Several studies (ID: 2, 15, 17) included subsets of data from intervention studies which had additional inclusion and exclusion criteria, including in some cases meeting cut-off scores on additional symptom measures. One study (Strauss et al., 2018) reported including data from larger protocols with one of these being interventional but did not give details. A summary of all additional relevant inclusion and exclusion criteria and are highlighted in Appendix 1.6.

A meta-analytic summary of participant demographics reported consistently across studies was conducted and results are outlined in Table 1.3. Overall, estimates were

widely spread (lowest I^2 = 85.63%) demonstrating the results were related to distribution of the true estimate as opposed to simply sampling error. The prediction intervals for the results ranged between a lower limit of 1.58 units from the mean (Years of Education) and an upper limit of 8.078 units from the mean (CAINS; one fifth of the possible score variance) demonstrating that there was some, but not substantial variability in demographics across studies, comparable for example to the variability in the original CAINS validation sample. Using the DerSimonian-Laird estimator for comparison identified no significant differences in model parameters (2% variance or less). Thomson (2019) appeared to include significantly younger participants. Stoilkovska et al. (2021) appeared visually to have a lower proportion of the sample whose marital status was single although this was statistically non-significant. Ristić et al. (2020) had a significantly older age of illness onset. Li et al. (2025) had a higher proportion of females and Vayisoğlu et al. (2024) had a higher proportion of males than other samples but neither were significant outliers. Finally, Vayisoğlu et al. (2024) had a significantly lower years of education than other samples.

Figure 1.2 and 1.3 demonstrate the spread of samples globally (categorised by country and by sample density respectively). The largest samples were the United States and China samples. When plotting demographic and clinical variables by country as a proxy measure for cultural differences, few clear patterns emerged. The Scottish sample and two Chinese samples appeared to have somewhat lower mean levels of negative symptoms (the latter two less substantially so); the Chinese and South East Asian samples had a higher proportion of females; and the Scottish and Albanian samples appeared lower in age. As none of these individual findings persisted across other demographics, it is unclear whether these findings could be attributed to cultural differences in those most likely to experience negative symptoms or whether this simply represents sampling variation. No other demographic variables had sufficiently complete data to identify any other trends.

Fig. 1.2 Sample location categorised by country

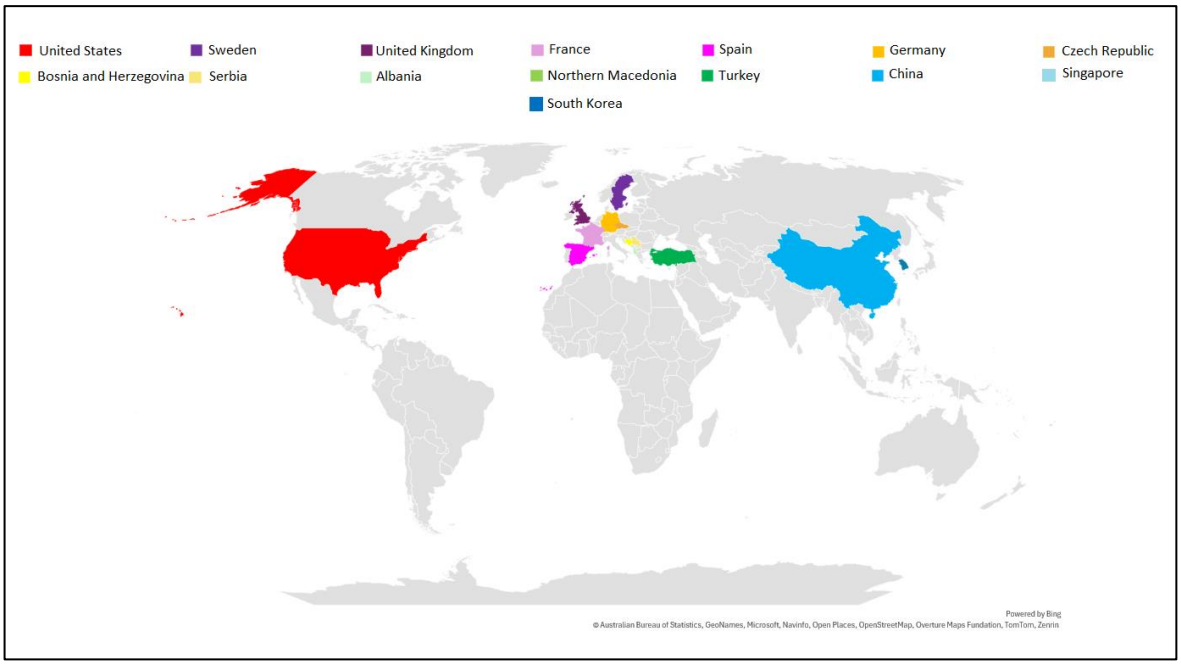


Fig. 1.3 Sample geographic distribution coded by sample density

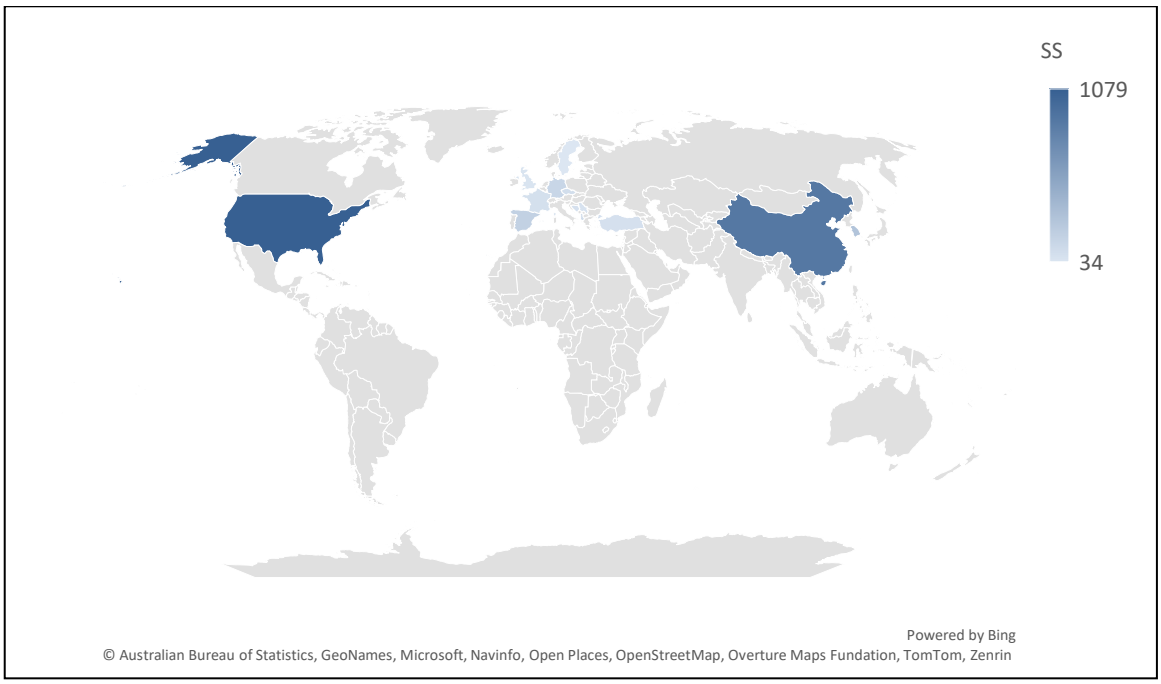


Table 1.3: Pooled estimates of demographic and clinical variables

Variable	Weighted Average	95% Confidence Interval	95% Prediction Interval
Age	37.361	34.454 – 40.269	23.272 – 51.451
Proportion Male	61.25%	56.99 - 65.51	42.44 – 80.66
Years of Education	11.491	10.932 – 12.050	9.915 – 13.067
Proportion single	74.80%	69.63 – 79.96	57.21 – 92.38
Duration of illness	10.703	7.690 – 13.716	0 – 21.440
Age of illness onset	24.866	23.385-26.347	20.719 – 29.013
CAINS Mean	21.044	18.878 – 23.211	12.337 – 29.752

1.3.4 CAINS factor structure

Across the 23 studies reporting CAINS factor structure data, six best fitting factor solutions were reported. There were two two-factor solutions representing the original (14 total - ID: 2, 3, 5, 6, 7, 8, 9, 11, 12, 15, 20, 22, 23, 24); and an alternate solution (2 total: Chan et al., 2015; Stoilkovska et al., 2021). A further three- and four- factor solutions were found by one (Ristić et al., 2020) and three (Kulenović et al., 2022; Rekhi et al., 2019; Uka et al., 2020) studies respectively to be of best fit. Finally, two five-factor solutions performed best across several studies: one based on NIMH consensus domains which performed better in one study as a single-order structure (Li et al., 2025) and better in another as a hierarchical two-order structure (Strauss et al., 2018); and a separate five-factor structure (Russo et al., 2021). The items contributing to each factor are outlined in Figure 1.4. There were no clear patterns of cultural variation in which factor structure was of best fit, with both the Chinese and North American samples ascribing to different factor structures equally. Of interest perhaps, the three- or four-factor structures were only of best-fit in south eastern European samples. However, due to the high proportion of studies reporting confirmatory factor analyses only it is impossible to say whether this is due to any cultural variation versus sampling differences.

Fig. 1.4 CAINS factor structures reported across studies

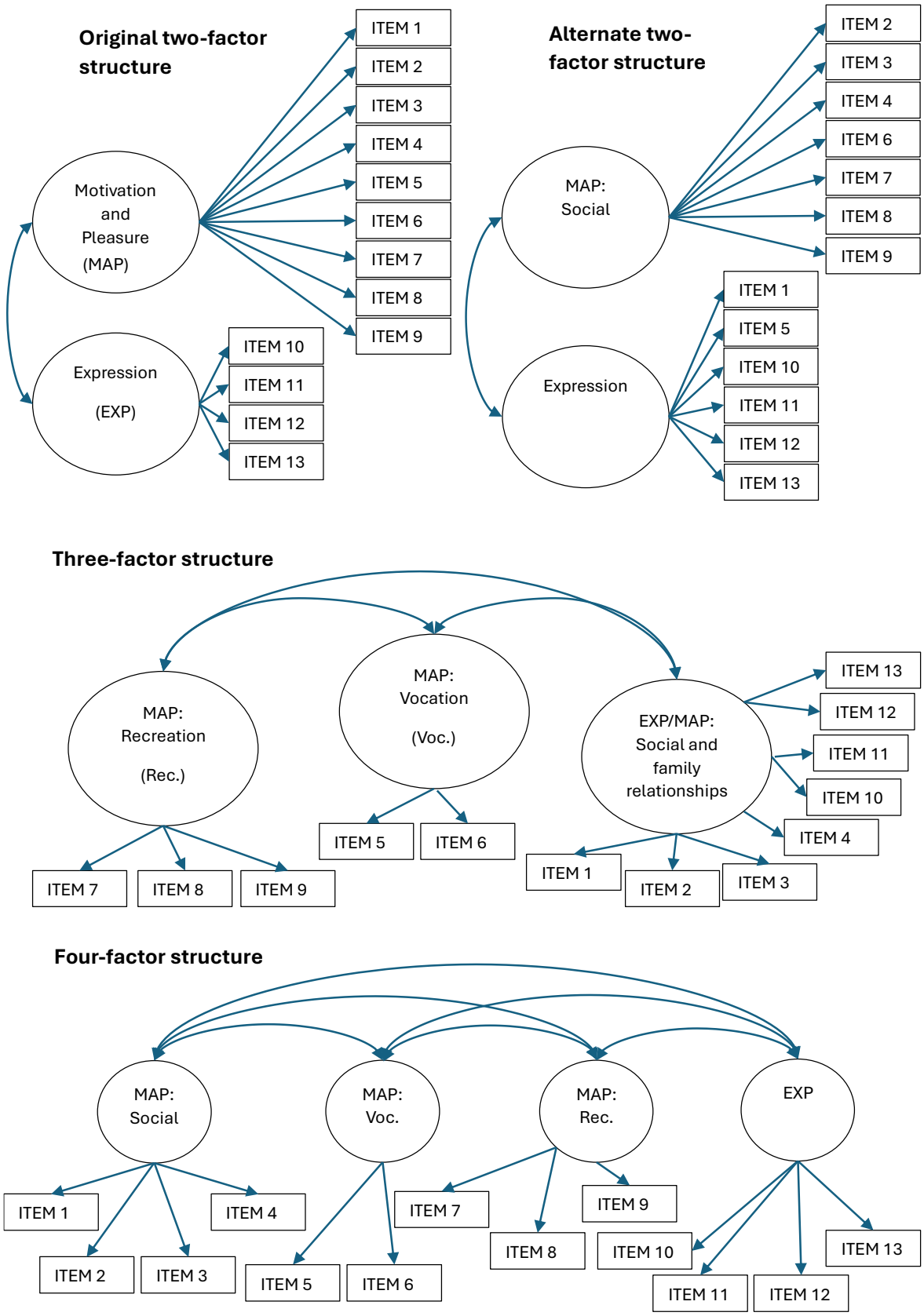
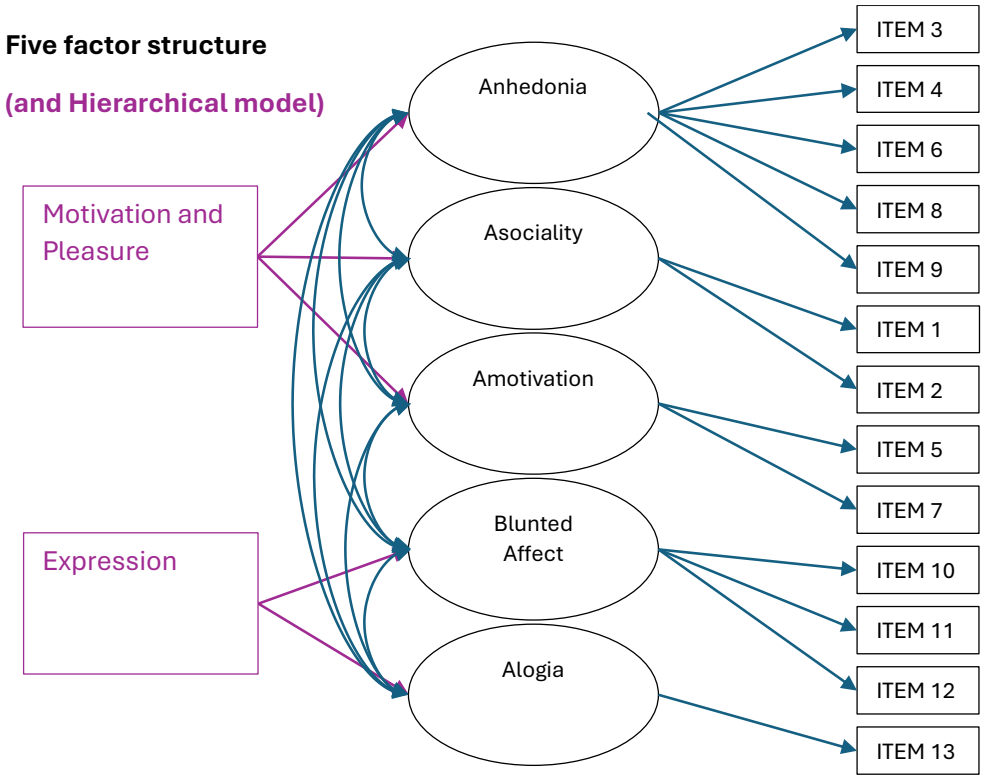
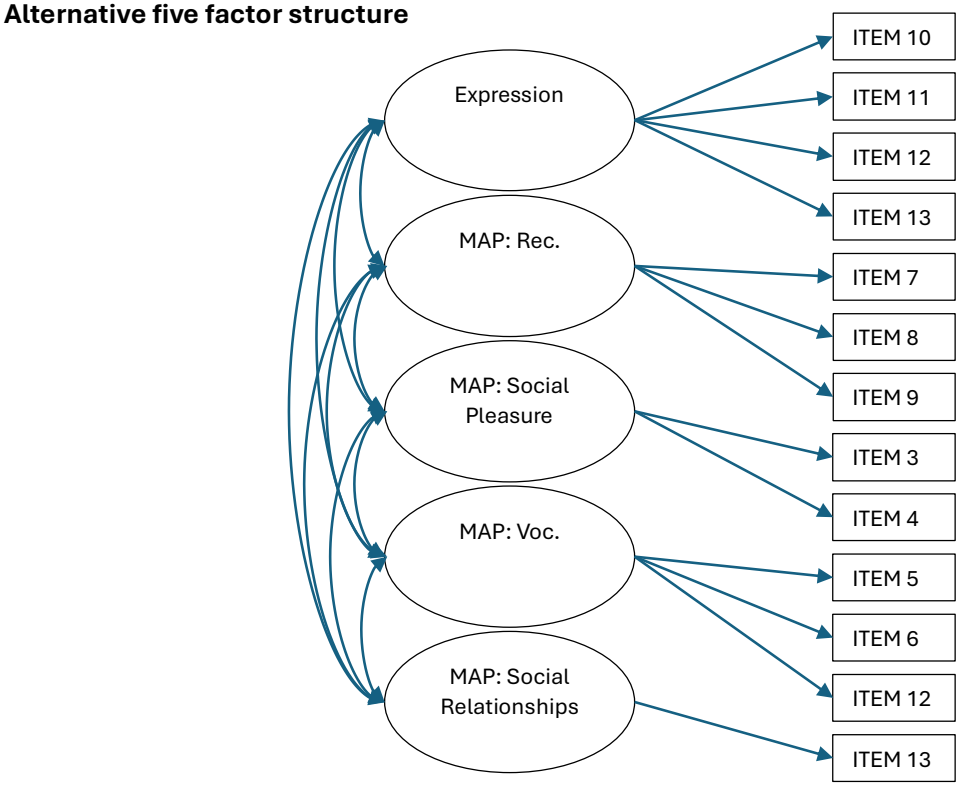


Fig. 1.4 contd


Five factor structure
(and Hierarchical model)
Motivation and Pleasure
Expression
Anhedonia
Asociality
Amotivation
Blunted Affect
Alogia
ITEM 3
ITEM 4
ITEM 6
ITEM 8
ITEM 9
ITEM 1
ITEM 2
ITEM 5
ITEM 7
ITEM 10
ITEM 11
ITEM 12
ITEM 13


Alternative five factor structure
Expression
MAP: Rec.
MAP: Social Pleasure
MAP: Voc.
MAP: Social Relationships
ITEM 10
ITEM 11
ITEM 12
ITEM 13
ITEM 7
ITEM 8
ITEM 9
ITEM 3
ITEM 4
ITEM 5
ITEM 6
ITEM 12
ITEM 13

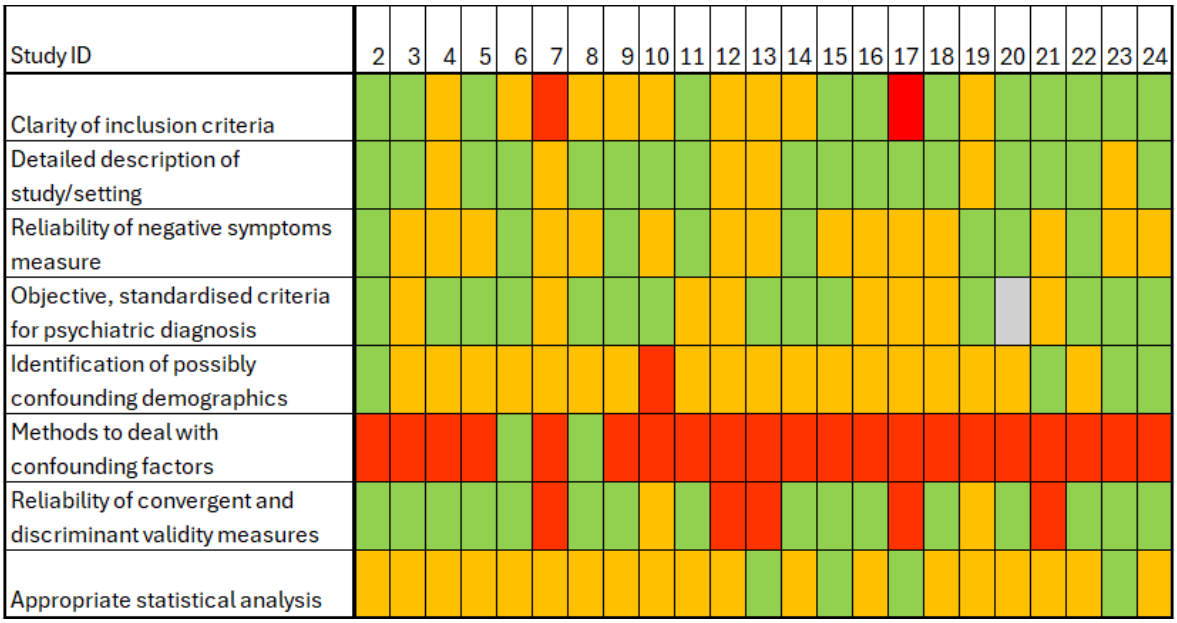
1.3.5 Risk of bias assessment

Most risk of bias ratings for the included studies were “unclear”, often due to incomplete information (see Figure 1.5 for a summary). Most studies did not explore whether there were confounding demographic factors and only two studies made any acceptable attempts to control for potentially confounding demographics statistically (Engel et al., 2014; Jang et al., 2017). Other sources of high risk of bias included no description of inclusion/exclusion criteria (Hosáková et al., 2017; Russo et al., 2021); lack of socioeconomic descriptors (Kulenović et al., 2022), and inadequate exploration of convergent and discriminant validity (ID: 7, 12, 13, 17, 21). It is worth noting that the highest risk of bias ratings came from a poster presentation (Hosáková et al., 2017) and it is likely that the non-English language publication contained relevant information that would have lowered these ratings.

Unclear ratings were especially pertinent for both descriptions of methods to ensure reliability of CAINS measurement and appropriate statistical analysis, with only half of the reports achieving a low risk of bias rating for any of these items. These ratings stemmed largely from only 13 reports giving detailed information on each of the following items relevant to CAINS application: level of qualification, training methods (including use of manuals) and supervision methods. Furthermore, five studies reported no efforts to ensure standardisation across raters. Only one study had some form of blinding to other convergent/discriminant validity measures amongst raters. Regarding statistical analysis only nine reports gave descriptions of efforts to ensure sampling adequacy (mostly the Kaiser–Meyer–Olkin test), seven reports described no methods to assess assumptions of normality necessary for the analyses conducted, only nine reports included model fit statistics, and only six studies gave descriptions of their analysis which included method of estimating the variance in the model.

When reviewed by country as a proxy measure for cultural differences, there were no clear trends for any differences in risk of bias although central and southeastern European samples and the Chinese samples were more often rated unclear on risk of bias than other studies for reliability of negative symptoms measure and use of objective standardised criteria for psychiatric diagnosis.

Fig. 1.5 Risk of bias summary



Colour Key: Red – high; Orange – unclear; Green – low; Grey – non-applicable

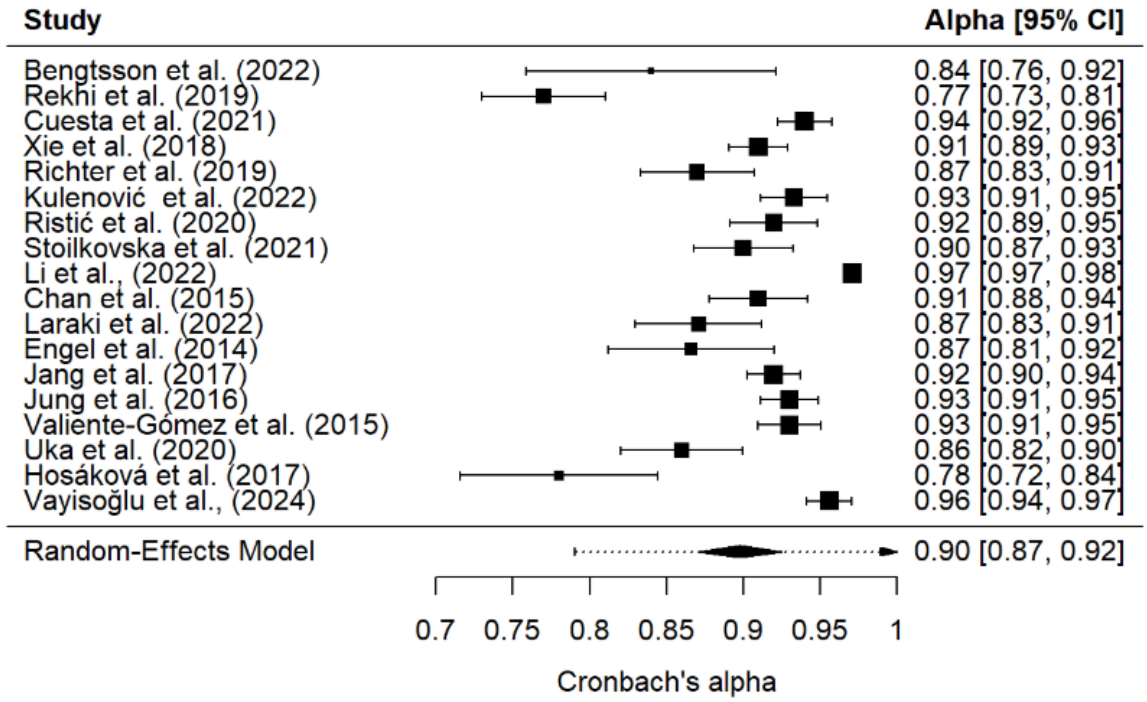
1.3.6 Meta analysis results

When reviewed by country as a proxy measure for cultural differences, there were no strong trends in Cronbach's alpha across the data, although some eastern samples (south east Europe, and Asia) did appear to have a higher estimates. Given that these studies were also more likely to have unclear risk of bias ratings for measurement of the CAINS, it is not possible to interpret whether this is the result of stringent administration processes or true differences in the internal consistency of the data in these samples. There were no trends in the ICC results.

Six studies did not provide estimates of internal consistency (ID: 1, 3, 13, 17, 20) and therefore did not contribute to the pooled estimate of Cronbach's alpha. The overall estimate was 0.898 (95% CI: 0.871 – 0.925), and variance across studies is depicted in the forest plot in Figure 1.6. Funnel plots and Egger's regression test were suggestive of asymmetry. If the CAINS has a true estimate of high internal consistency, this asymmetry would be expected. Trim and fill results did not estimate any missing studies due to publication bias suggesting the results were not strongly influenced by the asymmetry. The heterogeneity was I^2 = 96.04%, which indicates that variance observed is indicative of a true variance as opposed to sampling error. The 95% prediction interval was between 0.790 – 1, indicating a variability in the estimate of 0.21 when accounting for both

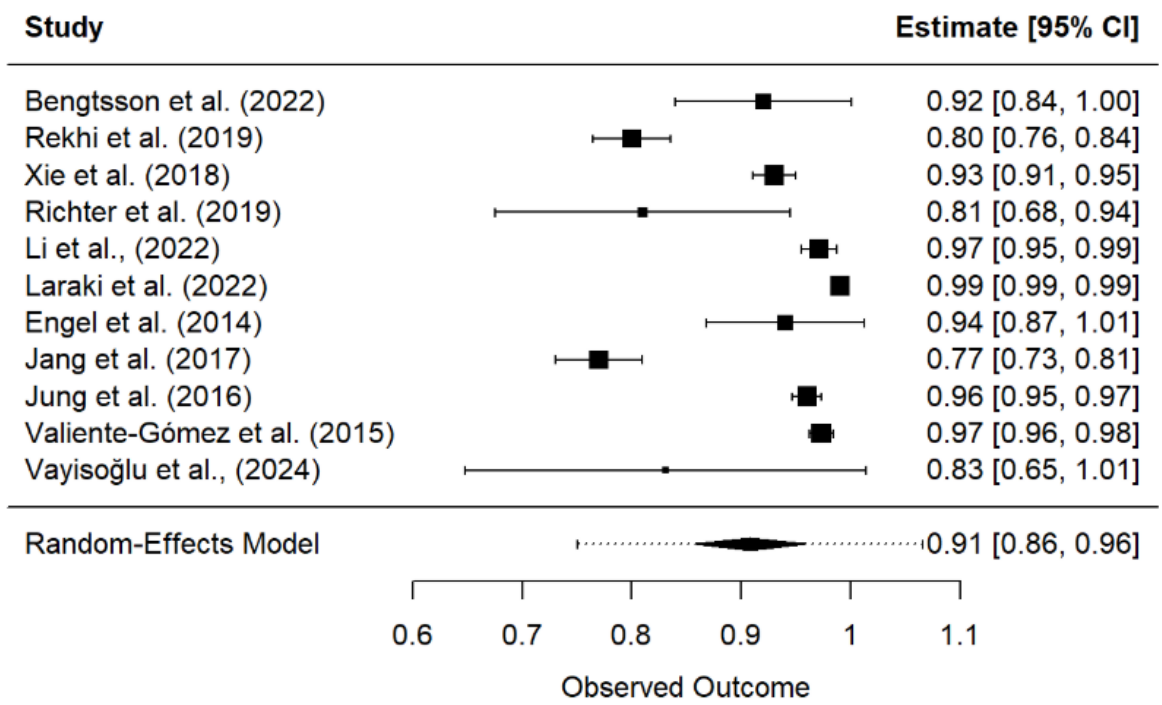
between and within study variability. The difference in results using transformed coefficients was minor, although variability in results was not detected using Bonnett's transformation.

Fig. 1.6 Forrest Plot of studies reporting Cronbach's Alpha



Eleven studies contributed to the pooled ICC estimate (ICC = 0.908, 95%CI: 0.859 – 0.957), with variance across studies depicted in Figure 1.7. Funnel plots and Egger's regression test were suggestive of asymmetry, whereby lower estimates of inter-rater reliability are unpublished. Trim and fill results estimated one study missing from the right side of the plot, with an amended estimate of 0.923 when statistically accounting for publication bias. The heterogeneity was I^2 = 98.83%, which indicates observed heterogeneity is mostly attributable to a true difference in the estimate as opposed to sampling error, and the 95% prediction interval of 0.750 – 1 (capped) suggests variation in the data when accounting for within and between study variance of 0.25.

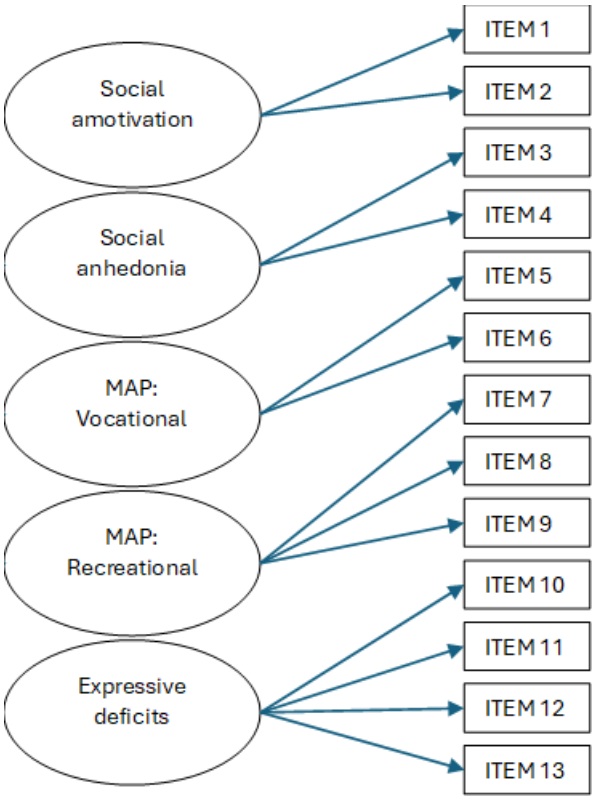
Fig. 1.7 Forrest Plot of studies reporting ICC



Sensitivity analyses included removal of demographic outliers (i.e. later age of illness onset (Ristić et al., 2020)) and clinical outliers (i.e. lower mean levels of negative symptoms (Li et al., 2022; Ristić et al., 2020)); and removal of (Ristić et al., 2020) from the ICC estimate, for failing to report an exact ICC (reporting only that above 0.80 was achieved). The results were negligible.

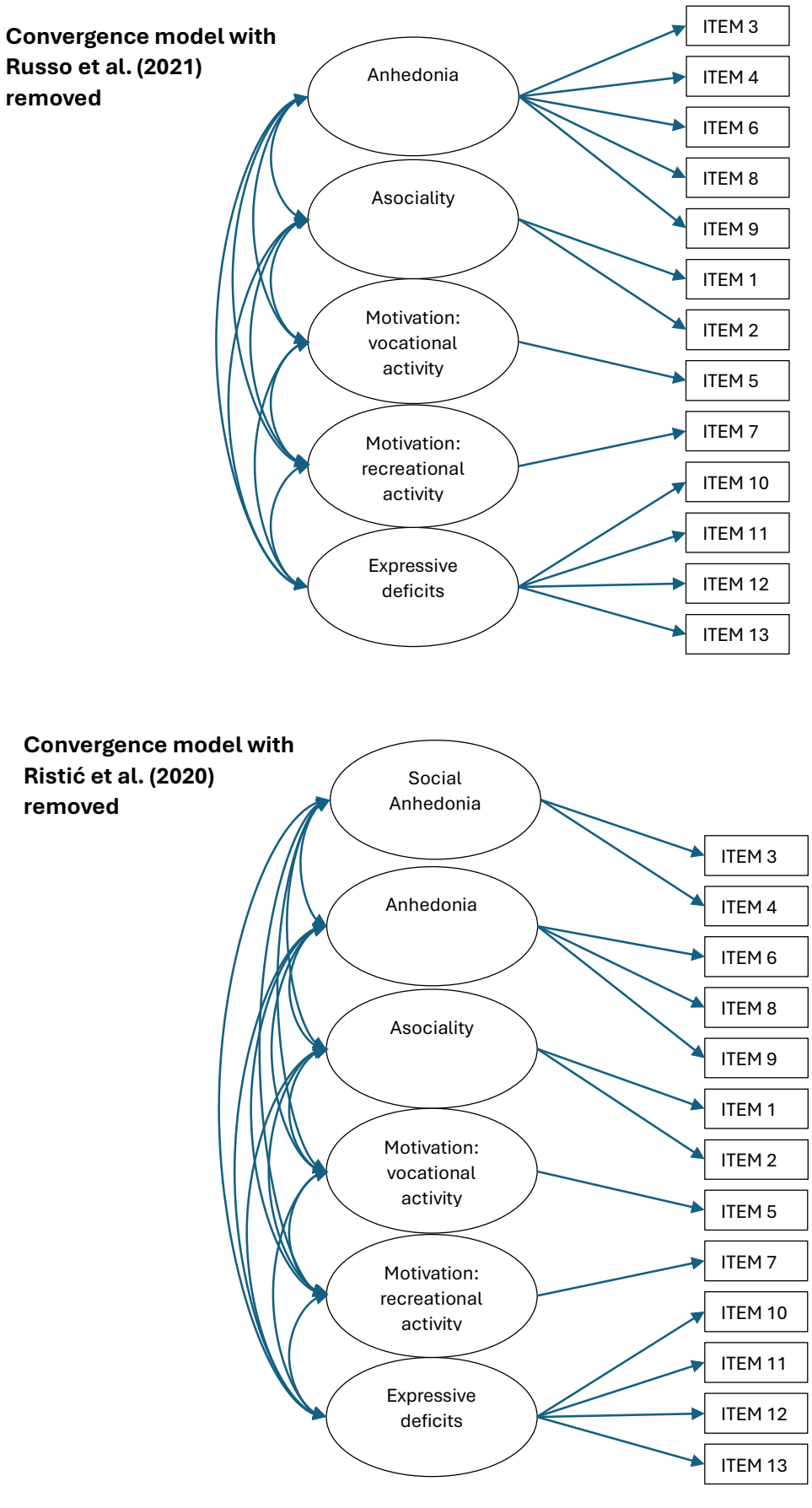
The meta-analytic EFA using co-occurrence matrices demonstrates that studies converge on the five-factor solution as the most common model. Although Unrelated Least Squares is recommended as the best estimator of between study variance for these analyses, REML was also computed due to the similarity of considerations as previous analyses in this review. The only impact this had was on whether item six (expected pleasure for vocational activity) loaded beside MAP recreational items or the other MAP vocational item. The more parsimonious factor structure was the REML structure as depicted in in Figure 1.8, which appears clearly associated with the different categories the CAINS is structured by.

Fig. 1.8 Convergence model with EFA constrained to six possible factors



Sensitivity analyses were conducted removing several outliers compared to the pooled sample including participants meeting different diagnostic criteria (Li et al., 2025; Thomson, 2019); having low negative symptoms (Li et al., 2022; Li et al., 2025; Rekhi et al., 2019; Thomson, 2019)(2,12,13,29), being younger in age (Thomson, 2019); older age of illness onset (Ristić et al., 2020); and having overlapping data (Russo et al., 2021). Removing the younger and low negative symptoms samples had no effect on the overall factor structure. Removing the overlapping data and separately the higher age of illness onset sample revealed a factor structure which grouped items to more closely resemble the anhedonia items of NIMH consensus five-factor structure; although the amotivation items did not load together and expressive deficits did not load separately. When removing the overlapping data this was presented as a five factor structure; whereas removing the older age of illness onset sample revealed a six-factor structure with social anhedonia loading separately to other anhedonia items (Figure 1.9). When constrained to two-factors; all analyses replicated the original factor structure.

Fig. 1.9 Differing convergence models in sensitivity analysis



1.4. Discussion

This review was designed to examine the sampling, methodological and analytical characteristics of studies exploring the CAINS factor structure. It aimed to conduct a critical exploration of relationships between these study components and the factor structures identified and their reliability. The CAINS has been investigated in mostly cross-sectional validation studies across north-western; European and south-east Asian and Chinese samples. Samples were predominantly similar to the range of participants contributing to validation of the original CAINS, with few transdiagnostic studies.

Like other reviews (Wehr et al., 2024) current findings demonstrate that cross-cultural replications of the CAINS are quantitatively similar to the original. Psychometrics including internal consistency and inter-rater reliability were robust, being minimally impacted by demographic factors as demonstrated by sensitivity analyses. Given these comparisons were spread relatively equally across eastern and western hemispheres, and in spite of the possibility of sub-optimal calibration methods given the unclear reporting, these results could be interpreted as somewhat reassuring that the CAINS appears to measure a consistent construct globally for these relevant research populations. Future studies which adhere to more rigorous reporting standards around the administration and statistical analysis of CAINS data would increase confidence in these findings.

The two-factor solution was the most commonly replicated structure, with consistently acceptable fit statistics. Although over 50% of studies included some form of confirmatory factor analyses, this was also true for several EFA analyses, and less remarkably, in PCA analyses (where the analyses aims to identify only the most essential factor structure components). Although this does not address whether the two-factor structure is of “best” fit, these results offer some reassurance of cross-cultural consistency to researchers and clinicians utilising the evidence-base investigating negative symptoms by these domains.

The more granulated factor structure models also presented with acceptable fit statistics and conceptually appeared to stratify the CAINS by expressive domains, and experiential deficits: grouped by either the CAINS interview structure or the NIMH consensus domains. No clear cultural differences seemed to underlie model selection, although cultural factors could not effectively be separated from sampling and risk of bias. Perhaps

researchers and clinicians should not yet then be dissuaded from utilisation of these models in further exploration of negative symptoms. However, given several of the proposed models conformed to an acceptably fitting factor structure, it remains important that researchers utilise a theoretical rationale to justify their use.

Experiential items, particularly motivation-related, were least stable in contributing to any factor. This is perhaps unsurprising given the possibility that motivation is a more multifaceted construct whose influences are not yet completely understood (Luther et al., 2019). Perhaps the CAINS, as an overarching measure of all negative symptoms, fails to assess this construct with sufficient granularity to identify a reliable factor structure separated by relevant underlying mechanisms.

Further work is required to validate the scale globally. Ability to explore cultural variance was limited due to geographical paucity; and lack of systemic cultural indices as opposed to the proxy measure of "country" (Strauss, 2021). Generalisation of conclusions beyond convenience samples from predominantly community-based psychiatric settings is also limited. Current samples largely screened out individuals with neurological and mental health comorbidities despite this being common in clinical practice (Lu et al., 2022) and individuals with higher levels of negative symptoms were underrepresented.

1.4.1 Strengths and limitations

These review findings suggest that the database searches had a high level of sensitivity to identify relevant papers and the overall search appears comprehensive to have identified relevant data. The analytical methods give a clearer indication of the convergence of factor structure and psychometric properties than would be possible with narrative synthesis alone. Limiting to English language publications only prevented inclusion of additional relevant papers. Reliability generalisation meta-analysis (Sánchez-Meca et al., 2021) may have been a more comprehensive methodology for developing pooled estimates. This review follows several steps relevant to this protocol; but should be recognised as not intending to perform reliability generalisation and therefore does not include all available estimates of Cronbach's alpha or ICCs. Meta-regression may have also more comprehensively investigated the impact of specific demographic variables on findings, but given the relative consistency of estimates and likely lack of power to conducted these analyses it was not utilised. Finally, the risk of bias assessment tool does

not formally quantify the certainty of recommendations, and within the lifespan of the review potentially more appropriate tools have been developed (Elsman et al., 2024).

1.4.2 Conclusions

This review comprehensively identified the relevant evidence base and provided an informative summary of the psychometric and factor-structure properties of CAINS validation studies across cultures. It reveals that CAINS psychometrics are relatively robust, and the factor structures which converge relate to existing negative symptom models. Results are caveated by a limited exploration samples deviating from traditional clinical trial populations; limited quantification of cultural factors; and limited geographical coverage. While researchers and clinicians may feel reassured in use of the CAINS globally, the evidence-base would benefit from clearer methodological reporting quality and including more diverse populations.

1.5 References

Abplanalp, S. J., Braff, D. L., Light, G. A., Nuechterlein, K. H., & Green, M. F. (2022). Understanding connections and boundaries between positive symptoms, negative symptoms, and role functioning among individuals with schizophrenia: A network psychometric approach. *JAMA psychiatry, 79*(10), 1014-1022. https://doi.org/https://doi.org/10.1001/jamapsychiatry.2022.2386

Ahmed, A. O., Kirkpatrick, B., Granholm, E., Rowland, L. M., Barker, P. B., Gold, J. M., Buchanan, R. W., Outram, T., Bernardo, M., Paz García-Portilla, M., Mane, A., Fernandez-Egea, E., & Strauss, G. P. (2022). Two factors, five factors, or both? External validation studies of negative symptom dimensions in schizophrenia. *Schizophrenia Bulletin, 48*(3), 620-630. https://doi.org/https://doi.org/10.1093/schbul/sbab148

Andreasen, N. C. (1989). The Scale for the Assessment of Negative Symptoms (SANS): Conceptual and theoretical foundations. *British Journal of Psychiatry, 155*(S7), 49-52. https://doi.org/10.1192/S0007125000291496

Bengtsson, J., Boden, R., Neider, D., & Cernvall, M. (2022). A blinded validation of the Swedish version of the Clinical Assessment Interview for Negative Symptoms (CAINS). *Nordic Journal of Psychiatry, 76*(1), 44-51. https://doi.org/https://doi.org/10.1080/08039488.2021.1933174

Blanchard, J. J., Bradshaw, K. R., Garcia, C. P., Nasrallah, H. A., Harvey, P. D., Casey, D., Csoboth, C. T., Hudson, J. I., Julian, L., Lentz, E., Nuechterlein, K. H., Perkins, D. O., Skale, T. G., Snowden, L. R., Tandon, R., Tek, C., Velligan, D., Vinogradov, S., & O'Gorman, C. (2017). Examining the reliability and validity of the Clinical Assessment Interview for Negative Symptoms within the Management of Schizophrenia in Clinical Practice (MOSAIC) multisite national study. *Schizophrenia Research, 185*, 137-143. https://doi.org/https://doi.org/10.1016/j.schres.2017.01.011

Bleuler, E. (1911). *Dementia praecox or the group of schizophrenias*. Deuticke.

Bonett, D. G. (2002). Sample size requirements for testing and estimating coefficient alpha. *Journal of Educational and Behavioral Statistics, 27*(4), 335-340. https://doi.org/https://doi.org/10.3102/10769986027004335

Borenstein, M. (2020). Research Note: In a meta-analysis, the I2 index does not tell us how much the effect size varies across studies. *Journal of physiotherapy, 66*(2), 135-139. https://doi.org/https://doi.org/10.1016/j.jphys.2020.02.011

Chan, R. C., Shi, C., Lui, S. S., Ho, K. K., Hung, K. S., Lam, J. W., Wang, Y., Cheung, E. F., & Yu, X. (2015). Validation of the Chinese version of the Clinical Assessment Interview for Negative Symptoms (CAINS): A preliminary report. *Frontiers in Psychology, 6*, 7. https://doi.org/https://doi.org/10.3389/fpsyg.2015.00007

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334. https://doi.org/https://doi.org/10.1007/BF02310555

Cuesta, M. J., Sánchez-Torres, A. M., Lorente-Omeñaca, R., Moreno-Izco, L., Peralta, V., & Group, S. (2021). Cognitive, community functioning and clinical correlates of the Clinical Assessment Interview for Negative Symptoms (CAINS) in psychotic

disorders. *European Archives of Psychiatry and Clinical Neuroscience, 271*(8), 1537-1546. https://doi.org/https://doi.org/10.1007/s00406-020-01188-x

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455-463. https://doi.org/https://doi.org/10.1111/j.0006-341x.2000.00455.x

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *bmj, 315*(7109), 629-634. https://doi.org/https://doi.org/10.1136/bmj.315.7109.629

Elsman, E. B. M., Mokkink, L. B., Terwee, C. B., Beaton, D., Gagnier, J. J., Tricco, A. C., Baba, A., Butcher, N. J., Smith, M., Hofstetter, C., Aiyegbusi, O. L., Berardi, A., Farmer, J., Haywood, K. L., Krause, K. R., Markham, S., Mayo-Wilson, E., Mehdipour, A., Ricketts, J., Szatmari, P., Touma, Z., Moher, D., & Offringa, M. (2024). Guideline for reporting systematic reviews of outcome measurement instruments (OMIs): PRISMA-COSMIN for OMIs 2024. *Journal of Clinical Epidemiology, 173*(111422), 1-19. https://doi.org/https://doi.org/10.1016/j.jclinepi.2024.111422

Engel, M., Fritzsche, A., & Lincoln, T. M. (2014). Validation of the German version of the clinical assessment interview for negative symptoms (CAINS). *Psychiatry Research, 220*(1-2), 659-663. https://doi.org/https://doi.org/10.1016/j.psychres.2014.07.070

Escaffi-Schwarz, M., Gempp, R., & Irmer, J. P. (2025). Meta-analysing the factor structure and reliability of measurement instruments: An R-based tutorial. *International Journal of Psychology, 60*(2), 1-9. https://doi.org/https://doi.org/10.1002/ijop.70003

Galderisi, S., Kaiser, S., Bitter, I., Nordentoft, M., Mucci, A., Sabé, M., Giordano, G. M., Nielsen, M. Ø., Glenthøj, L. B., Pezzella, P., Falkai, P., Dollfus, S., & Gaebel, W. (2021). EPA guidance on treatment of negative symptoms in schizophrenia. *European Psychiatry, 64*(1), 1-15. https://doi.org/https://doi.org/10.1192/j.eurpsy.2021.13

Galderisi, S., Mucci, A., Buchanan, R. W., & Arango, C. (2018). Negative symptoms of schizophrenia: New developments and unanswered research questions. *Lancet Psychiatry, 5*(8), 664-677. https://doi.org/10.1016/S2215-0366(18)30050-6

Giordano, G. M., Caporusso, E., Pezzella, P., & Galderisi, S. (2022). Updated perspectives on the clinical significance of negative symptoms in patients with schizophrenia. *Expert Review of Neurotherapeutics, 22*(7), 541-555. https://doi.org/10.1080/14737175.2022.2092402

Giordano, G. M., Sanmarchi, F., Mucci, A., Rucci, P., Brando, F., Caporusso, E., Giuliani, L., Melillo, A., Pezzella, P., Bucci, P., Rocca, P., Rossi, A., Bertolino, A., Rossi, R., Pergola, G., Galderisi, S., Maj, M., & Psychoses, I. N. f. R. o. (2023). External validation of the five domains of negative symptoms: Focus on cognition, functional capacity, and real-world functioning. *European Psychiatry, 67*(1), 1-9. https://doi.org/10.1192/j.eurpsy.2023.2478

Guessoum, S. B., Le Strat, Y., Dubertret, C., & Mallet, J. (2020). A transnosographic approach of negative symptoms pathophysiology in schizophrenia and depressive disorders. *Progress in Neuropsychopharmacology and Biological Psychiatry, 99*(109862), 1-12. https://doi.org/10.1016/j.pnpbp.2020.109862

Hakstian, A. R., & Whalen, T. E. (1976). A k-sample significance test for independent alpha coefficients. *Psychometrika, 41*(2), 219-231. https://doi.org/https://doi.org/10.1007/BF02291840

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2022). *Cochrane handbook for systematic reviews of interventions version 6.3*. Cochrane. https://www.training.cochrane.org/handbook

Hosáková, K., Viktorová, L., Lečbych, M., & Hosák, L. (2017). *Clinical Assessment Interview for Negative Symptoms: Validation of the Czech version* 13th World Congress of Biological Psychiatry, https://www.researchgate.net/publication/320518358_Clinical_Assessment_Interview_for_Negative_Symptoms_Validation_of_the_Czech_version

Jang, S.-K., Park, S.-C., Choi, K.-H., Yi, J.-S., Park, J.-K., Lee, J. S., & Lee, S.-H. (2017). Validation of the Korean version of the Clinical Assessment Interview for Negative Symptoms. *Psychiatry investigation, 14*(4), 413-419. https://doi.org/https://doi.org/10.4306/pi.2017.14.4.413

Joanna Briggs Institute. (2017). Checklist for analytical cross sectional studies. https://jbi.global/critical-appraisal-tools

Jung, S. I., Woo, J., Kim, Y.-T., & Kwak, S. G. (2016). Validation of the Korean-version of the Clinical Assessment Interview for Negative Symptoms of schizophrenia (CAINS). *Journal of Korean Medical Science, 31*(7), 1114-1120. https://doi.org/https://doi.org/10.3346/jkms.2016.31.7.1114

Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophrenia Bulletin, 13*(2), 261-276. https://doi.org/10.1093/schbul/13.2.261

Kirkpatrick, B., Fenton, W. S., Carpenter, W. T., & Marder, S. R. (2006). The NIMH-MATRICS consensus statement on negative symptoms. *Schizophrenia Bulletin, 32*(2), 214-219. https://doi.org/10.1093/schbul/sbj053

Kirkpatrick, B., Strauss, G. P., Linh, N., Fischer, B. A., Daniel, D. G., Cienfuegos, A., & Marder, S. R. (2011). The Brief Negative Symptom Scale: Psychometric properties. *Schizophrenia Bulletin, 37*(2), 300-305. https://doi.org/10.1093/schbul/sbq059

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in medicine, 22*(17), 2693-2710. https://doi.org/https://doi.org/10.1002/sim.1482

Kring, A. M., Gur, R. E., Blanchard, J. J., Horan, W. P., & Reise, S. P. (2013). The Clinical Assessment Interview for Negative Symptoms (CAINS): Final development and validation. *American Journal of Psychiatry, 170*(2), 165-172. https://doi.org/10.1176/appi.ajp.2012.12010109

Kulenović, A. D., Mešević, E. S., Ribić, E., Repišti, S., Radojičić, T., & Jovanović, N. (2022). Factor structure and internal consistency of the Clinical Assessment Interview for

Negative Symptoms (CAINS) on a sample of persons with psychotic disorders in Bosnia and Herzegovina. *Journal of Neural Transmission, 129*(7), 905-911. https://doi.org/https://doi.org/10.1007/s00702-021-02435-8

Laraki, Y., Lebrun, C., Merenciano, M., Eisenblaetter, M., Attal, J., Macgregor, A., Decombe, A., Capdevielle, D., & Raffard, S. (2022). Validation of the French clinical assessment interview for negative symptoms in a sample of stable French individuals with schizophrenia. *Frontiers in Psychiatry, 13*(836600), 1-8. https://doi.org/https://doi.org/10.3389/fpsyt.2022.836600

Li, S.-b., Liu, C., Zhang, J.-b., Wang, L.-l., Hu, H.-x., Chu, M.-y., Wang, Y., Lv, Q.-y., Lui, S. S., Cheung, E. F., Yi, Z.-H., & Chan, R. C. (2022). Revisiting the latent structure of negative symptoms in schizophrenia: Evidence from two second-generation clinical assessments. *Schizophrenia Research, 248*, 131-139. https://doi.org/https://doi.org/10.1016/j.schres.2022.08.016

Li, S.-B., Zhang, J.-B., Liu, C., Wang, L.-L., Hu, H.-X., Chu, M.-Y., Wang, Y., Lv, Q.-Y., Lui, S. S., Yi, Z.-H., & Chan, R. C. (2025). A transdiagnostic approach of negative symptoms in psychiatric disorders: Replication of a two-factor structure in major depressive disorder and bipolar disorder. *European Archives of Psychiatry and Clinical Neuroscience, 275*(4), 1197-1208. https://doi.org/https://doi.org/10.1007/s00406-024-01934-5

López-López, J. A., Botella, J., Sánchez-Meca, J., & Marín-Martínez, F. (2013). Alternatives for mixed-effects meta-regression models in the reliability generalization approach: A simulation study. *Journal of Educational and Behavioral Statistics, 38*(5), 443-469. https://doi.org/https://doi.org/10.3102/1076998612466142

Lu, C., Jin, D., Palmer, N., Fox, K., Kohane, I. S., Smoller, J. W., & Yu, K.-H. (2022). Large-scale real-world data analysis identifies comorbidity patterns in schizophrenia. *Translational psychiatry, 12*(1), 1-8. https://doi.org/https://doi.org/10.1038/s41398-022-01916-y

Luther, L., Fischer, M. W., Firmin, R. L., & Salyers, M. P. (2019). Clarifying the overlap between motivation and negative symptom measures in schizophrenia research: A

meta-analysis. *Schizophrenia Research, 206*, 27-36. https://doi.org/https://doi.org/10.1016/j.schres.2018.10.010

Mokkink, L. B., Prinsen, C. A. C., Bouter, L. M., de Vet, H. C. W., & Terwee, C. B. (2016). The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. *Brazilian Journal of Physical Therapy, 20*(2), 105-113. https://doi.org/10.1590/bjpt-rbf.2014.0143

Noble, S., Scheinost, D., & Constable, R. T. (2019). A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. *Neuroimage, 203*(116157), 1-15. https://doi.org/https://doi.org/10.1016/j.neuroimage.2019.116157

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *bmj, 372*(n160), 1-36. https://doi.org/10.1136/bmj.n160

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., Britten, N., Roen, K., & Duffy, S. (2006). Guidance on the conduct of narrative synthesis in systematic reviews. *A product from the ESRC methods programme Version 1*. https://doi.org/10.13140/2.1.1018.4643

Rekhi, G., San Ang, M., Yuen, C. K. Y., Ng, W. Y., & Lee, J. (2019). Assessing negative symptoms in schizophrenia: Validity of the Clinical Assessment Interview for Negative Symptoms in Singapore. *Schizophrenia Research, 206*, 177-182. https://doi.org/https://doi.org/10.1016/j.schres.2018.11.029

Richter, J., Hesse, K., Schreiber, L., Burmeister, C. P., Eberle, M.-C., Eckstein, K. N., Zimmermann, L., Wildgruber, D., & Klingberg, S. (2019). Evidence for two distinct domains of negative symptoms: Confirming the factorial structure of the CAINS.

Psychiatry Research, 271, 693-701. https://doi.org/https://doi.org/10.1016/j.psychres.2018.12.043

Ristić, I., Jerotić, S., Zebić, M., Savić, B., Vuković, V., Russo, M., Voskresenski, T., Jovanović, N., & Marić, N. P. (2020). Factorial structure of the Serbian version of the Clinical Assessment Interview for Negative Symptoms–Evidence for three factors of negative symptoms. *Frontiers in Psychology, 11*(570356), 1-7. https://doi.org/https://doi.org/10.3389/fpsyg.2020.570356

Russo, M., Repišti, S., Stoilkovska, B. B., Jerotić, S., Ristić, I., Smajic, E. M., Uka, F., Arenliu, A., Bajraktarov, S., Kulenović, A. D., Stevović, L. I., Priebe, S., & Jovanović, N. (2021). Structure of negative symptoms in schizophrenia: An unresolved issue. *Frontiers in Psychiatry, 12*(785144), 1-8. https://doi.org/https://doi.org/10.3389/fpsyt.2021.785144

Sánchez-Meca, J., Marín-Martínez, F., López-López, J. A., Núñez-Núñez, R. M., Rubio-Aparicio, M., López-García, J. J., López-Pina, J. A., Blázquez-Rincón, D. M., López-Ibáñez, C., & López-Nicolás, R. (2021). Improving the reporting quality of reliability generalization meta-analyses: The REGEMA checklist. *Research Synthesis Methods, 12*(4), 516-536. https://doi.org/https://doi.org/10.1002/jrsm.1487

Shafer, A. (2005). Meta-analysis of the brief psychiatric rating scale factor structure. *Psychological assessment, 17*(3), 324-335. https://doi.org/https://doi.org/10.1037/1040-3590.17.3.324

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological bulletin, 86*(2), 420-428. https://doi.org/https://doi.org/10.1037/0033-2909.86.2.420

Stevović Injac, L., Radojičić, T., Repišti, S., Ražnatović, A., Barac-Otašević, Z., Perunović Jovanović, T., Ljutica, I., Kalač, S., Brajković, I., & Purišić, A. (2019). Psihometrijska validacija intervjua za kliničku procjenu negativnih sindroma (CAINS) na uzorku osoba sa psihotičnim poremećajima u Crnoj Gori. *Psihijatrija, 1*(1), 68-78.

Stoilkovska, B. B., Bajraktarov, S., Simoska, S. M., Milutinović, M., Novotni, L., Arsova, S., Novotni, G., Samardziska, V. Č., Delova, S., Novotni, A., & Jovanović, N. (2021).

Psychometric characteristics of the macedonian version of Clinical Assessment Interview for Negative Symptoms (CAINS). *Psychiatria Danubina, 33*(3), 347-353. https://doi.org/https://doi.org/10.24869/psyd.2021.347

Strauss, G. P. (2021). A bioecosystem theory of negative symptoms in schizophrenia. *Frontiers in Psychiatry, 12*(655471), 1-13. https://doi.org/10.3389/fpsyt.2021.655471

Strauss, G. P. (2024). Environmental factors contributing to negative symptoms in youth at clinical high risk for psychosis and outpatients with schizophrenia. *Social Psychiatry and Psychiatric Epidemiology, 59*(7), 1167-1175. https://doi.org/https://doi.org/10.1007/s00127-023-02556-3

Strauss, G. P., & Cohen, A. S. (2017). A transdiagnostic review of negative symptom phenomenology and etiology. *Schizophrenia Bulletin, 43*(4), 712-729. https://doi.org/10.1093/schbul/sbx066

Strauss, G. P., Nuñez, A., Ahmed, A. O., Barchard, K. A., Granholm, E., Kirkpatrick, B., Gold, J. M., & Allen, D. N. (2018). The latent structure of negative symptoms in schizophrenia. *JAMA psychiatry, 75*(12), 1271-1279. https://doi.org/10.1001/jamapsychiatry.2018.2475

Tatsumi, K., Kirkpatrick, B., Strauss, G. P., & Opler, M. (2020). The brief negative symptom scale in translation: A review of psychometric properties and beyond. *European Neuropsychopharmacology, 33*, 36-44. https://doi.org/https://doi.org/10.1016/j.euroneuro.2020.01.018

Thomson, A. (2019). *Roles of attachment, reflective function and emotion regulation in the development of negative symptoms* [Doctoral Dissertation, University of Edinburgh]. http://hdl.handle.net/1842/36028

Topor, M., Pickering, J. S., Mendes, A. B., Bishop, D. V., Büttner, F., Elsherif, M. M., Evans, T. R., Henderson, E. L., Kalandadze, T., Nitschke, F. T., Staaks, J. P., R., v. d. A. O., Yeung, S. K., Zaneva, M., Lam, A., Madan, C. R., Moreau, D., O'Mahony, A., Parker, A. J., Riegelman, A., Testerman, M., & Westwood, S. J. (2023). An integrative framework for planning and conducting Non-Intervention, Reproducible, and

Open Systematic Reviews (NIRO-SR). *Meta-Psychology, 7*(7), 1-14. https://doi.org/https://doi.org/10.15626/MP.2021.2840

Uka, F., Repišti, S., Arenliu, A., Ramadani, F., Bërxulli, D., Konjufca, J., & Jovanović, N. (2020). Factor structure of the Albanian version of the Clinical Assessment Interview for Negative Symptoms (CAINS): Associations with the Brief Symptoms Inventory (BSI). *Global psychiatry, 3*(2), 169-191. https://doi.org/10.52095/gpa.2020.1375

Valiente-Gómez, A., Mezquida, G., Romaguera, A., Vilardebò, I., Andrés, H., Granados, B., Larrubia, J., Pomarol-Clotet, E., McKenna, P. J., Sarró, S., & Bernardo, M. (2015). Validation of the Spanish version of the Clinical Assessment for Negative Symptoms (CAINS). *Schizophrenia Research, 166*(1-3), 104-109. https://doi.org/https://doi.org/10.1016/j.schres.2015.06.006

Vayisoğlu, S., Karahan, S., Gürel, Ş. C., & Yağcioğlu, A. E. A. (2024). Validity and reliability study of Turkish version of Clinical Assessment Interview for Negative Symptoms (CAINS). *Archives of Neuropsychiatry, 61*(1), 59-66. https://doi.org/10.29399/npa.28438

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of statistical software, 36*, 1-48. https://doi.org/10.18637/jss.v036.i03

Wehr, S., Weigel, L., Davis, J., Galderisi, S., Mucci, A., & Leucht, S. (2024). Clinical Assessment Interview for Negative Symptoms (CAINS): A systematic review of measurement properties. *Schizophrenia Bulletin, 50*(4), 747-756. https://doi.org/10.1093/schbul/sbad137

Weigel, L., Wehr, S., Galderisi, S., Mucci, A., Davis, J., Giordano, G. M., & Leucht, S. (2023, Jul 27). The Brief negative Symptom Scale (BNSS): A systematic review of measurement properties. *Schizophrenia, 9*(1), 45-54. https://doi.org/10.1038/s41537-023-00380-x

Xie, D.-j., Shi, H.-s., Lui, S. S., Shi, C., Li, Y., Ho, K. K., Hung, K. S., Li, W.-x., Yi, Z.-h., & Cheung, E. F. (2018). Cross cultural validation and extension of the Clinical Assessment Interview for Negative Symptoms (CAINS) in the Chinese context:

Evidence from a spectrum perspective. *Schizophrenia Bulletin, 44*(S2), 547-555. https://doi.org/https://doi.org/10.1093/schbul/sby013

Федотов, И., Павличенко, А., Чумаков, Е., Леонова, А., Сорокин, М., Богоявленская, В., Власова, В., Кузнецова, А., & Петрова, Н. (2024). Результаты многоцентрового клинического исследования по адаптации и валидизации русскоязычной версии «Клинического интервью для оценки негативных симптомов»(CAINS). *Обозрение психиатрии и медицинской психологии имени ВМ Бехтерева, 58*(4-1), 107-119. https://doi.org/https://doi.org/10.31363/2313-7053-2024-971

Chapter Two

Investigating the impact of a six-session metacognitively focused therapy on individual negative symptoms: A Single Case Experimental Design Study

Prepared in accordance with the author requirements for Journal of Psychopathology and Behavioral Assessment (https://link.springer.com/journal/10862/submission-guidelines?utm_source=slink&utm_medium=journal_finder)

Nicola McGuire[1]*

School of Health and Wellbeing, University of Glasgow, Scotland
[1]email correspondence to: n.mcguire.1@research.gla.ac.uk
School of Health and Wellbeing,
College of Medical, Veterinary and Life Sciences
University of Glasgow
Clarice Pears Building
90 Byres Road
Glasgow, Scotland
G12 8TB
ORCiD: 0000-0001-5332-7259

Statements and declarations

The primary author was responsible for project administration, study conception and design, development of study-specific data collection documents, data curation and analysis, visualization and manuscript drafting, review and editing. RW and HM contributed to the execution of the research methodology by participating in data collection (for outcome measurements and in-therapy measures respectively). HM also developed the therapy modules; and provided supervision for the project and review of the manuscript.

This project was awarded a maximum £250 budget for costs incurred through the main researchers participation in the Doctorate of Clinical Psychology. This course is funded by NHS Education for Scotland (NES). NES played no role in the research design, conduct or write up processes.

The authors have no relevant financial or non-financial interests to disclose.

Ethics approval: This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the NRS West of Scotland Research Ethics Committee (Ref: 19/WS/0090, IRAS Project ID: 256369).

Informed consent was obtained from all individual participants included in the study. Patients signed informed consent regarding publishing their data.

Meta-data underlying the generation of this manuscript (i.e. protocols and relevant study documents) and the software code used in statistical analyses are available in the Open Science Framework at the following URL: https://doi.org/10.17605/OSF.IO/ZWFJR. All other study materials are not publicly available due to their resource density, but may be made available upon request. Raw data for the study are not publicly available to preserve individuals' privacy under the European General Data Protection Regulation. Data may be made available upon request.

Abstract

Negative Symptoms (often part of schizophrenia spectrum diagnoses) are a significant source of disability. Metacognitive deficits may maintain experiences of negative symptoms through increased difficulties with self-understanding. This study explored the acceptability, feasibility and effectiveness of a six-session metacognitive therapy targeting these domains.

The study used a multiple-baseline Single Case Experimental Design, delivering therapy to schizophrenia rehabilitation inpatients with moderate to severe primary negative symptoms. Changes in metacognition; manifestations of negative symptoms (CAINS scores, activity levels); and secondary outcomes including personal and social performance and recovery were quantified. Multi-level modelling with metacognition as a moderator, as appropriate for more complete data, was also illustrated.

Three participants were recruited and retained in the study, with positive changes in therapeutic alliance; suggesting intervention acceptability and feasibility. Research protocol engagement was variable with reduced response variability on self-report measures and poor adherence to actigraphy use suggesting high burden. Most outcomes showed change in the expected direction, excepting actigraphy data which had high-level missingness. Multi-level models were mostly non-significant.

Preliminary evidence suggests psychological therapies targeting metacognition are relevant for further exploration in people with moderate to severe negative symptoms. Considering researcher and measurement factors which could improve protocol adherence is important for future studies.

Keywords

Negative symptoms, anhedonia, apathy, psychosis, schizophrenia, metacognition

Plain language summary

Understanding the impact of a six-session psychological therapy on negative symptoms

Negative symptoms are a set of difficulties commonly seen in people experiencing psychosis. These include experiencing less motivation and pleasure, seeking less social activity, and using fewer emotional expressions, gestures and language. Difficulties with metacognition (the ability to think about one's own thinking about the self, others, the world and how we respond to psychological problems) have been found in people with negative symptoms (McLeod et al., 2014) and may contribute to negative symptoms remaining a problem.

This study aimed to explore whether a six-session psychological therapy targeting metacognition was acceptable to participants; was feasible to deliver; and was capable of creating change in metacognition (evaluated through therapy transcripts) and therefore negative symptoms. To do this, individual people were recruited to different study conditions (i.e. how long before receiving treatment was varied). This is called single-case experimental design, and it allows greater confidence that when a change is observed, it is the result of the intervention, instead of improvements that person may have experienced over time anyway.

The study recruited three people who met study criteria (being 18+; receiving mental health care from inpatient rehabilitation psychology; having a psychosis-related diagnosis; and whose scores on a negative symptoms measure were classed as moderate to severe). People who had negative symptoms which appeared to be caused by other psychosis symptoms were not eligible to take part. Everyone who took part provided informed consent to participate. People were asked to wait a randomly selected amount of time before beginning therapy, and theirs and researcher's ratings on questionnaires and interviews across the study were compared to see if there were changes in their symptoms, functioning and sense of mental health recovery.

Those recruited did not drop out, suggesting the treatment was feasible. Their relationship with the therapist improved over time, suggesting the treatment was acceptable. It was also feasible to rate metacognition with therapy transcripts. People found completing the research measures tiring and repetitive, and this might explain why people provided less varied responses over time, and decided not to wear a watch

designed to measure physical activity. Most study measures showed people experienced improvement, for example improved metacognition and less experiences of negative symptoms, but watch data were hard to evaluate due to lots of missing data. Some statistical analysis was carried out, but confidence in these results is limited because of these issues. Most of these analyses did not identify any statistically significant changes.

This early-stage evidence suggests that metacognitive therapy for people with negative symptoms was acceptable, feasible, and showed benefits. Researchers should continue to investigate therapies for this group of people. However, understanding and improving on what made the study challenging for participants is an important learning point from this study.

References

McLeod, H. J., Gumley, A. I., MacBeth, A., Schwannauer, M., & Lysaker, P. H. (2014, Jul). Metacognitive functioning predicts positive and negative symptoms over 12 months in first episode psychosis. *Journal of Psychiatric Research, 54*, 109-115. https://doi.org/10.1016/j.jpsychires.2014.03.018

2.1. Introduction

Negative symptoms encapsulate diminished or absent typical experiences, such as pleasure and motivation (Elis et al., 2013). While evidence suggests these occur transdiagnostically, they are commonly conceptualised within schizophrenia spectrum diagnoses (Kaiser et al., 2011). Persistent experiences of negative symptoms can cause distress and disability (Barch et al., 2017). Research describing maintenance factors for negative symptoms implicate several mechanisms, including reward valuation deficits (Strauss & Cohen, 2017). However, negative symptom subdomains are heterogeneously associated with functioning (i.e. activity levels; Docx et al., 2013), suggesting different processes may be in operation. Limited understanding of maintenance mechanisms leaves a poor theoretical basis for treatment developments, and current treatments for negative symptoms have limited effects (Lincoln et al., 2017). Improved causal mechanistic models of negative symptom maintenance factors are therefore crucial.

Literature suggests a strong and predictive relationship between negative symptoms and metacognition (McLeod et al., 2014), with metacognition forecasting levels of negative symptoms (i.e. motivation) over relatively stable periods (i.e. six months; Luther et al., 2016). Metacognition entails the ability to synthesise affective and interpersonal experiences to create a complex picture about oneself and others; to evaluate the utility of this knowledge in social contexts; and to use this information to respond to psychological distress and interpersonal problems (Lysaker et al., 2013). Reduced metacognition could be associated with metacognition through a decrease in self-referential processing required to make sense of experiences. Individuals with negative symptoms demonstrate difficulties consistent with this, including accessing coherent self-defining memories (Bennouna-Greene et al., 2012), and processing reward and pleasure (Kring & Barch, 2014; Oorschot et al., 2013). How changes in metacognition might temporally influence experiences of negative symptoms has not been investigated through experimental manipulation to date.

This study explores whether improving metacognition through a six-session psychological therapy affects persistent experiential negative symptoms. Therapeutic techniques focus on increasing the patient's ability to synthesise perceptions about themselves and others; and to utilise these in a values-focused way to address psychosocial challenges. They are derived from studies examining negative symptom formation and maintenance, but have not been tested in a structured therapy protocol.

Using single-case experimental design (SCED), the study will test the efficacy of this intervention using a small number of participants; undertaking repeated measurements of individual change under experimental conditions, allowing individuals to serve as their own control (Krasny-Pacini & Evans, 2017). This design is particularly suited to exploratory research, which is relevant given the following novel research questions:

1. Is therapy targeting metacognitive processes acceptable to people experiencing negative symptoms?

2. Do negative symptoms (and secondary outcomes of recovery) improve following treatment focused on improving metacognition?

3. Do metacognitive improvements relate to changes in negative symptoms, following treatment?

4. Can changes in metacognitive functioning such as greater awareness of ones own mind, the mind of others, and awareness of social-interpersonal processes be detected in therapy transcripts?

2.2 Methods

The study was pre-registered on https://clinicaltrials.gov (ID: NCT03999112). Following significant revision due to unforeseen changes in the study team the final protocol was approved (https://osf.io/zwfjr/files/osfstorage/6911d7919b3745c027fdc5f7). This report adheres to SCED guidelines (Tate et al., 2016; see Appendix 2.1 for a checklist).

2.2.1 Study design

Appropriate to treatments with potential sustained effects (i.e. consolidation and continue use of learning beyond the treatment period), a *multiple-baseline* SCED was employed. This approach signals whether presence of the intervention explains the degree of change observed by examining if effects are replicated across persons, irrespective of baseline length (in which variation in key variables could naturally occur; Krasny-Pacini & Evans, 2017). Due to recruitment difficulties, baselines were administered *non-concurrently* (i.e. commenced on different calendar days). Several research papers highlight benefits of this approach and how threats to internal validity are controlled (Slocum et al., 2022).

2.2.2 Participants

Eligibility criteria was broad given that target recipients have increased risk of co-occurring physical and additional mental health difficulties. Eligible participants were:

- Aged 18 or above
- Receiving rehabilitation psychology inpatient services in NHS Greater Glasgow and Clyde.
- Meeting criteria for a schizophrenia spectrum diagnosis, defined by diagnostic codes F20-29 from the International Classification of Diseases – 10th Edition (ICD-10; World Health Organization, 1992), as identified by the clinical team and confirmed by medical records after recruitment.
- Able to provide informed consent, as determined by treating clinician.
- Experiencing moderate to severe motivation and pleasure deficits (scored 18 or above) or expressive deficits (scored eight or above) on the CAINS (Kring et al., 2013).
- Experiencing negative symptoms not better explained by the presence of positive symptoms.

All participants recruited to the study provided written informed consent to participate.

2.2.3 Setting

Clyde Ward and Kelvin House in Gartnavel Royal Hospital and the Rehabilitation Ward in Leverndale Hospital offer mental health rehabilitation services for inpatients working towards greater independence. They typically address the needs of individuals who have more longstanding or complex mental health difficulties.

2.2.4 Recruitment

Three opportunities to demonstrate an effect are required to meet multiple-baseline SCED quality criteria (Tate et al., 2013), therefore a recruitment target of three-six participants was set, allowing for possible attrition. Clinical teams identified and initially approached those considered to meet inclusion criteria with an information leaflet (https://osf.io/zwfjr/files/osfstorage/68c7ce4e3979636bac1e451b). The information sheet (https://osf.io/zwfjr/files/osfstorage/68c7ce93841df7bce9b7abe1) was

subsequently discussed with the primary researcher for those interested with the opportunity to ask questions. Twenty-four hours was given to consider and discuss participation with others at each stage. Those agreeable completed an informed consent form (https://osf.io/zwfjr/files/osfstorage/68c7cf2ba58dbcc7fddae98d) and screening to confirm eligibility. Recruitment occurred between 5th January and 15th March 2023.

2.2.5 Procedure

Randomisation

Before recruitment, baseline length was predetermined (sequence generator: https://www.random.org); known as case randomisation this minimises threats to internal and external validity (including researcher bias in allocation or analysis; Kratochwill et al., 2013). A period of four, six, or eight weeks was determined to be of sufficient length, within the study time constraints, to explore stability of measures preceding intervention. Potential participants were approached in order of identification by staff teams, blinded to the allocation schedule to minimise bias. The baseline for the participants recruited was four, six and four weeks respectively.

Intervention

The intervention comprised six, one-hour sessions of CBT-based one-one therapy targeting metacognitive processes relevant to negative symptoms maintenance. Standard NICE recommended CBT for psychosis involves 16+ sessions of psychotherapy which is rarely feasible in inpatient settings (NICE, 2014; Raphael et al., 2021). Short-term CBT-based interventions for psychosis, particularly those focusing on single, as opposed to multiple mechanisms of distress, may be as effective as traditional-length CBT interventions (Hazell et al., 2016).

Sessions focused on psychoeducation, guided discovery, behavioural goal setting, and skills training to enhance metacognitive functions relevant to negative symptom maintenance. The modular focus (https://osf.io/zwfjr/files/osfstorage/68c7cff708d27686651e450f) draws from Metacognitive and Reflective Insight Therapy which emphasises creating a coherent narrative of understanding for oneself and others in order to make sense of and act within the social world (Lysaker & Klion, 2017). It additionally addresses motivational

mechanisms linked to behavioural change including self-defeating beliefs (Beck & Rector, 2005) and engagement with personally-valued goals (White et al., 2015).

The therapist (HM) was experienced in delivery of therapies for individuals with predominant negative symptom profiles in rehabilitation settings. The procedures for developing competence in rating metacognition (described below) were equally pertinent to increasing conceptual and clinical competence in the intervention delivery. Treatment formulations and treatment strategies were discussed in monthly peer supervision.

2.2.6 Materials and measures

Data collection

Baseline measures were completed with the primary researcher (N.M) one week following the screening appointment (excluding the CAINS where screening data was used for baseline), and repeated at onset, middle and end of therapy (example schedule Appendix 2.2). The primary researcher had one year's leave from training, so some data collection (endpoint data (participant two) and intervention data (participant three)) was completed by a second researcher (RW) with a linked project assessing therapeutic mechanisms of action (Whyte, 2025). Both researchers were doctoral level clinical psychology trainees with prior experience and training in administering research measures. Treatment sessions were audio-recorded and transcribed to facilitate analysis.

Negative symptoms

Negative symptoms were the primary outcome of interest, assessed via measures described below. Negative symptom data has been previously correlated with actigraphy data indicating the validity of this approach (Docx et al., 2013). Only the actigraphy protocol provided enough repeated measurements to meet SCED outcome assessment criteria (Kratochwill et al., 2013). Objective *and* self-report measures were used to address possible issues of rater bias (including the researcher developing an inaccurate understanding of participants' intrinsic experience, or participants' difficulties with introspection).

1. **Activity data**
    a. ***Actigraphy***

Participants were given a GENEActiv Original wristwatch. Manufacturer recommendations (ActivInsights Ltd, 2025) highlight its suitability for research around activity and lifestyle and registration under the Medical Device Directive 93/42/EEC as a Class I Non-Sterile Medical Device. Participants were asked to wear the watch 24 hours a day on their non-dominant wrist. Recording was set at 10 cycles every second (10 Hz).

b. Report of time use

The Time Use Budget (TUB; Jolley et al., 2006), with adaptations to explore social and volitional information (Velligan et al., 2016), was employed to explore how participants spent their time over the past seven days. Days are divided into four time-blocks lessening the cognitive demands on participants compared with other measures; and allowing increased social activity which not accompanied by increased physical activity to be captured (Velligan et al., 2016).

2. Researcher-rated Negative Symptoms

The CAINS (Kring et al., 2013) was selected as a gold standard measure of subjective and objectively observable manifestations of negative symptoms (Kaiser et al., 2017). It has positive reliability and validity properties (Wehr et al., 2024). It is a 13-item structured clinical interview and rating scale based on activity reported over the past week and behaviour throughout the interview. Rating covers motivation and pleasure (9 items) and expression (4 items) deficits, with higher scores representing greater impairment, based on considerations given in the manual. Researchers referred to the manual in their scoring. The primary researcher completed CAINS training provided by the original developers and was calibrated acceptably against gold standard ratings in the assessment training materials (two-way mixed effect absolute agreement average rating ICC = 1, $p>0.01$).

3. Self-evaluation of Negative Symptoms (SNS)

The SNS scale (Dolfus et al., 2015) is a 20-item measure with five subdomains; social withdrawal, diminished emotional range, avolition, anhedonia, and alogia which are measured by four items each on a three point Likert scale representing strength of agreement.

Metacognition

Metacognition was also measured via self- and observer- report measures.

1. Observer ratings

The Metacognitive Assessment Scale – Adapted (MAS-A; Lysaker et al., 2005; Semerari et al., 2003) was used to rate metacognition across four components:

- self-reflectivity (understanding one's own wishes and the complexities of these)
- understanding the mind of the other (intuiting other's beliefs and that they might differ from one's own)
- decentration (recognising that our own and others' beliefs occur in particular contexts and may change across people or time)
- mastery (ability to use metacognitive knowledge to deal with psychological distress)

Each component is indexed to a hierarchical scale, with higher scores representing abilities of increasing complexity. Administration and scoring is guided by a detailed manual and codebook created by the MAS-A developers (Lysaker et al., 2021). Preparation for use of the scale includes a two day training course followed by calibration against gold standard ratings for cases from MAS-A training materials.

a. The primary researcher rated metacognition for each timepoint. Due to resource constraints, use of an independent second rater was not possible. For baseline ratings the Indiana Psychiatric Illness Interview (the IPII; Lysaker et al., 2002) was conducted. It asks questions which facilitate participants' reflections on their mental health difficulties and is phrased to allow iterative demonstration of level of metacognition without prompting. In line with previous research showing their effectiveness (Maillard et al., 2017); therapy transcripts were used for intervention phase ratings.

b. The therapist also rated metacognition immediately following each session of the intervention.

2. Self-report metacognition

The Beck Cognitive Insight Scale (BCIS; Beck et al., 2004) was used to measure self-reported metacognition. This 15 item-scale is separated into self-reflectiveness and self-certainty domains with higher scores representing greater identification with capacities highlighted. While alternative self-report measures related to the MAS have been

developed (Pedone et al., 2017), the BCIS has previously been used successfully in the study target population and correlates well with the MAS-A (Lysaker et al., 2013; 2008).

Secondary outcomes

1. Personal and social functioning

Keyworkers involved in participants' care completed the Personal and Social Performance scale (PSP; Patrick et al., 2009), which gives an impression of the social interactions, personal care, and aggressive or disturbing behaviours a person might exhibit. Categorical responses across each of the listed areas are used to inform a 10-point score range between 0 and 100, from which the rater then selects the value they believe best fits the persons' functioning overall. The original validation study validates its sensitivity to detect change over time in people with psychosis. Training was given on the use of the tool prior to ratings being established.

2. Perception of recovery

a. The Personal Questionnaire Rapid Scaling Technique (PQRST), introduced by Brett-Jones et al. (1987) was used to assess subjective recovery from self-selected statements, meaningful to each participant, about their difficulties. Participants could formulate 1-5 statements with the therapist and rated each week how much their difficulty in this domain had changed since last discussed.

b. The Questionnaire about the Process of Recovery (QPR; Neil et al., 2009), supplemented the PQRST as a more in-depth and standardised measure of recovery. This 22-item scale measures individuals' perceptions of recovery in intrapersonal (17 items) and interpersonal (5 items) domains, which are focused on personal and social role functioning and not necessarily the amelioration of symptoms.

Therapeutic Effectiveness

Metacognitive therapies are thought to be most effective when the therapist only offers interventions which stimulate metacognition at a level proximal to the participants' current level of functioning (Lysaker & Klion, 2017). The primary researcher rated therapist metacognition for each session and thus the difference across the therapy dyad. *A priori*, it was reasoned that two points higher on each metacognitive subscale would be a tolerable difference between therapist and participant level of metacognition, as this

offers a possible one-point variance for mis-estimation of participants' level of functioning when attempting an intervention which would be optimally one-point higher on each scale.

Additionally, therapeutic alliance was assessed using the Working Alliance Inventory – Short Revised form (Hatcher & Gillaspy, 2006) participant and therapist versions (wording described in Hatcher et al., 2020). Higher ratings indicate a greater prevalence of alliance factors in therapy across domains of therapeutic bond; and orientation to goals and task respectively.

2.2.7 Governance considerations

As documented in Appendix 2.3, this research received ethical approval from the NRS West of Scotland Research Ethics Committee (Ref: 19/WS/0090, IRAS Project ID: 256369) and managerial approval from NHSGGC Research and Innovation Department (Ref: UGN19MH107). The benefits and harms to potential participants and the researchers were considered, and individuals participating were able to withdraw from the study without any impact on their ongoing care and treatment. The research was audited by NHS Sponsors part way through the trial. Data was stored in accordance with Health Research Authority guidelines and GDPR, following a data management plan (https://osf.io/zwfjr/files/osfstorage/68c7d7b1d68d01f6671e46f7) stipulating anonymised data be retained over a 10-year period following study completion. Additionally, health and safety procedures (https://osf.io/zwfjr/files/osfstorage/68c7d875b87ead2115ca7b9e) and resource cost planning (https://osf.io/zwfjr/files/osfstorage/68c7d89e5b17bf1f79dae9c1) were monitored.

2.2.8 Analysis

A priori, the researcher intended to follow recommendations by Manolov and Moeyaert (2017) in selection of visual and statistical analyses, then computed in R (4.5.0; code provided in Appendix 2.4).

Pre-processing

Actigraphy outcomes were computed with the GGIR R package (van Hees et al., 2025). Its functions compute several actigraphy metrics (including estimates of non-wear, time

spent asleep, and levels of light, moderate and vigorous physical activity) based on epochs and cut-offs of the analysts choice. The default settings are recommended in most cases by the package author based on their evidence-base. Only the criterion for valid days (based on percentage of hours the watch is worn across the day) was changed in the current analyses: being lowered to a period of four hours to attempt to maximise data availability. These analyses revealed that the proportion of missing data for the actigraph watches was high across participants.

Additionally, the PSP had missing data (due to clinician unavailability) as did therapist-rated metacognition (due to a data storage issue resulting in data loss). As per recommendations by Aydin (2024), the first value carried back was imputed for these data due to its relatively acceptable performance and the expectation that this might be the least impactful to use for missing baseline and Session One data. For the more extensively missing actigraphy data, multiple imputation was employed using the mice package extension for multi-level modelling (Audigier & Resche-Rigon, 2023). Twenty datasets were imputed using two-level predictive mean matching with 50 iterations per imputation. The variables included in the planned multi-level model constituted the predictor matrix (i.e. the actigraphy outcome of interest, the total CAINS and MAS-A (researcher rated) scores, and a weighted average of different activity level data from the time use budget was included as an ancillary variable). For the sedentary behaviour and sleep duration data, which contained less inherent variability, fixed effect predictor matrices of the same variables were computed using predictive mean matching.

Statistical methods

Descriptive summaries and visual analysis were conducted to inform hypotheses as to whether there was variance in the outcome measures attributable to implementation of the intervention, predominantly utilising the scan, scdhlm, and nlme R packages (Pinheiro et al., 2025; Pustejovsky et al., 2024; Wilbert & Lüke, 2025). Percentage of Non-Overlapping Data (PND), and Percentage of All Non-Overlapping data (PAND) were used to compute proportion of change attributable to intervention delivery. To quantify the level of change across phases and the size of effect the following were also computed: Mean BaseLine Reduction (MBLR), Glass's delta, Cohen's *d*, the Hedges, Pustejovsky, and Shadish *d* statistic (henceforth described as HPSd; Hedges et al., 2013) and a model of treatment effect controlling for autocorrelation and case variability resulting in a *d* statistic comparable to Cohen's *d* developed by Pustjovsky, Hedges and Shadish

(henceforth described as PHSd; Pustejovsky et al., 2014). Statistical information relevant to the computation these outputs is described in (Manolov et al., 2016). Due to the nature of missingness, only one participant's data was available to compute these statistics. No additional parameters could be added to modelling when computing the PHS *d* statistic due to insufficient data to support model convergence. It was intended that slope and level change would also be quantified for actigraphy data however there were insufficient intervention phase datapoints to render this possible.

As per the pre-specified analyses plan, correlations between negative symptom measures were calculated. Finally, multi-level modelling was intended to explore the change in actigraphy data in baseline versus intervention, when controlling for the relationship with negative symptom measures, with change in metacognition being computed as a predictor. Multi-level models did not converge when measurement time and negative symptom variables were included, due to lack of a sufficiently variable sample. The final model analysed measured metacognition and treatment as fixed effects, expected to vary randomly across cases, controlling for autocorrelation for all actigraphy outcomes excepting measures of sedentary behaviour and sleep duration, where only fixed effect models would converge. The multi-level models were estimated under Restricted Maximum Likelihood. Pooled estimates with standard error computed according to Rueben's Rules for the imputed datasets were compared to the null model.

2.3 Results

2.3.1 Participants

Three participants (described in Table 2.1) were recruited to and completed the study. No instances of drop-out or expressed interest which did not result in recruitment occurred. There were no adverse events. All participants' reported ethnicities were White Scottish and their marital status reported as single.

Table 2.1: Participant demographics

Participant Pseudonym	**Age**	**Sex**	**CTO**	**Psychoactive medication** (daily unless stated)	**Admission period (when recruited)**	**ICD Diagnosis**
Sara	34-37	Female	No	Clozapine 225mg Aripiprazole 10mg Sodium valproate 1200mg PRN: Diazepam 5mg Lorazepam 1mg	1 year, 4 months	Schizoaffective disorder
Matt	28-32	Male	Yes	Clozapine 250mg PRN: Lorazepam 2mg	1 year, 5 months	Paranoid Schizophrenia
John	48-52	Male	No	Promethazine 75mg Monthly depot: Aripiprazole 400mg PRN: Lorazepam max 5 mg two hourly	1 year	Paranoid schizophrenia

CTO – Community Treatment Order; PRN – as required; ICD – International Classification of Diseases

2.3.2 Descriptive statistics

Tables 2.3 (actigraphy data; excepting data for John, and Matt's intervention period which was unavailable) and 2.4 (all other outcomes) summarise study data. In general correlations between variables (described fully in Appendix 2.5) were non-significant. Significant results ranged from r = -0.52, (p=0.05, for both the relationship between duration of sustained inactivity bouts and minutes of moderate activity per day; and the relationship between sleep duration and treatment onset), and r = 1, (p=0.04 for the relationship between patient-rated working alliance and BCIS score; and p>0.01 for the relationship between therapist rated working alliance and current activity satisfaction). No small effect sizes were significant.

Table 2.3: Actigraphy data summary

Variable	Matt		Sara			
	Baseline		Baseline		Intervention	
	Mean(SD)	Median	Mean (SD)	Median	Mean (SD)	Median
Percentage of non-wear (day period)	55.48(25.68)	48.13	35.04(27.41)	35.10	89.82(16.83)	100.00
Percentage of non-wear (24 hours)	70.59(15.09)	65.63	47.19(22.33)	56.25	89.16(18.23)	100.00
Number of sedentary periods per day	4.67(1.97)	5.00	6.83(3.06)	5.50	9.00(2.83)	9.00
Sustained inactivity bouts (day period)	3.30(2.45)	2.77	1.35(0.46)	1.27	2.28(2.37)	2.28
Sustained inactivity bouts of minimum 15 minutes (day period)	3.11(2.39)	2.60	0.64(0.32)	0.75	1.63(2.31)	1.63
Duration of estimated sleep period (hours)	12.82(6.10)	13.79	9.99(7.77)	13.64	0.78(0.94)	0.78
Inactivity per day (minutes)	671.67(339.81)	635.00	613.43(231.99)	638.00	970.00(452.55)	970.00
Light activity minutes per day	49.17(34.75)	40.50	81.57(19.16)	82.00	44.50(4.95)	44.50
Moderate activity minutes per day	11.00(11.00)	9.00	34.57(24.04)	40.00	11.00(4.24)	11.00
Average acceleration (least active 5 hours per day)	0.00	0.00	2.23(1.79)	2.81	0.00	0.00
Average acceleration (most active 5 hours per day)	21.56(5.79)	21.44	38.33(12.07)	40.22	19.15	19.15

Table 2.4: Key study outcomes

Variable	Sara			Matt			John		
	Baseline (M)	Intervention		Baseline (M)	Intervention		Baseline (M)	Intervention	
		M(SD)	Median		M(SD)	Median		M(SD)	Median
BCIS	21	26(5.29)	24	15	18(23.81)	9	14	22.67(1.53)	23
CAINS EXP	10	8(2)	8	10	9.667(0.58)	10	10	4.33(2.31)	3
CAINS MAP	20	16(4)	16	22	22.333(1.52)	22	25	20(1.73)	21
CAINS Total	30	24(5.29)	22	32	32(1.73)	31	35	24.33(3.51)	24
SNS Total	15	13(5.57)	12	15	0(0)	0	22	19(3.46)	21
Total MAS-A - RR	8.5	20.5(2.60)	22	11	14.5(2.60)	16	9	15.83(0.29)	16
Total MAS-A - TR		17.75(1.13)	17.5		11.5(2.12)	12.5		14.17(1.53)	14.5
PSP	55	57.67(14.19)	55	78	73.33(5.77)	70	64	59.33(23.35)	64
QPR	47	54(2.65)	55	30	55(8.66)	60	30	32(4)	32
Weighted TUB Activity	40	25(8.19)	27	58	49.33(7.51)	46	37	26.67(3.51)	27
TUB No. of activities with individuals	17	6.33(4.04)	7	21	8.67(3.06)	21	20	6.67(1.16)	6
Current Activity Satisfaction	2	1.67(0.58)	2	2	2(1)	2	2	2.33(0.58)	2

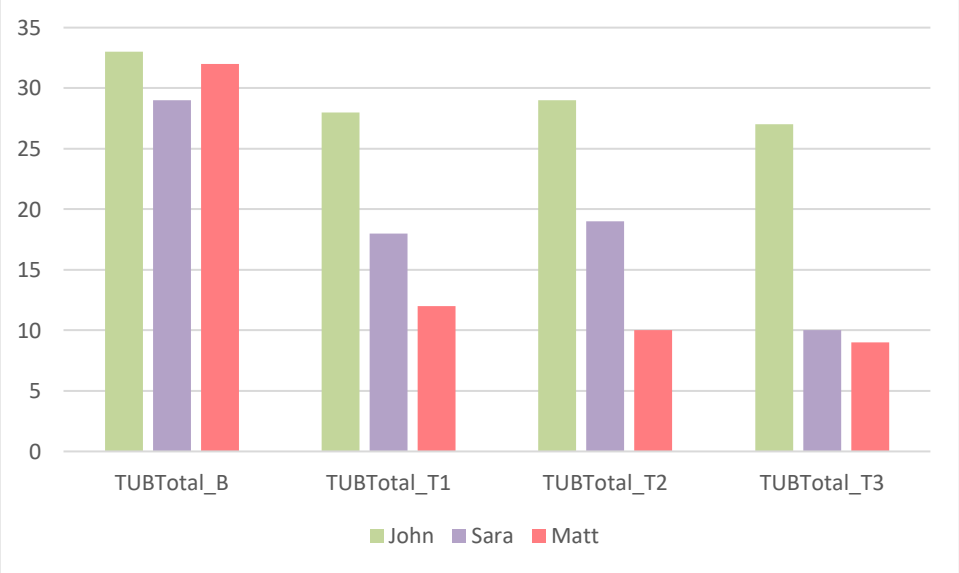
BCIS – Beck Cognitive Insight Scale; CAINS – Clinical Assessment Interview for Negative Symptoms; EXP – Experiential deficits; MAP – Motivation and Pleasure deficits; SNS – Self-evaluation of Negative Symptoms scale; MAS-A;

Metacognition Assessment Scale (Adapted); RR – Researcher Rated; TR – Therapist Rated; PSP – Personal and Social Performance scale; QPR – Questionnaire about the Process of Recovery scale; TUB – Time Use Budget

2.3.3 Data quality

All participants provided feedback that the data collection measures were fatiguing and repetitive. Coefficients of variability for self-report data were low (lower variability than the mean response given) for all participants, with variability for John being 0 across all but three measurements. Discussion and observations during data collection did not indicate simple lack of engagement. Response invariability means regression models are less likely to be informative, and a homogenous response style can also impact the reliability of observed statistical relationships between variables. Similarly, all participants reported fewer activities on the TUB at each timepoint (see Figure 2.1). It is less likely that participants, who continued to take part in a rehab programme in which keyworker rated scores of functioning did not indicate overall decline, were participating in less activity, but rather that these activities were no longer being reported despite maximal prompting from researchers.

Fig. 2.1 Number of Activities Reported in Time Use Budget (TUB) across time



Additionally, all participants reported struggling with motivation to wear actigraph watches for a sustained period but could not clearly articulate what they found difficult. The non-wear percentage for all participants was high, both on individual days where the watches were worn and across the study period (ranging from 35.04-100%) with John having no suitable actigraphy data. Figures 2.2-2.4 respectively show graphed output for wear periods for Sara's baseline and intervention phase, and baseline for Matt. These were used to assess any possible misattributions of the algorithm.

The data for both participants demonstrate patterns of light to moderate activity in some phases, however, for Matt in particular the timing of the observed periods and their relative inactivity are estimated to reflect inaccurate coding of wear periods which were actually non-wear. Additionally, where Sara appears to have worn the actigraph watch overnight, it appears that the time when it was coded as removed (approximately 9.30am) has resulted in bouts of sedentary behaviour prior to this being recorded as sleep, which has skewed the available sleep data. Actigraphy data are therefore unlikely to be a meaningful reflection of the true outcomes, and particularly for Matt where no intervention phase data was collected, did not have sufficient datapoints for meaningful interpretation. The subsequent reporting and analyses of data are therefore to demonstrate application of the analyses methods and whether these findings might signal the relevance of future similar analyses in a novel sample.

Fig. 2.2 Baseline accelerometer data for Sara

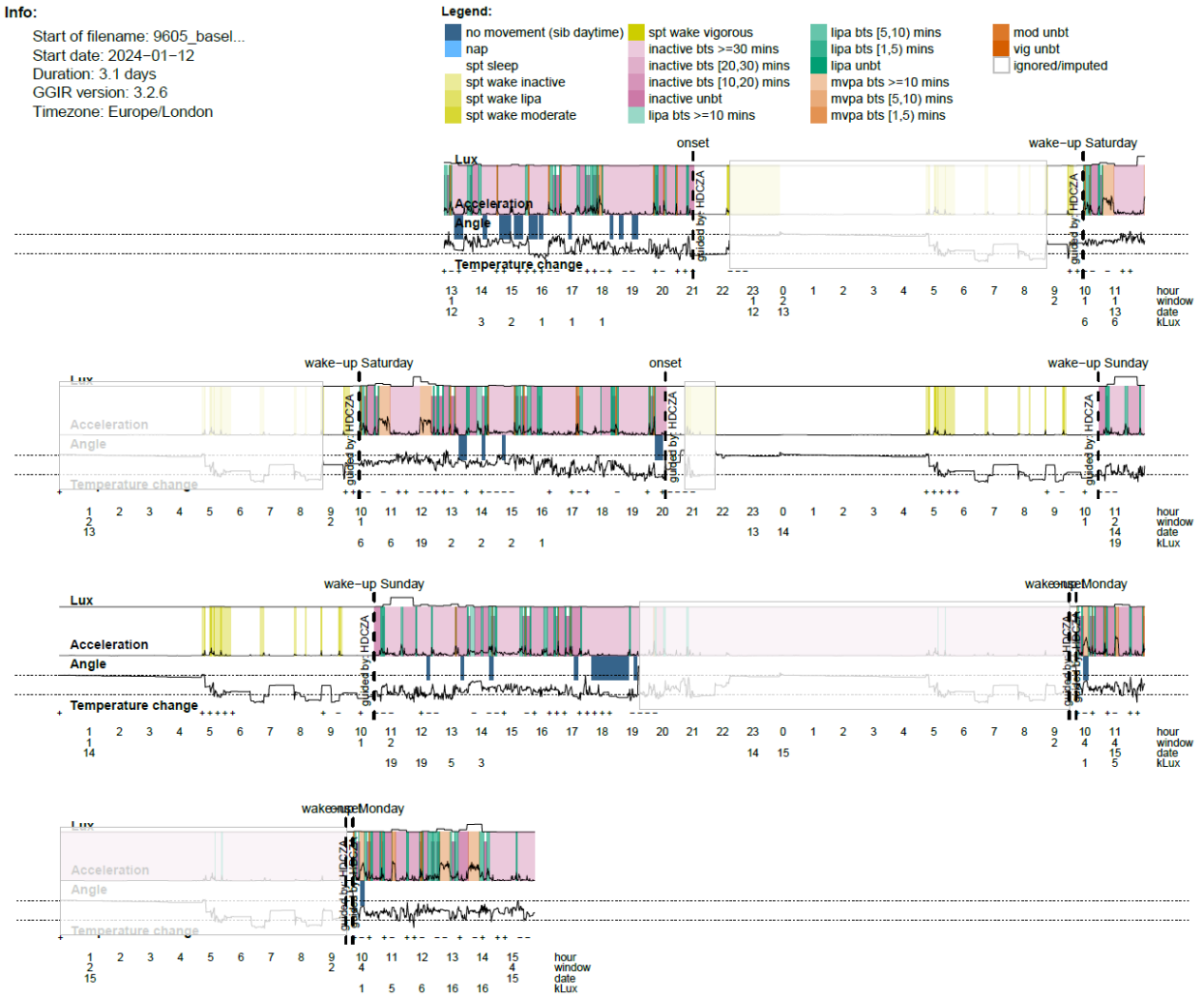


Fig. 2.3 Intervention phase accelerometer data for Sara

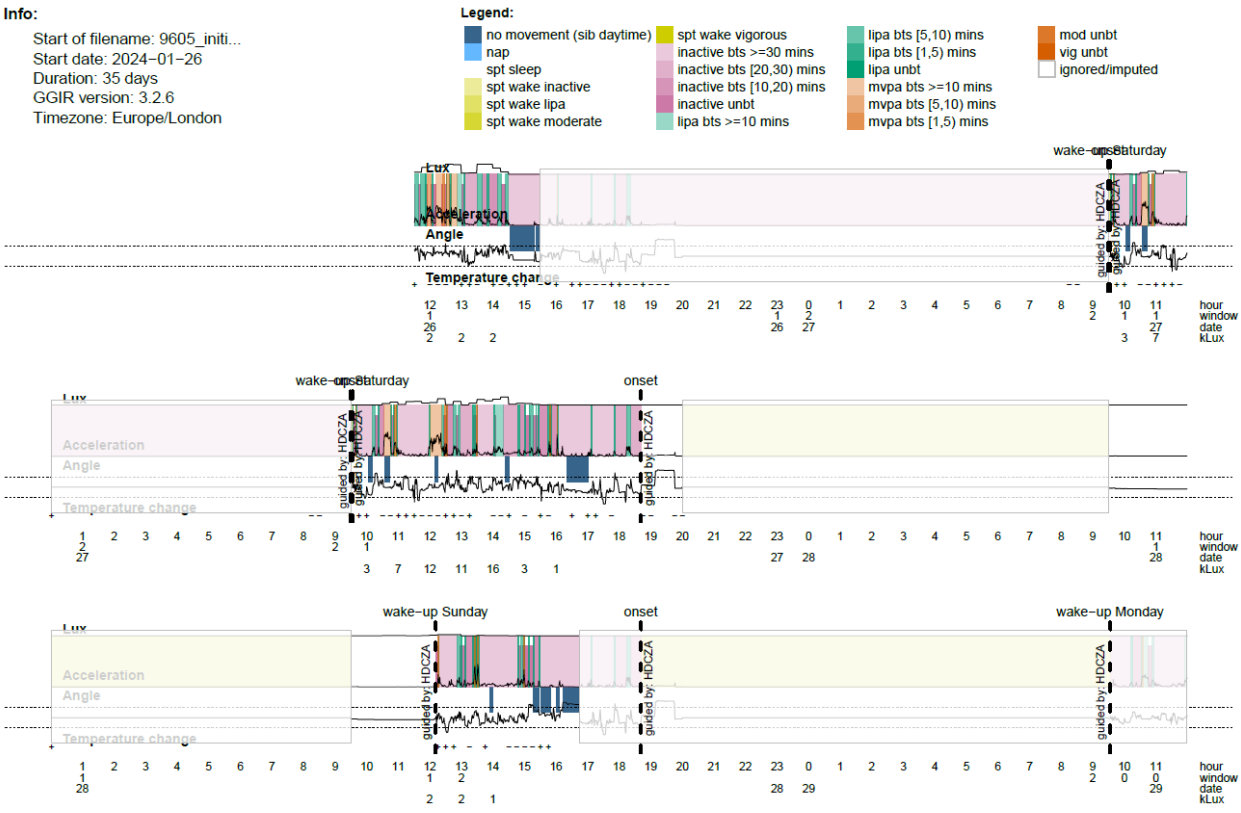


Fig. 2.4 Baseline accelerometer data for Matt

2.3.4 Visual analyses

Visual analyses were plotted using the scPlot R package (Wilbert, 2025). To conserve space, only the main analytical predictors and average level of daily light activity (as an example of actigraphy outcomes) are displayed in figures 2.5-2.8 for the baseline (A) and Intervention (B) phases, with the plots for all other variables included in Appendix 2.6.

Fig. 2.5 Weighted average activity levels from Time Use Budget

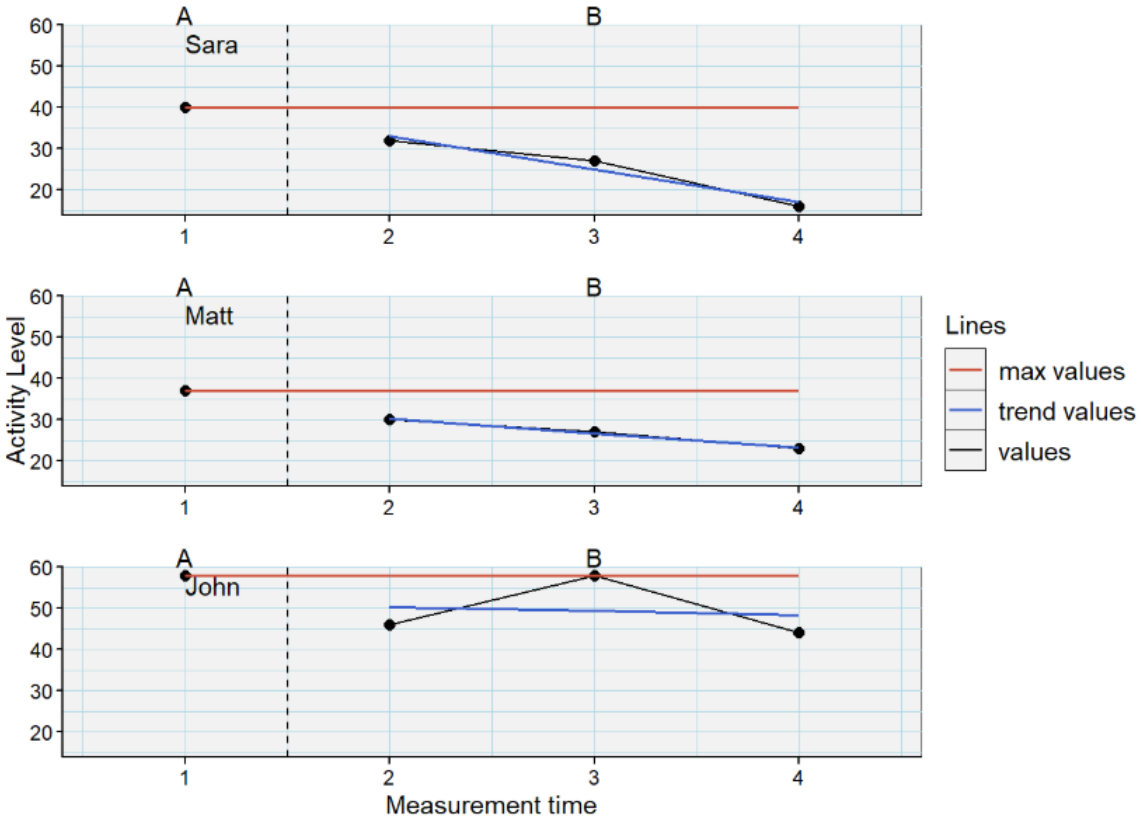


Fig. 2.6 Total metacognition (researcher-rated)

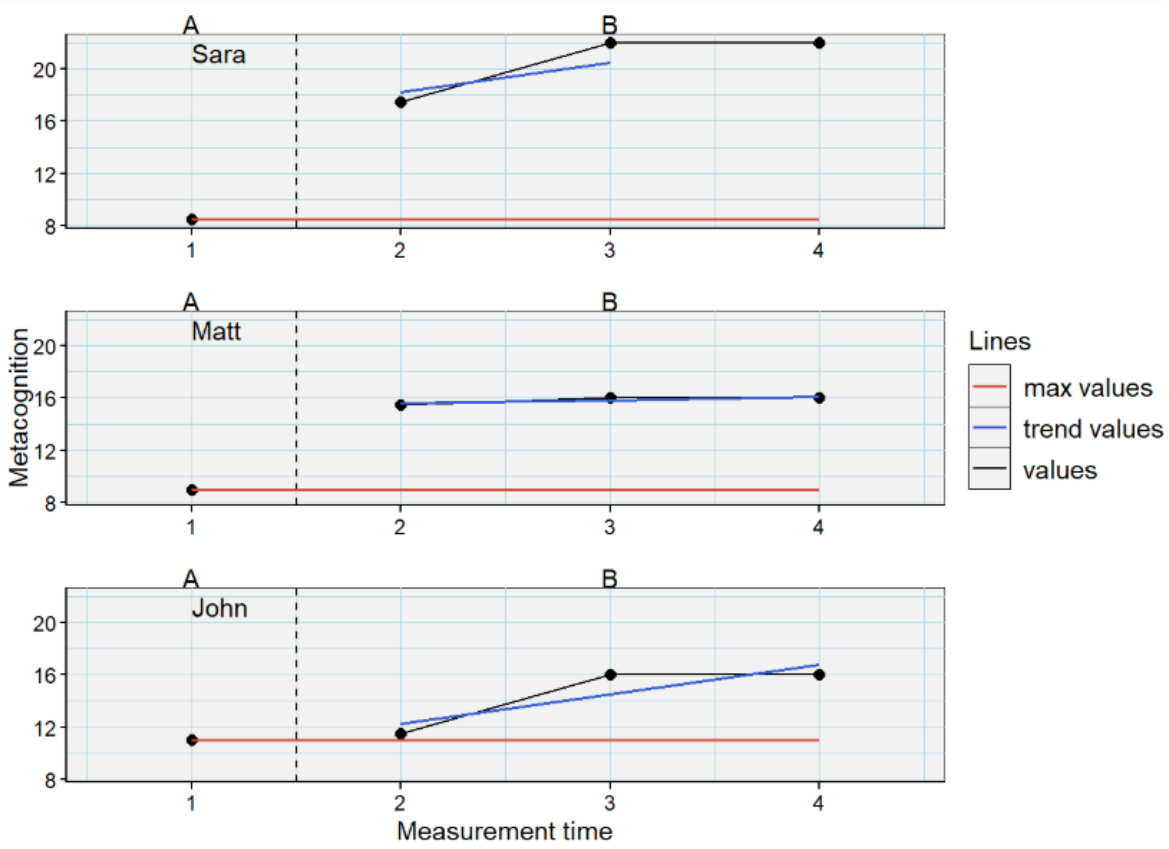


Fig. 2.7 Total CAINS

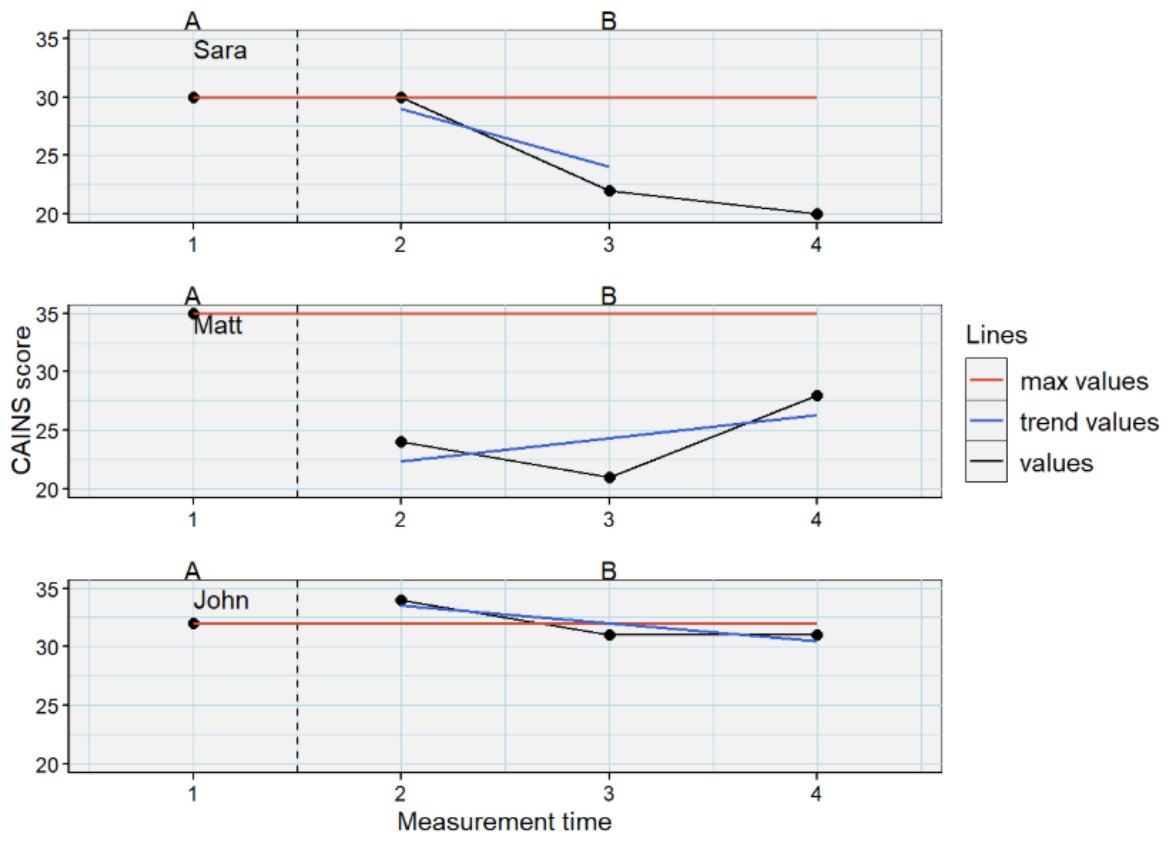
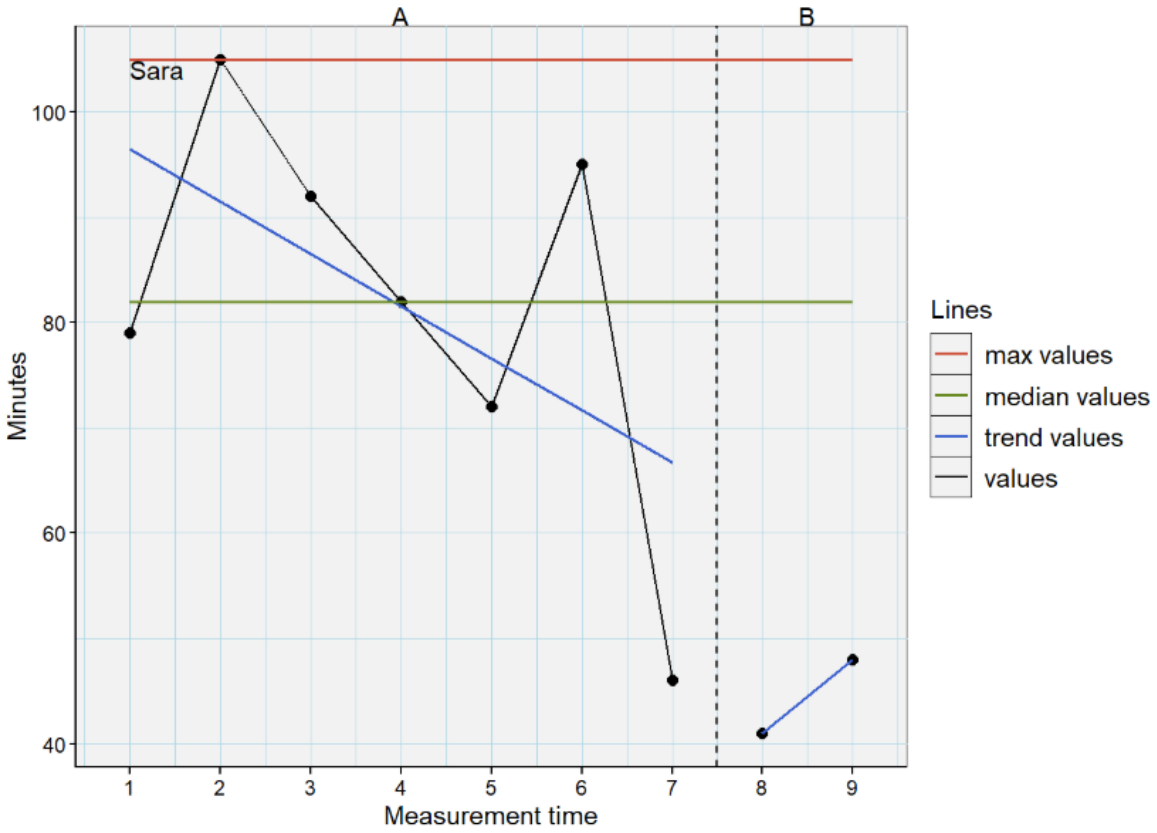


Fig. 2.8 Average amount of light activity minutes per day



Variables that showed improvements compared to baseline included the BCIS, SNS, CAINS, researcher-rated MAS-A, and the frequency of level 0 activity in the TUB (periods of no activity i.e. sitting thinking). TUB variables including satisfaction with activity (past week), and weighted average of activity levels, showed changes oppositional to the intended effect for all participants (e.g. less satisfaction and lower activity). Notably the QPR showed no substantial change upon visual inspection. Of actigraphy data for Sara (containing only two intervention datapoints), all variables demonstrated change opposite to the expected effect excluding sleep duration which gave an implausible result (sleep duration of one hour or less).

Although there are no baseline comparison data, reviewing the change therapist ratings of participant metacognition (Figure 2.9) suggests that only John experienced a clear improvement in metacognition. Participant- and therapist- rated therapeutic alliance (Figure 2.10) suggested that therapeutic alliance did not meaningfully reduce over the course of therapy, and in the case of John increased. For PQRST outcomes, only motivation for activity (2 points); and sleep (1 point) for Sara; and depression (1 point, but then returned to baseline rating) showed any positive change for participants (full data Appendix 2.7). Given the attendance patterns, change in several outcomes, and positive

ratings of therapeutic alliance, it is possible that participants did indeed find the intervention acceptable.

Fig. 2.9 Therapist-rated metacognition

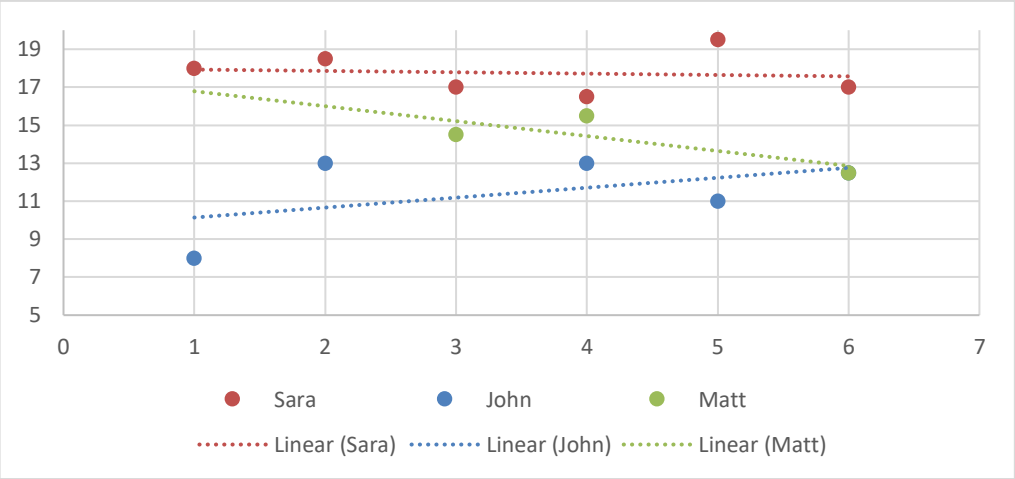
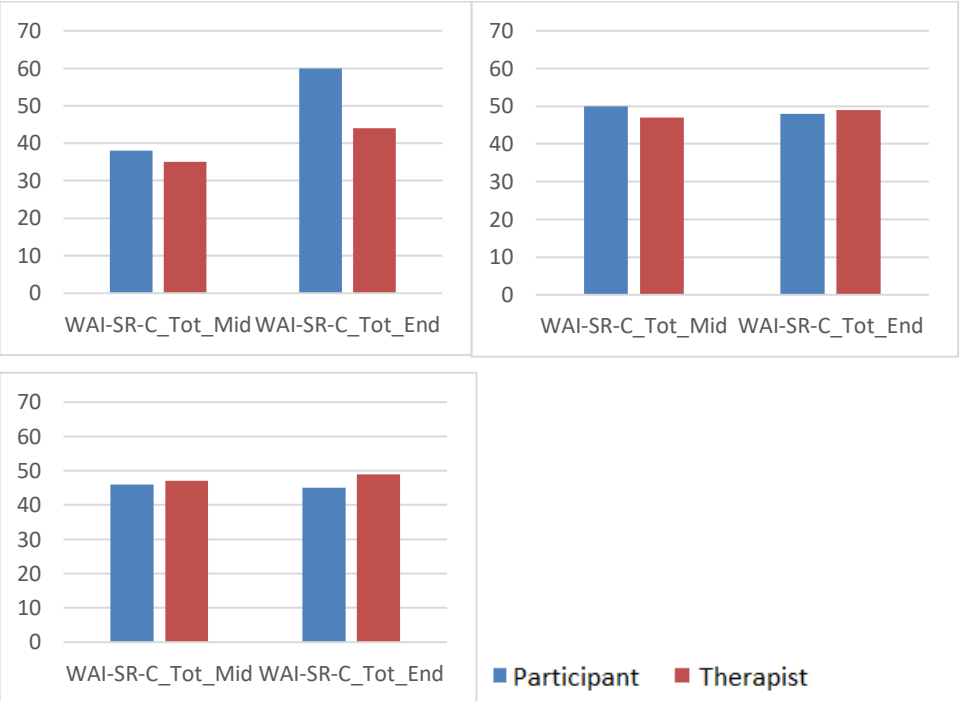


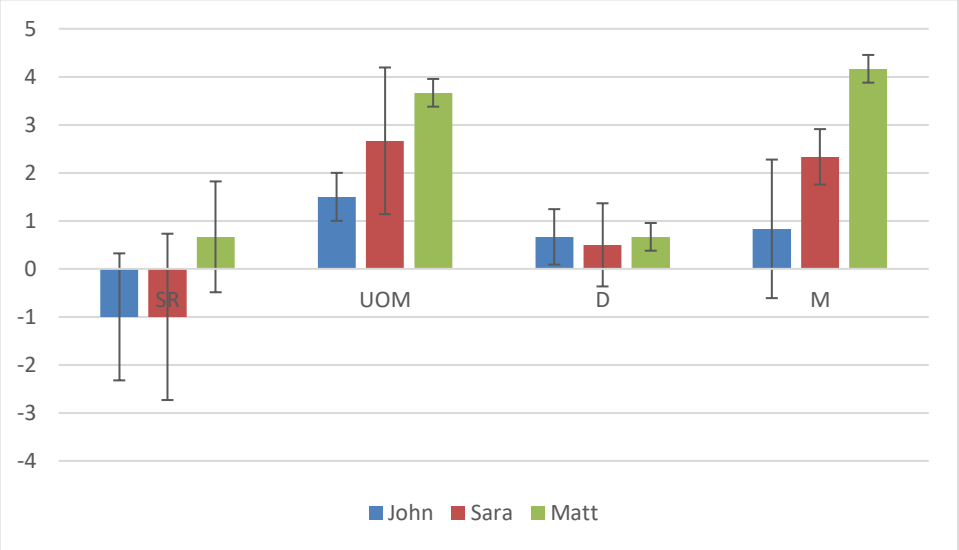
Fig. 2.10 Therapeutic alliance



The metacognition ratings of both therapist and the primary researcher were highly and significantly correlated, although the researcher rated data showed clearer improvement and was more strongly correlated with other outcomes (time spent with others and CAINS MAP). This is suggestive that it is indeed possible to rate metacognition with therapeutic transcripts. Therapist metacognition was demonstrated above the two-point tolerance threshold compared to participant ratings for one third of the data (maximum difference four points). Metacognitive discrepancies (Figure 11) outwith tolerance were

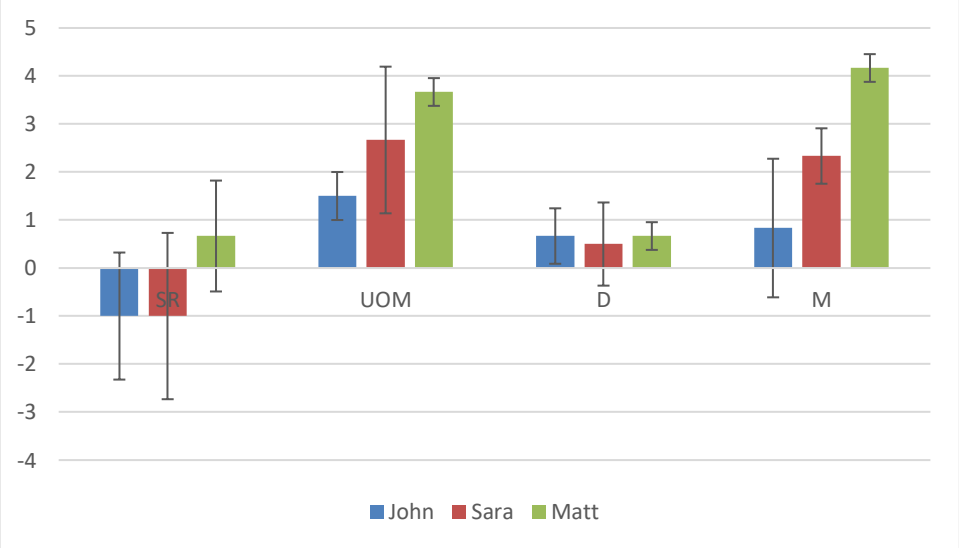
more common for the understanding others' minds, and mastery subscales, and in the case of self-reflectivity, participants occasionally demonstrated higher levels of metacognition than the therapist, perhaps appropriately given that participants would be engaged in greater levels of self-disclosure. All decentration ratings were within tolerance. When the variance was accounted for, only clear differences across time were that the therapist demonstrated greater self-reflectivity in early sessions (where therapist expectations were discussed as part of contracting); and decentration appeared to become less discrepant over the course of therapy (Figure 12).

Fig. 2.11 Metacognitive discrepancies (therapist versus participant) by domain



SR – Self-Reflectivity; UOM – Understanding Others' Minds; D – Decentration; M - Mastery

Fig. 2.12 Metacognitive discrepancies over time



SR – Self-Reflectivity; UOM – Understanding Others' Minds; D – Decentration; M - Mastery

2.3.5 Statistical analyses

Common SCED metrics are summarised in Tables 2.6 (actigraphy data) and 2.7 (all other data). For most outcomes a degree of non-overlap is evident, which generally corroborated visual analyses. The MBLR data was not always concordant with visual analyses, for example citing that expressive deficits have not changed in the expected direction for the intervention. This may be because these metrics fail to reflect tied datapoints between baseline and intervention and discrepancies across participants. Only change in metacognition was statistically significant ($X^2(1, N=3) = 12.00$, p = 0.01, Fisher's exact test $p = 0.01$). No randomisation tests were significant.

Effect sizes estimate of the strength of relationships observed included Cohen's *d,* Glass' Δ and HPS and PHS *d* statistics (a statistically comparable metric to Cohen's *d*). They ranged from moderate to large, with only current activity satisfaction, actigraphy acceleration metrics, and the PSP showing small or inconsistent effect sizes across analyses. Notably these data were more intrinsically invariant (i.e. the satisfaction with current activity being a single 1-5 Likert scale item) or had less available datapoints than the other outcomes. Cohen's *d* and Glass' Δ were generally incompatible with each other; and the PHS *d* statistic was more conservative than these estimates. HPS and PHS *d* statistics were somewhat concordant*,* and concordance appeared impacted by levels of autocorrelation and within subject variance. These statistics, while having large effect sizes, also often had a similar size of standard errors to the computed statistic, and no models were found to be a significant improvement on the null model, indicating they were relatively uninformative findings.

Table 2.6: SCED metrics (actigraphy data – Sara only)

	Percentage of non-overlapping data	Percentage of all non-overlapping data	Mean BaseLine Reduction	Cohen's *d*	Glass' Δ	Randomisation Test (T, *p*)	PHS *d* (SE)	PHS β (magnitude of unit change)	Variance	
									Within case	Between case
Number of sedentary periods per day	0.00	50.00	60.42	0.74	0.71	-2.17, 1	0.84 (0.82)	2.67 (2.10)	6.81 (3.12)	1.24 (3.61)
Sustained inactivity bouts (day period)	50.00	75.00	218.24	0.55	2.02	-0.93, 1	-0.12 (0.87)	-0.44 (1.68)	6.74 (10.97)	0.00 (10.61)
Sustained inactivity bouts of minimum 15 minutes (day period)	0.00	75.00	469.10	0.65	3.64	-1.03, 1	-0.14 (1.14)	-0.62 (1.62)	8.16 (17.86)	0.00 (17.52)
Duration of estimated sleep period (hours)	0.00	75.00	-83.19	-1.67	-1.19	9.21, 0.2	-1.54 (0.68)	-10.81 (3.27)	41.09 (19.69)	0.00 (4.01)
Inactivity per day (minutes)	0.00	55.60	58.13	0.99	1.54	-356.57, 1	1.10 (0.89)	360.66 (245.94)	90609.61 (45233.64)	>0.00 (33505.14)
Light activity minutes per day	50.00	77.80	-45.45	-2.65	-1.94	37.07, 0.17	-0.89 (0.81)	-37.07 (19.15)	659.27 (270.86)	449.15 (755.13)
Moderate activity minutes per day	0.00	55.60	-68.18	-1.37	-0.98	23.57, 0.17	-0.87 (0.79)	-28.44 (10.43)	297.77 (128.74)	298.98 (453.49)
Average acceleration (least active 5 hours per day)	0.00	75.00	-57.89		-0.85	-1.38, 0.2	-0.03 (8.48)	-0.18 (1.37)	6.60 (37.55)	0.00 (35.52)
Average acceleration (most active 5 hours per day)	0.00	75.00	-50.05		-1.59	-19.18, 1	-0.58 (1.66)	-15.25 (9.32)	92.91 (48.85)	168.79 (246.31)

Table 2.7: SCED metrics (all other data)

	Percentage of non-overlapping data	Percentage of all non-overlapping data	Mean BaseLine Reduction			HPS *d* (SE)	HPS β	Auto-correlation	Within case variance	Sample variance	PHS *d* (SE)	PHS β	Variance	
			Sara	Matt	John								Within-case	Between-case
BCIS	77.78	83.30	23.81	61.91	20.00	0.43 (0.69)	5.556	0.24	241.44	131.25	0.37 (0.68)	4.72 (7.56)	134.59 (90.98)	0.00 (69.69)
CAINS EXP	66.67	66.67	-20.00	-56.67	-3.33	-0.79 (0.64)	-2.667	0.03	3.29	7.42	-0.86 (0.68)	-2.43 (1.45)	0.42 (0.63)	0.00 (5.89)
CAINS MAP	66.67	66.67	-20.00	-20.00	1.52	-0.72 (0.65)	-2.889	0.20	8.32	11.83	-0.37 (0.55)	-1.74 (1.77)	9.98 (12.36)	4.56 (13.73)
CAINS Total	77.78	66.70	-20.00	-30.48	0.00	-0.99 (0.69)	-5.556	0.15	16.15	25.58	-0.82 (0.67)	-4.46 (2.80)	21.71 (20.78)	1.76 (18.90)
SNS	88.89	83.30	-13.33	-100.00	-13.64	-0.51 (0.63)	-6.667	0.14	81.50	15.87	-0.59 (0.61)	-7.39 (2.86)	18.66 (9.46)	63.97 (67.34)
Total MAS-A - RR	100.00	100.00	141.18	75.93	31.82	2.10 (1.08)	7.444	0.41	5.15	8.77	1.75 (0.87)	6.28 (1.71)	9.72 (12.38)	0.00 (11.31)
PSP	44.44	50.00	4.85	-7.29	-5.98	-0.12 (0.65)	-2.222	-0.24	228.33	272.67	-0.28 (0.52)	-4.71 (7.47)	159.98 (84.82)	73.45 (89.71)
QPR	88.89	83.30	14.89	6.67	83.33	0.59 (0.68)	11.333	-0.28	174.33	28.16	0.61 (0.44)	10.20 (2.36)	45.02 (31.55)	121.16 (122.22)
Weighted TUB Activity	0.00	50.00	-37.50	-27.93	-14.94	-0.61 (0.63)	-11.333	0.04	184.42	46.95	-0.55 (0.65)	-11.04 (4.13)	42.01 (33.60)	153.45 (169.51)
TUB Level 0 Activity	88.89	83.30	-47.62	100.00	-100.00	-0.62 (0.59)	-2.111	-0.12	8.00	3.83	-0.37 (0.45)	-1.29 (0.97)	2.60 (1.36)	5.08 (5.36)
TUB Level 1 Activity	22.22	50.00	-52.38	-98.55	-26.19	-0.83 (0.79)	-8.889	-0.07	13.39	58.58	-1.22 (0.66)	-10.33 (3.62)	33.30 (16.86)	22.12 (26.51)
TUB Level 2 Activity	11.11	50.00	-63.89	-47.62	-26.67	-1.32 (0.81)	-4.56	-0.17	9.67	9.98	-1.30 (0.74)	-4.53 (2.04)	9.26 (5.08)	1.01 (3.92)
TUB Level 3 Activity	66.67	83.30	44.44	Inf	-12.50	0.60 (0.91)	2.22	-0.23	0.49	5.58	0.71 (0.62)	2.03 (1.19)	4.93 (6.80)	0.86 (6.64)

Table 2.7 Contd.

	Percentage of non-overlapping data	Percentage of all non-overlapping data	Mean BaseLine Reduction			HPS *d* (SE)	HPS β	Auto-correlation	Within case variance	Sample variance	PHS *d* (SE)	PHS β	Variance	
			Sara	Matt	John								Within-case	Between-case
Current Activity Satisfaction	22.22	66.70	-16.67	16.67	0.00	0.00 (0.70)	0.000	-0.03	0.54	0.42	-0.08 (0.65)	-0.06 (0.40)	0.40 (0.20)	0.00 (0.07)
TUB No of activities with individuals	0.00	50.00	-62.75	-66.67	-58.73	-4.39 (1.59)	-12.11	0.31	11.80	6.00	-3.82 (1.28)	-11.59 (1.74)	7.57 (5.91)	0.00 (4.83)

BCIS – Beck Cognitive Insight Scale; CAINS – Clinical Assessment Interview for Negative Symptoms; EXP – Experiential deficits; MAP – Motivation and Pleasure deficits; SNS – Self-evaluation of Negative Symptoms scale; MAS-A – Metacognition Assessment Scale (Adapted); RR- Researcher Rated; PSP – Personal and Social Performance scale; QPR – Questionnaire about the Process of Recovery scale; TUB – Time Use Budget

Multiple imputation was used to examine whether metacognitive changes predicted changes in actigraphy as a proxy measure for negative symptoms over the course of treatment. The imputed means were all generally plausible and somewhat comparative to the original data only with standard errors that were either surprisingly invariant, or implausibly large (see Appendix 2.8). The imprecision of these estimates reflects the extent of missing data. The regression output is summarised in Table 2.9. Only treatment and metacognition showed significant effects in fixed effect regression models for the number of sedentary periods per day and duration of sedentary periods lasting 15 minutes or longer in the day period (summarised Table 2.8, for all other regression results see Appendix 2.9). The estimated effects of metacognition were small, and similar to other analyses, it is likely that these estimates are not truly informative.

Table 2.8: Regression output

Variable	Phase (CI)	Total metacognition (researcher rated; CI)	Case
Number of sedentary periods per day (A)	-0.78(-4.87-3.31)	**0.37(0.07-0.66), p=0.02**	0.72(-0.50-1.94)
Sustained inactivity bouts of minimum 15 minute (day period; B)	**-2.64 (-4.82- -0.46), p=0.02**	**0.23(0.04-0.41), p=0.02**	0.13(-0.68-0.95)

A - Comparison with model of treatment effect only: F(1,22.82) =6.68, p=0.02, comparison with null model: F(2,29.19) =6.65, p>0.01.

B - Comparison with model of treatment effect only: F(1,18.88) =6.81, p=0.02, comparison with null model: F(2,23.42) =3.41, p=0.05.

2.4 Discussion

This study assessed the acceptability, feasibility and effectiveness of a six-session psychotherapy for negative symptoms. Participant engagement and retention as well as therapeutic alliance ratings suggest that the therapy protocol was acceptable and the majority of outcome measures and sessions delivered appeared feasible. There was some preliminary evidence of effectiveness including change in the expected direction for several of the main outcome variables (i.e. metacognition, negative symptoms and participant-identified psychological problems). Rating the MAS-A via therapy transcripts also proved feasible and ratings appeared to be robust given their correlation with other metacognition measures and theoretically consistent associations with other variables of interest. There were some discrepancies in metacognitive matching to the participants,

and these may be explained by differences in level of self-disclosure, and perhaps true difficulties in only demonstrating understanding of others' minds and mastery proximal to participants levels. It is unclear from these preliminary data whether adherence to this tolerance threshold would have significantly improved therapy outcomes.

The rigour and comprehensiveness of this evaluation was limited due to several factors. Actigraphy data, the main outcome measured with frequency to meet SCED requirements, was largely unavailable, and there were queries about the reliability of some of the raw data. This may be explained by participant burden, as was reflected in response invariance across repeated measures. These data collection issues also impacted ability of statistical models to reliably detect effects. Some outcomes which showed changes in an unanticipated direction (e.g. reductions as opposed to increased actigraphy data) are difficult to evaluate due to the imbalance of available data compared to baseline.

These results are somewhat comparable with the evidence base, where brief (8-15 sessions or less) therapies have been found to be similarly effective as other lengths of treatment for negative symptoms (Riehle et al., 2020; Xu & Xu, 2024). Brief treatments may be effective provided they adequately target a specific cognitive or behavioural difficulty (Brabban et al., 2009), and are perhaps more likely to be successful when a variety of change methods are used (Velthorst et al., 2015). The change in relevant constructs was modest in some instances in this study, perhaps indicating that more sessions to create change in these relatively stable constructs (McLeod et al., 2014), or to build engagement and a more individualised preliminary formulation may be of benefit as indicated in other psychosis-specific CBT protocols (Wood et al., 2025). Results also varied across individuals; a larger pool of participants in similar future studies would help to explore relationships between constructs that might serve as predictors for change. It would also be helpful to explore how these data may have interacted with passive ambulatory data, including actigraphy and also perhaps mobile phone use (Daniel et al., 2023).

2.4.1 Clinical and research implications

These results showed some clinical change despite data collection difficulties. This indicates the relevance and importance of conducting psychological therapies with this

clinical group who are often labelled as ‘treatment-resistant’, and for symptoms which have shown a non-uniform impact of psychological therapy (Salahuddin et al., 2024). The evidence base for brief, inpatient, and negative symptom specific treatments are all more-so limited than other psychosis interventions and therefore this research is an important contribution. The pattern of challenges in data gathering (reduced response variability from repeated measures, and actigraphy nonadherence), offer insight into the factors that might improve data quality in subsequent studies. Researchers should pay close attention research protocol burden and factors to improve this including researcher engagement and frequency of contact (Robiner, 2005). Use of momentary and sampling assessment methodologies may also help increase response variability by offering a shorter recall period for relevant information and potentially shorter lengths of assessment (Bell et al., 2024).

2.4.2 Strengths and Limitations

This was a comprehensive study using SCED guidelines in the design and analyses, allowing it to address the relevant research questions, particularly regarding the main aims of exploring acceptability and feasibility of the intervention and data collection methods. Compared with a large-scale trial, the small sample size is ethically beneficial in minimising the number of participants completing the protocol before any issues with acceptability, feasibility or the intervention’s efficacy might have been identified and incorporated into any further iterations of the protocol. Additionally, the study findings demonstrate the capacity for SCED studies to be conducted in this population with limited clinical and financial resources, suggesting it can be a cost-effective and evidence-based method for clinicians and researchers to consider in future studies.

The data collection plan was limited from the outset, with actigraphy being the only measure capable of demonstrating baseline trend for comparison with actual intervention data as per SCED guidelines, and only one baseline datapoint for all other measures. Given participants were participating in other therapeutic activities in baseline, there were anticipated to be trend in outcome data towards the desired effect (i.e. downward for negative symptoms, upward for activity outcomes). It was also anticipated that therapy effects may be small to moderate, creating overlap between baseline and intervention phases. Therefore, nonoverlap indices and descriptive indices for changes in level and slope, including standardised mean differences, were expected to be

suboptimal for establishing true intervention effects. Regression and randomisation tests were anticipated to be more relevant.

All statistical analyses were rendered underpowered by unavailability of sufficient datapoints: this was anticipated for the majority of outcomes where these outcomes are thought to be relatively stable (McLeod et al., 2014) and analyses were intended to be supplemental, but was particularly detrimental in regards to actigraphy data. Had analyses been sufficiently powered to conduct further multiple comparisons, the impact of utilising self-reported negative symptoms as opposed to the CAINS may have been of interest.

The use of different raters who did not overlap on any outcome assessments limited calibration and inter-rater agreement assessment, and there were discrepancies between CAINS ratings for one participant (9 points from baseline to intervention start point; Matt) which were outliers, and might have been explained more by use of different raters than true symptom variation. Given the IPII is a less elaborated narrative due to limited use of prompts, and generates shorter speech samples than therapy transcripts, it is possible that this inflated intervention effects on metacognition.

2.4.3 Conclusions

As a contribution to the preliminary evidence base about the effectiveness of targeting metacognition for treatment of negative symptoms, this study demonstrates some signals that this is a relevant treatment target worthy of further consideration. The impact of methodologies on the data collected serves as important learning for researchers considering effective ways of generating reliable data in this population.

2.5 References

ActivInsights Ltd. (2025). Digital Health Wearable for Clinical Trials. https://activinsights.com/clinical-trials-digital-health-technology/digital-health-technology-wearables/geneactiv-raw-data-wearable/

Audigier, V., & Resche-Rigon, M. (2023). *MICEMD: Multiple Imputation by Chained Equations with Multilevel Data.*

Aydin, O. (2024). A description of missing data in single-case experimental designs studies and an evaluation of single imputation methods. *Behavior Modification, 48*(3), 312-359. https://doi.org/10.1177/01454455241226879

Barch, D. M., Gold, J. M., & Kring, A. M. (2017). Paradigms for assessing hedonic processing and motivation in humans: Relevance to understanding negative symptoms in psychopathology. *Schizophrenia Bulletin, 43*(4), 701-705. https://doi.org/10.1093/schbul/sbx063

Beck, A. T., Baruch, E., Balter, J. M., Steer, R. A., & Warman, D. M. (2004). A new instrument for measuring insight: the Beck Cognitive Insight Scale. *Schizophrenia Research, 68*(2-3), 319-329. https://doi.org/10.1016/s0920-9964(03)00189-0

Beck, A. T., & Rector, N. A. (2005). Cognitive approaches to schizophrenia: Theory and therapy. *Annual Review of Clinical Psychology, 1*, 577-606. https://doi.org/10.1146/annurev.clinpsy.1.102803.144205

Bell, I. H., Eisner, E., Allan, S., Cartner, S., Torous, J., Bucci, S., & Thomas, N. (2024). Methodological characteristics and feasibility of ecological momentary assessment studies in psychosis: A systematic review and meta-analysis. *Schizophrenia Bulletin, 50*(2), 238-265. https://doi.org/10.1093/schbul/sbad127

Bennouna-Greene, M., Berna, F., Conway, M. A., Rathbone, C. J., Vidailhet, P., & Danion, J. M. (2012). Self-images and related autobiographical memories in schizophrenia. *Consciousness and Cognition, 21*(1), 247-257. https://doi.org/10.1016/j.concog.2011.10.006

Brabban, A., Tai, S., & Turkington, D. (2009). Predictors of outcome in brief cognitive behavior therapy for schizophrenia. *Schizophrenia Bulletin, 35*(5), 859-864. https://doi.org/https://doi.org/10.1093/schbul/sbp065

Brett-Jones, J., Garety, P., & Hemsley, D. (1987). Measuring delusional experiences - A method and its application. *British Journal of Clinical Psychology, 26*(4), 257-265. https://doi.org/10.1111/j.2044-8260.1987.tb01359.x

Daniel, D. G., Cohen, A. S., Velligan, D., Harvey, P. D., Alphs, L., Davidson, M., Potter, W., Kott, A., Schooler, N., Brodie, C. R., Moore, R. C., Lindenmeyer, P., & Marder, S. R.

(2023). Remote assessment of negative symptoms of schizophrenia. *Schizophrenia bulletin open, 4*(1), 1-9. https://doi.org/https://doi.org/10.1093/schizbullopen/sgad001

Docx, L., Sabbe, B., Provinciael, P., Merckx, N., & Morrens, M. (2013). Quantitative psychomotor dysfunction in schizophrenia: A loss of drive, impaired movement execution or both? *Neuropsychobiology, 68*(4), 221-227. https://doi.org/10.1159/000355293

Elis, O., Caponigro, J. M., & Kring, A. M. (2013). Psychosocial treatments for negative symptoms in schizophrenia: Current practices and future directions. *Clinical Psychology Review, 33*(8), 914-928. https://doi.org/10.1016/j.cpr.2013.07.001

Hatcher, R. L., & Gillaspy, J. A. (2006). Development and validation of a revised short version of the Working Alliance Inventory. *Psychotherapy research, 16*(1), 12-25. https://doi.org/https://doi.org/10.1080/10503300500352500

Hatcher, R. L., Lindqvist, K., & Falkenström, F. (2020). Psychometric evaluation of the Working Alliance Inventory—Therapist version: Current and new short forms. *Psychotherapy research, 30*(6), 706-717. https://doi.org/https://doi.org/10.1080/10503307.2019.1677964

Hazell, C. M., Hayward, M., Cavanagh, K., & Strauss, C. (2016). A systematic review and meta-analysis of low intensity CBT for psychosis. *Clinical Psychology Review, 45*, 183-192. https://doi.org/10.1016/j.cpr.2016.03.004

Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2013). A standardized mean difference effect size for multiple baseline designs across individuals. *Research Synthesis Methods, 4*(4), 324-341.

Jolley, S., Garety, P. A., Ellett, L., Kuipers, E., Freeman, D., Bebbington, P. E., Fowler, D. G., & Dunn, G. (2006). A validation of a new measure of activity in psychosis. *Schizophrenia Research, 85*(1-3), 288-295. https://doi.org/10.1016/j.schres.2006.03.012

Kaiser, S., Heekeren, K., & Simon, J. J. (2011). The negative symptoms of schizophrenia: Category or continuum? *Psychopathology, 44*(6), 345-353. https://doi.org/10.1159/000325912

Kaiser, S., Lyne, J., Agartz, I., Clarke, M., Mørch-Johnsen, L., & Faerden, A. (2017). Individual negative symptoms and domains - Relevance for assessment, pathomechanisms and treatment. *Schizophrenia Research, 186*, 39-45. https://doi.org/10.1016/j.schres.2016.07.013

Krasny-Pacini, A., & Evans, J. (2017). Single-case experimental designs to assess intervention effectiveness in rehabilitation: A practical guide. *Annals of Physical and Rehabilitation Medicine, 61*(3). https://doi.org/https://doi.org/10.1016/j.rehab.2017.12.002

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education, 34*(1), 26-38. https://doi.org/https://doi.org/10.1177/0741932512452794

Kring, A. M., & Barch, D. M. (2014). The motivation and pleasure dimension of negative symptoms: neural substrates and behavioral outputs. *European Neuropsychopharmacology, 24*(5), 725-736. https://doi.org/10.1016/j.euroneuro.2013.06.007

Kring, A. M., Gur, R. E., Blanchard, J. J., Horan, W. P., & Reise, S. P. (2013). The Clinical Assessment Interview for Negative Symptoms (CAINS): Final development and validation. *American Journal of Psychiatry, 170*(2), 165-172. https://doi.org/10.1176/appi.ajp.2012.12010109

Lincoln, T. M., Riehle, M., Pillny, M., Helbig-Lang, S., Fladung, A. K., Hartmann-Riemer, M., & Kaiser, S. (2017). Using functional analysis as a framework to guide individualized treatment for negative symptoms. *Frontiers in Psychology, 8*(2108), 1-15. https://doi.org/10.3389/fpsyg.2017.02108

Luther, L., Firmin, R. L., Minor, K. S., Vohs, J. L., Buck, B., Buck, K. D., & Lysaker, P. H. (2016). Metacognition deficits as a risk factor for prospective motivation deficits in

schizophrenia spectrum disorders. *Psychiatry Research, 245*, 172-178. https://doi.org/10.1016/j.psychres.2016.08.032

Lysaker, P. H., Carcione, A., Dimaggio, G., Johannesen, J. K., Nicolò, G., Procacci, M., & Semerari, A. (2005). Metacognition amidst narratives of self and illness in schizophrenia: Associations with neurocognition, symptoms, insight and quality of life. *Acta Psychiatrica Scandinavica, 112*(1), 64-71. https://doi.org/10.1111/j.1600-0447.2005.00514.x

Lysaker, P. H., Clements, C. A., Plascak-Hallberg, C. D., Knipscheer, S. J., & Wright, D. E. (2002). Insight and personal narratives of illness in schizophrenia. *Psychiatry: Interpersonal and Biological Processes, 65*(3), 197-206. https://doi.org/https://doi.org/10.1521/psyc.65.3.197.20174

Lysaker, P. H., Gumley, A., Luedtke, B., Buck, K. D., Ringer, J. M., Olesek, K., Kukla, M., Leonhardt, B. L., Popolo, R., & Dimaggio, G. (2013). Social cognition and metacognition in schizophrenia: Evidence of their independence and linkage with outcomes. *Acta Psychiatrica Scandinavica, 127*(3), 239-247. https://doi.org/10.1111/acps.12012

Lysaker, P. H., & Klion, R. E. (2017). *Recovery, meaning-making, and severe mental illness: A comprehensive guide to metacognitive reflection and insight therapy*. Routledge. https://doi.org/https://doi.org/10.4324/9781315447001

Lysaker, P. H., Warman, D. M., Dimaggio, G., Procacci, M., LaRocco, V. A., Clark, L. K., Dike, C. A., & Nicolò, G. (2008). Metacognition in schizophrenia - Associations with multiple assessments of executive function. *Journal of Nervous and Mental Disease, 196*(5), 384-389. https://doi.org/10.1097/NMD.0b013e3181710916

Lysaker, P. H., Wiesepape, C. N., Hamm, J. A., Leonhardt, B. L., Hillis, J. D., Musselman, A., & Hasson-Ohayon, I. (2021). *Metacognition Assessment Scale – Abbreviated (MAS-A): A brief overview and coding manual*. Indiana University.

Maillard, P., Dimaggio, G., de Roten, Y., Berthoud, L., Despland, J. N., & Kramer, U. (2017). Metacognition as a predictor of change in the treatment for borderline personality disorder: A preliminary pilot study. *Journal of Psychotherapy Integration, 27*(4), 445-459. https://doi.org/10.1037/int0000090

Manolov, R., & Moeyaert, M. (2017). Recommendations for choosing single-case data analytical techniques. *Behavior Therapy, 48*(1), 97-114. https://doi.org/https://doi.org/10.1016/j.beth.2016.04.008

Manolov, R., Moeyaert, M., & Evans, J. (2016). Single-case data analysis: Software resources for applied researchers.

McLeod, H. J., Gumley, A., & Schwannauer, M. (2014). *The impact of metacognition on the development and maintenance of negative symptoms*. https://doi.org/10.1016/B978-0-12-405172-0.00007-7

Neil, S. T., Kilbride, M., Pitt, L., Nothard, S., Welford, M., Sellwood, W., & Morrison, A. P. (2009). The questionnaire about the process of recovery (QPR): A measurement tool developed in collaboration with service users. *Psychosis-Psychological Social and Integrative Approaches, 1*(2), 145-155. https://doi.org/10.1080/17522430902913450

NICE. (2014). Psychosis and schizophrenia in adults: Prevention and management. https://www.nice.org.uk/guidance/cg178

Oorschot, M., Lataster, T., Thewissen, V., Lardinois, M., Wichers, M., van Os, J., Delespaul, P., & Myin-Germeys, I. (2013). Emotional experience in negative symptoms of schizophrenia - No evidence for a generalized hedonic deficit. *Schizophrenia Bulletin, 39*(1), 217-225. https://doi.org/10.1093/schbul/sbr137

Patrick, D. L., Burns, T., Morosini, P., Rothman, M., Gagnon, D. D., Wild, D., & Adriaenssen, I. (2009). Reliability, validity and ability to detect change of the clinician-rated Personal and Social Performance scale in patients with acute symptoms of schizophrenia. *Current Medical Research and Opinion, 25*(2), 325-338. https://doi.org/10.1185/03007990802611919

Pedone, R., Semerari, A., Riccardi, I., Procacci, M., Nicolò, G., & Carcione, A. (2017). Development of a self-report measure of metacognition: The metacognition self-assessment scale (MSAS). Instrument description and factor structure. *Clinical Neuropsychiatry, 14*(3), 185-194. <Go to ISI>://WOS:000406301700001

Pinheiro J., Bates D., & Team, R. C. (2025). *nlme: Linear and Nonlinear Mixed Effects Models.* In R, package version 3.1-168.

Pustejovsky J., Chen M., & B., H. (2024). *scdhlm: Estimating Hierarchical Linear Models for Single-Case Designs.* In R, package version 0.7.3.

Pustejovsky, J. E., Hedges, L. V., & Shadish, W. R. (2014). Design-comparable effect sizes in multiple baseline designs: A general modeling framework. *Journal of Educational and Behavioral Statistics, 39*(5), 368-393.

Raphael, J., Hutchinson, T., Haddock, G., Emsley, R., Bucci, S., Lovell, K., Edge, D., Price, O., Udachina, A., Day, C., Cross, C., Peak, C., Drake, R., & Berry, K. (2021). A study on the feasibility of delivering a psychologically informed ward-based intervention on an acute mental health ward. *Clinical Psychology & Psychotherapy, 28*(6), 1587-1597. https://doi.org/https://doi.org/10.1002/cpp.2597

Riehle, M., Böhl, M. C., Pillny, M., & Lincoln, T. M. (2020). Efficacy of psychological treatments for patients with schizophrenia and relevant negative symptoms: A meta-analysis. *Clinical psychology in Europe, 2*(3), 1-23. https://doi.org/https://doi.org/10.32872/cpe.v2i3.2899

Robiner, W. N. (2005). Enhancing adherence in clinical research. *Contemporary clinical trials, 26*(1), 59-77. https://doi.org/https://doi.org/10.1016/j.cct.2004.11.015

Salahuddin, N. H., Schütz, A., Pitschel-Walz, G., Mayer, S. F., Chaimani, A., Siafis, S., Priller, J., Leucht, S., & Bighelli, I. (2024). Psychological and psychosocial interventions for treatment-resistant schizophrenia: A systematic review and network meta-analysis. *The Lancet Psychiatry, 11*(7), 545-553. https://doi.org/10.1016/S2215-0366(24)00136-6

Semerari, A., Carcione, A., Dimaggio, G., Falcone, M., Nicolò, G., Procacci, M., & Alleva, G. (2003). How to evaluate metacognitive functioning in psychotherapy? The metacognition assessment scale and its applications. *Clinical Psychology & Psychotherapy, 10*(4), 238-261. https://doi.org/https://doi.org/10.1002/cpp.362

Slocum, T. A., Pinkelman, S. E., Joslyn, P. R., & Nichols, B. (2022). Threats to internal validity in multiple-baseline design variations. *Perspectives on Behavior Science, 45*(3), 619-638.

Strauss, G. P., & Cohen, A. S. (2017). A transdiagnostic review of negative symptom phenomenology and etiology. *Schizophrenia Bulletin, 43*(4), 712-729. https://doi.org/10.1093/schbul/sbx066

Tate, R. L., Perdices, M., Rosenkoetter, U., Shadish, W., Vohra, S., Barlow, D. H., Horner, R., Kazdin, A., Kratochwill, T., McDonald, S., Simpson, M., Shamseer, L., Togher, L., Albin, R., Backman, C., Douglas, J., Evans, J. J., Gast, D., Manolov, R., Mitchell, G., Nickels, L., Nikles, J., Ownsworth, T., Rose, M., Schmid, C. H., & Wilson, B. (2016). The Single-Case Reporting guideline In Behavioural Interventions (SCRIBE) 2016 statement. *Physical therapy, 96*(7), e1-e10. https://doi.org/https://doi.org/10.2522/ptj.2016.96.7.e1

Tate, R. L., Perdices, M., Rosenkoetter, U., Wakim, D., Godbee, K., Togher, L., & McDonald, S. (2013). Revision of a method quality rating scale for single-case experimental designs and n-of-1 trials: the 15-item Risk of Bias in N-of-1 Trials (RoBiNT) Scale. *Neuropsychol Rehabil, 23*(5), 619-638. https://doi.org/10.1080/09602011.2013.824383

van Hees, V., Migueles, J., Fang, Z., Zhao, J., Heywood, J., Mirkes, E., & Sabia, S. (2025). *GGIR: Raw Accelerometer Data Analysis.* In R, package version 3.2-6.

Velligan, D. I., Mintz, J., Sierra, C., Martin, M. L., Fredrick, M., Maglinte, G. A., & Corey-Lisle, P. K. (2016). The Daily Activity Report (DAR) a novel measure of functional outcome for serious mental illness. *Schizophrenia Bulletin, 42*(3), 579-587. https://doi.org/10.1093/schbul/sbv185

Velthorst, E., Koeter, M., Van Der Gaag, M., Nieman, D., Fett, A.-K., Smit, F., Staring, A., Meijer, C., & de Haan, L. (2015). Adapted cognitive–behavioural therapy required for targeting negative symptoms in schizophrenia: meta-analysis and meta-regression. *Psychological medicine, 45*(3), 453-465. https://doi.org/10.1017/S0033291714001147

Wehr, S., Weigel, L., Davis, J., Galderisi, S., Mucci, A., & Leucht, S. (2024). Clinical Assessment Interview for Negative Symptoms (CAINS): a systematic review of measurement properties. *Schizophrenia Bulletin, 50*(4), 747-756.

White, R., Gumley, A., McTaggart, J., Rattrie, L., McConville, D., Cleare, S., McLeod, H., & Mitchell, G. (2015). Acceptance and commitment therapy for depression following psychosis: An examination of clinically significant change. *Journal of Contextual Behavioral Science, 4*(3), 203-209. https://doi.org/https://doi.org/10.1016/j.jcbs.2015.06.004

Whyte, R. (2025). *Exploring metacognition and social connection as mechanisms of change in individuals with negative symptoms of psychosis.* (Publication Number glathesis:2025-85423) University of Glasgow]. Glasgow.

Wilbert, J. (2025). *scplot - Single-Case Data Plots. (0.5.1).* In https://CRAN.R-project.org/package=scplot

Wilbert, J., & Lüke, T. (2025). *Scan: Single-Case Data Analyses for Single and Multiple Baseline Designs. (0.64.0).* In CRAN. https://CRAN.R-project.org/package=scan

Wood, L., Morrison, A. P., Birken, M., Dare, C., Guerin, E., Nyikavaranda, P., Malde-Shah, N., Persaud, K., Ford, P., & Nebo, C. (2025). Examining the feasibility of a crisis-focused Cognitive Behaviour Therapy (CBT)–informed psychological intervention for inpatients experiencing psychosis (the CRISIS study): Results from a pilot randomised controlled trial. *EClinicalMedicine, 86*(103380), 1-12. https://doi.org/10.1016/j.eclinm.2025.103380

World Health Organization. (1992). *ICD-10: International statistical Classification of Diseases and related health problems*. World Health Organization.

Xu, F., & Xu, S. (2024). Cognitive-Behavioral Therapy for negative symptoms of schizophrenia: A systematic review and meta-analysis. *Medicine, 103*(36), 1-7. https://doi.org/10.1097/MD.0000000000039572

Appendix 1.1

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	P10
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	P12
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	P13-14
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	P14-15
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	P16
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	P16-17
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Protocol (cited p15)
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	P17
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	P17-18
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Online resource (cited p17)
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Online resource (cited p17)

Section and Topic	Item #	Checklist item	Location where item is reported
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	P18-19
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	P19-20
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	P18-20
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	P29
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	P18
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	P19-20
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	P19-20
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	P19-20
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	P20
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	None
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	P20-22
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	P20, 22
Study characteristics	17	Cite each included study and present its characteristics.	P24-28
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	P36
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	P29-30, 32, 36-40
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	P36-40
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	P29-30, 36-40
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	p29-30, 36-

Section and Topic	Item #	Checklist item	Location where item is reported
			40
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	p29-30, 36-40
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	p29-30, 36-40
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	None
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	P41-42
	23b	Discuss any limitations of the evidence included in the review.	P41-43
	23c	Discuss any limitations of the review processes used.	P42-43
	23d	Discuss implications of the results for practice, policy, and future research.	P41-42
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	P15
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	P15
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	Pre-registration (cited p15)
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	P10
Competing interests	26	Declare any competing interests of review authors.	P11
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	P11

Appendix 1.2

1. Motivation for close family/spouse/partner relationships
2. Motivation for close friendships and romantic relationships
3. Frequency of pleasurable social activities – past week
4. Frequency of pleasurable social activities - next week
5. Motivation for work and school activities
6. Frequency of expected pleasurable work and school activities – next week
7. Motivation for recreational activities
8. Frequency of pleasurable recreational activities – past week
9. Frequency of expected pleasurable recreational activities – next week
10. Facial expression
11. Vocal expression
12. Expressive gestures
13. Quantity of speech

Appendix 1.3

```
##check WD correct

getwd()


## load necessary packages

library(tidyverse)

library(readxl)

library(metafor)

library(foreach)


##load data

metadata <- read_excel("Summarydata.xlsx")

metadata[metadata == ""] <- NA

View(metadata)


##Example of calculation for raw mean demographics:

CAINS_MN <- escalc(measure="MN", mi=CAINSM, sdi=CAINSSD, ni=SS, data=metadata)

print(CAINS_MN)

summary.escalc(CAINS_MN)

#meta-analysis

CAINSMN_WA <- rma.uni(yi, vi, data=CAINS_MN, method="REML")

print(CAINSMN_WA)

summary(CAINSMN_WA)


##Example of calculationh for proportional demographics:

MVF_PR <- escalc(measure = "PR", xi=MaleN, mi=FemaleN, data=metadata)

print(MVF_PR)

summary.escalc(MVF_PR)
```

```r
#meta-analysis

MVF_WA <- rma.uni(yi, vi, data=MVF_PR, method="REML")

print(MVF_WA)

summary(MVF_WA)


##Calculate Raw Cronbach's alpha:

RCA <- rma(measure="ARAW", ai=CA, mi=mi, ni=SS, data=metadata, method="REML",
test="knha")

print(RCA)

summary(RCA)

forest(RCA)

funnel (RCA)

rstudent(RCA)

infRCA <- influence(RCA)

infRCA

plot(infRCA, plotdfb= TRUE)

qqnorm(RCA, main="Cronbach's Alpha")

leave1out(RCA, digits=3)

regtest(RCA)

trimfill(RCA)


##Calculate Hakstian & Whalen, 1976 transformed CA:

HWCA <- rma(measure="AHW", ai=CA, mi=mi, ni=SS, data=metadata, method="REML",
test="knha")

print(HWCA)

summary(HWCA)

predict(HWCA, transf=transf.iahw)

forest(HWCA)

funnel (HWCA)

rstudent(HWCA)
```

```
infHWCA <- influence(HWCA)

infHWCA

plot(infHWCA, plotdfb= TRUE)

qqnorm(HWCA, main="Cronbach's Alpha")

leave1out(HWCA, digits=3)

regtest(HWCA)

trimfill(HWCA)


##Calculate Bonnett, 2002 transformed CA:

BCA <- rma(measure="ABT", ai=CA, mi=mi, ni=SS, data=metadata, method="REML",
test="knha")

print(BCA)

summary(BCA)

predict(BCA, transf=transf.iahw)

forest(BCA)

funnel (BCA)

rstudent(BCA)

infBCA <- influence(BCA)

infBCA

plot(infBCA, plotdfb= TRUE)

qqnorm(BCA, main="Cronbach's Alpha")

leave1out(BCA, digits=3)

regtest(BCA)

trimfill(BCA)


##Calculate ICC pooled estimate:

#normality distribution calculation

# (k is number observations of each object, n is number objects)

calc_icc_var <- function(icc,n,k) {

  return(  2 * (1-icc)^2 * ( 1 + (n-1) * icc )^2 / ( k * n * (n-1) )  )
```

```
}

# Estimate ICC variance

metadata$ICC.Var_Est <- foreach(i=1:nrow(metadata), .combine = 'rbind') %do%
calc_icc_var(metadata$ICC[i], metadata$ICCK[i], metadata$ICCNo[i])

# Meta-analysis: pooled estimate of ICC nested by dataset

# nesting reference:
https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/fitting-a-three-
level-model.html

ICC_MN <- rma.mv(ICC, ICC.Var_Est, data=metadata,

                 random = list( ~1 | ID ),

                 method="REML")


##COEFA meta-analysis (note seperate WD/RMdown file):

##check WD correct

getwd()


## load necessary packages

library(tidyverse)

library(readxl)

library(metafor)

library(mixtools)

library(gdata)

library(foreach)

library(coefa)


##datasetup:

COEFAmatrices<-coefa_read(type = "xlsx")

print(COEFAmatrices)


##co-occurence matrices EFA
```

```
GCMAT <- coefa_gcm(COEFAmatrices)

print(GCMAT)

sz <- c(34, 274, 98, 185, 105, 501, 81, 67, 82, 305, 52, 257, 400, 68, 84, 53,
180, 119, 100, 106, 100, 331, 79)

ACMAT <- coefa_acm(GCMAT, sz, samplesized=TRUE)

print(ACMAT)

COEFMASCREEN <- coefa_summary(ACMAT, fa="fa")

print(COEFMASCREEN)

COEFAPCASCREEN <- coefa_summary(ACMAT, fa="pc")

print(COEFAPCASCREEN)

COEFAMA6fac <- coefa_fa(ACMAT, nfactors=6, methodcoefa="EFA", rotate="varimax",
fm="uls")

print(COEFAMA6fac)

summary(COEFAMA6fac)

COEFAMA6fac <- coefa_fa(ACMAT, nfactors=6, methodcoefa="EFA", rotate="promax",
fm="reml")

print(COEFAMA6fac)

summary(COEFAMA6fac)


print(summary(ICC_MN))
```

Appendix 1.4

Table A1 describes the additional reports of included studies in the systematic review.

Table A1.1: Additional reports of included studies

Additional report	ID link to table 1.2 in main text
Bengtsson, J. (2022). *Negative Symptoms, Repetitive Transcranial Magnetic Stimulation and Heart Rate Variability in Schizophrenia and Depression*. (1618075) [Doctoral Dissertation, Uppsala Universitet (Sweden)]. Retrieved from https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-460774	2
Gil Berrozpe, G., Sánchez-Torres, A., Moreno-Izco, L., Peralta, V., Cuesta, M., 2020. The cains scale in schizophrenia spectrum psychosis. factorial and concurrent validity and its relationship with psychosocial functioning. European Psychiatry 63.	5
Engel, M., Lincoln, T. (2015). Validation of the German version of The Clinical Assessment Interview for Negative Symptoms (CAINS). *Eur Arch Psychiatry Clin Neurosci* **265** (Suppl 1), 1–138. https://doi.org/10.1007/s00406-015-0627-8	6
Li, S.-b., Liu, C., Zhang, J.-b., Wang, L.-l., Hu, H.-x., Chu, M.-y., Wang, Y., Lv, Q.-y., Lui, S.S., Cheung, E.F., 2023. Corrigendum to “Revisiting the latent structure of negative symptoms in schizophrenia: Evidence from two second-generation clinical assessments”[Schizophr. Res. 248 (2022) 131–139].	12
Richter, J., 2019. Decoding the" Meh"-Assessment of Negative Symptoms in Psychosis. Eberhard Karls Universität Tübingen.	15

Appendix 1.5

Table A1.2: Data provided by original authors

Study	Review ID	Data
Cuesta et al. (2021)	5	CAINS mean (SD); Proportion of participants with "single" relationship status; Duration of illness
Hosáková et al. (2017)	7	CAINS mean (SD); Age of illness onset; Chronbach's alpha estimate
Li et al. (2022)	12	Proportion of participants with "single" relationship status
Li et al., (2024)	13	Proportion of participants with "single" relationship status
Ristić et al. (2020)	16	Age of illness onset
Strauss et al. (2018)	19	CAINS mean (SD)
Thomson (2019)	20	Duration of illness
Vayisoğlu et al., (2024)	23	Age of illness onset

Appendix 1.6

Table A1.3: Additional sample criteria for included studies

Study	Review ID	Additional Inclusions	Exclusions
Ahmed et al. (2022)	1	Enrolled in studies which administered a measure of negative symptoms and other variables "Most": Receiving antipsychotic medications Clinically stable No changes in medication type or dose (prior weeks)	
Bengtsson et al. (2022)	2	Intervention study (16 participants): 40+ MAP-SR score	Severe medical conditions Epilepsy Metal or cochlea implant in head Medication changes (last month) Addiction Pregnancy
Blanchard et al. (2017)	3	CAINS baseline available	Clinical treatment trial participation at enrolment to study Unexpected to be able to complete study assessments
Chan et al. (2015)	4		Comorbid diagnosis of major depressive disorder Substance dependence (6 months) or abuse (1 month) IQ<70 Head injury Neurological disorder

Cuesta et al. (2021)	5		Substance abuse/dependence (past five years) Head injuries with loss of consciousness Seizures CNS infection Intellectual disability
Engel et al. (2014)	6		Neurological disorder Head injury with loss of consciousness Acute substance use disorder Inability to effectively agree and participate in the assessment due to severe psychiatric symptoms
Hosáková et al., (2017)	7		Organic brain disorder Serious cognitive deficit (unable to provide requested questionnaire responses)
Jang et al. (2017)	8		Head injury Electro-convulsive treatment Comorbid substance dependence or abuse (3 months) Serious medical conditions Low intelligence without the ability to understand instructions presented in the study
Jung et al. (2016)	9		Mood episode (1 month) Substance dependence (6 months) Head injury Neurological disorder Clinically severe medical condition
Kulenović et al. (2022)	10	Recruitment: Clusters from eight sites of intervention RCT	Organic brain disorder Serious cognitive deficit (unable to provide requested questionnaire responses)

Laraki et al. (2022)	11	No hospitalisations (6 months Judged to be in a stable phase of illness No expected changes in pharmacological treatments (recruitment period and month preceding)	Traumatic brain injury Unable to participate in the assessment (e.g. due to severe psychiatric symptoms).
Li et al. (2022)	12		Comorbid psychiatric diagnosis Organic disorder Neurological disorder Head injury Substance or alcohol abuse.
Li et al., (2024)	13		Comorbid psychiatric diagnosis Psychiatric diagnosis secondary to organic disease Head injury Neurological condition Alcohol or substance abuse/dependence
Rekhi et al. (2019)	14		Concurrent alcohol or substance use disorder IQ <70 History of head injury Neurological disorder
Richter et al. (2019)	15	(Corrected) normal vision Intervention study (70 participants): Stable symptoms 10+ PANSS sub-score Intervention study (25 participants): Unspecified	Substance dependence as lead clinical problem IQ<70 (approximated based on education) Treatment sample: Severe depressive symptoms Structural brain lesions Severe extrapyramidal side effects Engaged in psychotherapeutic treatment

Ristić et al. (2020)	16	No planned discharge (12 months) Minimum one hospital admission (lifetime)	Organic brain disorder Serious cognitive deficit (based on clinician judgement)
Russo et al. (2021)	17	Recruitment: Cluster randomised controlled trial of intervention 268 from UK 257 from Bosnia and Herzegovina, Kosovo* by UN resolution, Montenegro, North Macedonia, and Serbia. UK sample: PANSS (18+) No antipsychotic medication changes (6 weeks) South-Eastern Europe sample: Minimum one hospital admission (lifetime) No planned discharge (12 months)	South-Eastern Europe sample: Organic brain disorder Severe cognitive deficits (based on clinician judgement)
Blaževska Stoilkovska et al. (2021)	18	Minimum one psychiatric admission (lifetime)	Organic brain disorder Severe cognitive deficits Unable to provide informed consent or reliable information on study measures
Strauss et al. (2018)	19		Participants data already contributed to EFA study
Thomson (2019)	20		Those deemed unable to give informed consent by those involved in their care Moderate to severe learning disability Neurological condition
Uka et al. (2020)	21	In psychiatric treatment (3 months) Planned continuance (3 months)	

		Capable for giving informed consent	
Valiente-Gómez et al. (2015)	22		Brain trauma Neurological disease Alcohol or substance abuse (12 months)
Vayisoğlu et al., (2024)	23		Brain trauma Neurological disease Alcohol or substance abuse (12 months)
Xie et al. (2018)	24	Han ethnicity	Neurological disorder Head injury Substance abuse or dependence IQ<80

Appendix 2.1

The Single-Case Reporting guideline In BEhavioural interventions (SCRIBE) 2016 Checklist

Item number	Topic	Item description	Notes
TITLE and ABSTRACT			
1	Title	Identify the research as a single-case experimental design in the title	P57
2	Abstract	Summarise the research question, population, design, methods including intervention/s (independent variable/s) and target behaviour/s and any other outcome/s (dependent variable/s), results, and conclusions	P59
INTRODUCTION			
3	Scientific background	Describe the scientific background to identify issue/s under analysis, current scientific knowledge, and gaps in that knowledge base	P62
4	Aims	State the purpose/aims of the study, research question/s, and, if applicable, hypotheses	P62-63
METHODS			
	DESIGN		
5	Design	Identify the design (e.g., withdrawal/reversal, multiple-baseline, alternating-treatments, changing-criterion, some combination thereof, or adaptive design) and describe the phases and phase sequence (whether determined *a priori* or data-driven) and, if applicable, criteria for phase change	P63, 65
6	Procedural changes	Describe any procedural changes that occurred during the course of the investigation after the start of the study	P63
7	Replication	Describe any planned replication	P63
8	Randomisation	State whether randomisation was used, and if so, describe the randomisation method and the elements of the study that were randomized	P65
9	Blinding	State whether blinding/masking was used, and if so, describe who was blinded/masked	P65
	PARTICIPANT/S or UNIT/S		
10	Selection criteria	State the inclusion and exclusion criteria, if applicable, and the method of recruitment	P64-65
11	Participant characteristics	For each participant, describe the demographic characteristics and clinical (or other) features relevant to the research question, such that anonymity is ensured	P73
	CONTEXT		
12	Setting	Describe characteristics of the setting and location where the study was conducted	P64
	APPROVALS		
13	Ethics	State whether ethics approval was obtained and indicate if and how informed consent and/or assent were obtained	P70
	MEASURES and MATERIALS		
14	Measures	Operationally define all target behaviours and outcome measures, describe reliability and validity, state how they were selected, and how and when they were measured	P66-70
15	Equipment	Clearly describe any equipment and/or materials (e.g., technological aids, biofeedback, computer programs, intervention manuals or other material resources) used to measure target behaviour/s and other outcome/s or deliver the interventions	P66-70
	INTERVENTIONS		
16	Intervention	Describe intervention and control condition in each phase, including how and when they were actually administered, with as much detail as possible to facilitate attempts at replication	P65
17	Procedural fidelity	Describe how procedural fidelity was evaluated in each phase	P70-71
	ANALYSIS		
18	Analyses	Describe and justify all methods used to analyse data	P70-72
RESULTS			
19	Sequence completed	For each participant, report the sequence actually completed, including the number of trials for each session for each case. For participant/s who did not complete, state when they stopped and the reasons	P72
20	Outcomes and estimation	For each participant, report results, including raw data, for each target behaviour and other outcome/s	P73-75
21	Adverse events	State whether or not any adverse events occurred for any participant and the phase in which they occurred	P72
DISCUSSION			
22	Interpretation	Summarise findings and interpret the results in the context of current evidence	P91-92
23	Limitations	Discuss limitations, addressing sources of potential bias and imprecision	P93-94
24	Applicability	Discuss applicability and implications of the study findings	P92-93
DOCUMENTATION			
25	Protocol	If available, state where a study protocol can be accessed	P63
26	Funding	Identify source/s of funding and other support; describe the role of funders	P58

Appendix 2.2

Table A2.1: four-week baseline measurement schedule

	Time period (weeks)									
	1	2	3	4	5	6	7	8	9	10
Phase	B	B	B	B	**I_1**	I_2	I_3	I_4	I_5	I_6
Actigraphy	X	X	X	X	**X**	X	X	X	X	X
Report of time use	X				**X**		X			X
SNS	X				**X**		X			X
CAINS	X				**X**		X			X
BCIS	X				**X**		X			X
PSP	X				**X**		X			X
QPR	X				**X**		X			X
Personal recovery (PQRST)					**X**	X	X	X	X	X
MAS-A – Participant (measured by IPII)	X									
MAS-A – Participant (measured by therapy transcripts)					**X**		X			X
MAS-A - Therapist (therapy transcripts)					**X**		X			X
MAS-A T-R (Therapist)					**X**	X	X	X	X	X
MAS-A T-R (Participant)					**X**	X	X	X	X	X
WAI-SR-C							X			X
WAI-SR-T							X			X

*B = baseline, I = intervention (sessions numbered 1-6), X denotes when a measure is taken, with blank segments representing no measure being completed that week. **SNS:** Self-evaluation of Negative Symptoms; **CAINS:** Clinical Assessment Interview for Negative Symptoms; **BCIS:** Beck Cognitive Insight Scale; **QPR:** Questionnaire about the Process of Recovery; **PSP:** Personal and Social Performance scale; **PQRST:** Personal Questionnaire Rapid Scaling Technique; **MAS-A:** Metacognition Assessment Scale – Adapted; **IPII:** Indiana Psychiatric Illness Interview; **WAI-SR-C:** Working Alliance Inventory – Short Revised – Client; **WAI-SR-T:** Working Alliance Inventory – Short Revised – Therapist; **MAS-A T-R:** MAS-A – Therapist Rated. **Colour key:** - negative symptoms measures; - metacognition measures; - secondary outcomes; - therapeutic effectiveness measures

Appendix 2.3

Appendix 2.3 removed due to confidentiality issues.

Appendix 2.4:

```
##check WD correct
getwd()


##Pre-processing - GGIR data:
library("GGIR")
library("GGIRread")
##code to run
GGIR(mode = 1:5,


     ##file being analysed


     datadir = c("C:/Users/nicol/OneDrive - University of Glasgow/DClin
Coursework/MRP/Data collection/NSSCEDActigraphy/FILE"),


     ##file output location


     outputdir = c("C:/Users/nicol/OneDrive - University of Glasgow/DClin
Coursework/MRP/Data collection/NSSCEDOutput"),


     #labels


     studyname = c("FILE"),
     idloc = 2,
     ##report functions
     do.report = c(2,4,5),


     ##report parameters
     ## - part 2:
     includedaycrit = c(4,4), # actigraph wore for one hour in that day results
in it being counted for analysis
     includdaycrit.part5 = 0.333, # actigraph being wore for one hour in the
waking period results in it being counted for analysis
     frag.metrics = "all", # tells GGIR to report when participants are in
different levels of physical activity throughout day and night
     part5_agg2_60seconds = TRUE) #uses 1 minute resolution for data as in
behavioural fragmentation literature


## load necessary packages
```

```r
library(tidyverse)
library(readxl)
library(scan)
library(mice)
library(scplot)
library(scdhlm)
library(SCDA)
library(micemd)
library(miceadds)
library(broom)
library(broom.mixed)
library(mitools)
library(nlme)
library(lmeInfo)
library(msm)
library(mitml)
library(psych)
library(dplyr)

##load data and compute descriptives
##actigraphy dataset
sceddataone <- read_excel("Analysesdataset.xlsx")
sceddataone[sceddataone == ""] <- NA
View(sceddataone)
summary(sceddataone)
##NON-actigraphy dataset
sceddatatwo <- read_excel("sceddata.xlsx")
sceddatatwo[sceddatatwo == ""] <- NA
View(sceddatatwo)
summary(sceddatatwo)

#descriptives per participant (N)
VAR_descriptives_N <- filter(sceddata, ID == "N")
VAR_Mean_N <- aggregate(VAR ~ phasecode, FUN=mean, data= VAR_descriptives_N)
VAR_Median_N <- aggregate(VAR ~ phasecode, FUN=median, data=
VAR_descriptives_N)
VAR_SD_N <- aggregate(VAR ~ phasecode, FUN=sd, data= VAR_descriptives_N)


VAR_Mean_N
```

```
VAR_Median_N

VAR_SD_N


##correlation matrices


corr_subset_ACTI<- subset(sceddataone, select = c('case', 'time', 'phasecode',
'dur_day_total_IN_min', 'VAR', 'dur_day_total_MOD_min',
'dur_day_total_VIG_min', 'L5_ENMO_mg_0to24hr', 'M5_ENMO_mg_0to24hr',
'duration_sib_wakinghours', 'number_sib_wakinghours',
'duration_sib_wakinghours_atleast15min', 'SptDuration', 'VAR',
'TUB_Total_weighted', 'TUBwiIndTot', 'QPR', 'BCIS_Tot', 'PSP_Tot', 'MAPTot',
'ExpTot', 'CAINSTot', 'P_MASA_Tot_TR', 'T_MASA_Tot_TR', 'P_MASA_Tot_RR',
'T_MASA_Tot_RR', 'WAISRC_Tot_CR', 'WAISRC_Tot_TR', 'CurrActSatis'))

corr.test(as.matrix(corr_subset_ACTI), short=FALSE)


corr_subset<- subset(sceddatatwo, select = c('case', 'time', 'phasecode',
'VAR', 'TUB_Total_weighted', 'TUBwiIndTot', 'QPR', 'BCIS_Tot', 'PSP_Tot',
'MAPTot', 'ExpTot', 'CAINSTot', 'P_MASA_Tot_TR', 'T_MASA_Tot_TR',
'P_MASA_Tot_RR', 'T_MASA_Tot_RR', 'WAISRC_Tot_CR', 'WAISRC_Tot_TR',
'CurrActSatis'))

corr.test(as.matrix(corr_subset), short=FALSE)


##create MLM subset


mlmsubset_VAR <- subset(sceddata, select = c('VAR', 'CAINSTot',
'TUB_Total_weighted', 'P_MASA_Tot_RR', 'phasecode', 'case' (##and
'PhaseTimecentr' for actigraphy data)))

  view (mlmsubset_VAR)


  ##create SCDF object for VAR


  Sara_VAR<- scdf(c(A = ..., B = ...))

  Matt_VAR<- scdf(c(A = ..., B = ...))

  John_VAR<- scdf(c(A = ..., B = ...))

  MBD_VAR <- c(Sara_VAR, Matt_VAR, John_VAR)

  names(MBD_VAR) <- c("Sara", "Matt", "John")


  ##autocorrelation per participant

  acf(VAR_descriptives_X$VAR, na.action = na.pass)


  ##visualisation

  ##actigraphy data

  ##visualise participant actigraphy data
```

```
  VARplot <- scplot(Participant_VAR) |>
    add_statline("max", phase = "A") |>
    add_statline("trend") |>
    add_statline("median", phase = "A") |>
    add_legend() |>
    set_ylabel("VAR")
  print(VARplot)


  ##visualise non-actigraphy data
  VARplot <- scplot(MBD_VAR) |>
    add_statline("max", phase = "A") |>
    add_statline("trend") |>
    add_legend() |>
    set_ylabel("VAR")
  print(VARplot)


  ##statistical analysis
  #percentage non-overlapping data
  VARPND <- pnd('SELECTED_VAR') ##add decreasing = TRUE if hope for a reduction
  print(VARPND)
  pandVAR <- pand('SELECTED_VAR') ##add decreasing = TRUE if hope for a
reduction
  print(pandVAR)


  #calculate MBR_Participant
  # data input
  score_PARTICIPANT_VAR <- PARTICIPANT_VAR[[1]]$values
  n_a_Participant <- X#length of baseline
  aim <- "reduce" ##or "increase" as appropriate
  #Data Manipulations
  n_b_Participant <- length(score_PARTICIPANT_VAR)-n_a_Participant
  phaseA_Participant <- score_PARTICIPANT_VAR[1:n_a_Participant]
  phaseB_Participant <- score_PARTICIPANT_VAR[(n_a_Participant
+1):length(score_PARTICIPANT_VAR)]
  mblr <- ((mean(phaseB_Participant)-
mean(phaseA_Participant))/mean(phaseA_Participant))*100
  # Graphical representation
  # Plot data
  indep_Participant <- 1:length(score_PARTICIPANT_VAR)
```

```
  plot(indep_Participant,score_PARTICIPANT_VAR,
xlim=c(indep_Participant[1],indep_Participant[length(indep_Participant)]),
xlab="Measurement time", ylab="Score", font.lab=2)


lines(indep_Participant[1:n_a_Participant],score_PARTICIPANT_VAR[1:n_a_Particip
ant],lty=2)


lines(indep_Participant[(n_a_Participant+1):length(score_PARTICIPANT_VAR)],scor
e_PARTICIPANT_VAR[(n_a_Participant+1):length(score_PARTICIPANT_VAR)],lty=2)

  abline (v=(n_a_Participant+0.5))

  points(indep_Participant, score_PARTICIPANT_VAR, pch=24, bg="black")

  title(main="MBLR")

  # Split-middle trend & projections with limits


lines(indep_Participant[1:n_a_Participant],rep(mean(phaseA_Participant),n_a_Par
ticipant),col="blue")


lines(indep_Participant[(n_a_Participant+1):length(score_PARTICIPANT_VAR)],rep(
mean(phaseB_Participant),n_b_Participant),col="blue")

  # Print the results

  if (aim == "reduce") {print("Mean baseline reduction"); print(mblr)} else
{print("Mean baseline increase"); print(mblr)}


  ##calculate Cohen's D and Glass's _

  VARmeandiffstats <- smd(SELECTED_VAR)

  print(VARmeandiffstats)


  ##randomisation test

  randtestVAR<- rand_test(SELECTED_VAR, limit = c(1,3), complete = TRUE,
statistic = "Mean A-B") ##Alternative reverse if expect an increase in scores

  print(randtestVAR)


  ##Calculate HPSD statistic

  effectsize_VAR <- effect_size_MB(sceddata$VAR, sceddata$phasecode,
sceddata$case, sceddata$time)

  print(effectsize_VAR)

  summary(effectsize_VAR)


  ##Calculate PHSD statistic: model 1 and 2 from report chosen due to higher
sample size required for Model 4

  PHSDVAR.1 <- lme(fixed = VAR ~ phasecode,

                   random = ~ 1 | case,

                   correlation = corAR1(0, ~ time | case),

                   data = sceddata, na.action = na.omit)
```

```
  summary(PHSDVAR.1)

  PHSDVAR.1_g <- g_mlm(PHSDVAR.1, p_const = c(0,1), r_const = c(1,0,1),
returnModel=FALSE)

  print(PHSDVAR.1_g)

  summary(PHSDVAR.1_g)

  PHSDVAR.1_g$p_beta

  PHSDVAR.1_g$phi


  ##Multiple imputation for each outcome

  #set imputation method

  impmethod_VAR <- character(ncol(mlmsubset_VAR))

  names(impmethod_VAR) <- colnames (mlmsubset_VAR)

  impmethod_VAR["VAR"] <- "2l.pmm"

  impmethod_VAR

  pm_VAR <- make.predictorMatrix(mlmsubset_VAR)

  pm_VAR[1,1] <- 0

  pm_VAR[, "PhaseTimecentr"] <- 0

  pm_VAR[, "TUB_Total_weighted"] <- 0

  pm_VAR[1,"case"] <- -2

  pm_VAR

  mlmsubset_VAR[mlmsubset_VAR == "NA"] <- ""

  res.mice.md_VAR <- mice(mlmsubset_VAR, m = 20, predictorMatrix = pm_VAR,
method = impmethod_VAR, maxit=50, printFlag = FALSE)

  print(res.mice.md_VAR)

  impdat_VAR <- complete(res.mice.md_VAR, action="long", include=FALSE)

  compute_median_VAR <- tapply(impdat_VAR$VAR, impdat_VAR$.imp, median)

  compute_mean_VAR <- tapply(impdat_VAR$VAR, impdat_VAR$.imp, mean)

  var_VAR <- tapply(impdat_VAR$VAR, impdat_VAR$.imp, var) /
nrow(impdat_VAR$data)

  pool_median_VAR <- pool.scalar(compute_median_VAR, var_VAR,
nrow(impdat_VAR$data))

  pooled_median_VAR_result <- data.frame(pooled.median = pool_median_VAR$qbar,
median.var = pool_median_VAR$ubar, btw.var = pool_median_VAR$b, pooled.var =
pool_median_VAR$t)

  print(pooled_median_VAR_result)

  pool_mean_VAR <- pool.scalar(compute_mean_VAR, var_VAR,
nrow(impdat_VAR$data))

  pooled_mean_VAR_result <- data.frame(pooled.mean = pool_mean_VAR$qbar,
mean.var = pool_mean_VAR$ubar, btw.var = pool_mean_VAR$b, pooled.var =
pool_mean_VAR$t)

  print(pooled_mean_VAR_result)
```

```
  res.mi.cor_VAR <- micombine.cor(mi.res = impdat_VAR, variables = c("VAR",
"CAINSTot", "TUB_Total_weighted"))

  res.mi.cor_VAR

  fit_m1 <- with(data= res.mice.md_VAR, expr=lme(fixed = VAR ~ 1 + phasecode +
P_MASA_Tot_RR, random = ~ 1 | case, correlation = corAR1(form=~1| case), method
= "REML", na.action = na.pass))

  summary_m1 <- summary(fit_m1)

  summary_m1

  summary(pool(fit_m1), conf.int = T)

  randefcomp_m1 <- sapply(fit_m1$analyses, FUN = function(x) VarCorr(x))

  randeftab_m1 <- randefcomp_m1[3,]

  randeftab_m1 <- as.numeric(randeftab_m1)

  randefvar_m1 <- sapply(fit_m1$analyses, FUN = function(x) getVarCov(x))

  pooledRE_m1 <- pool.scalar(Q= randeftab_m1, U = randefvar_m1, n = 20, k = 2)

  pooledRE_m1

  fit_m0 <- with(res.mice.md_VAR, lme(fixed = VAR ~ 1 + phasecode, random = ~ 1
| case, correlation = corAR1(form=~1| case), method = "REML", na.action =
na.pass))

  summary_m0 <- summary(fit_m0)

  summary_m0

  summary(pool(fit_m0), conf.int = T)

  randefcomp_m0 <- sapply(fit_m0$analyses, FUN = function(x) VarCorr(x))

  randeftab_m0 <- randefcomp_m0[3,]

  randeftab_m0 <- as.numeric(randeftab_m0)

  randefvar_m0 <- sapply(fit_m0$analyses, FUN = function(x) getVarCov(x))

  pooledRE_m0 <- pool.scalar(Q= randeftab_m0, U = randefvar_m0, n = 20, k = 2)

  pooledRE_m0

  fit_mnull <- with(res.mice.md_VAR, lme(fixed = VAR ~ 1, random = ~ 1 | case,
correlation = corAR1(form=~1| case), method = "REML", na.action = na.pass))

  summary_mnull <- summary(fit_mnull)

  summary_mnull

  summary(pool(fit_mnull), conf.int = T)

  randefcomp_mnull <- sapply(fit_mnull$analyses, FUN = function(x) VarCorr(x))

  randeftab_mnull <- randefcomp_mnull[3,]

  randeftab_mnull <- as.numeric(randeftab_mnull)

  randefvar_mnull <- sapply(fit_mnull$analyses, FUN = function(x) getVarCov(x))

  pooledRE_mnull <- pool.scalar(Q= randeftab_mnull, U = randefvar_mnull, n =
20, k = 2)

  pooledRE_mnull

  anova(fit_m1, fit_m0)

  anova(fit_m1, fit_mnull)
```

Appendix 2.5:

Table A2.2: Correlation coefficients

	Case	Time	Treatment (Y/N)	Inactivity/day (mins)	Light Activity/day (mins)	Moderate Activity/day (mins)	Lowest acceleration value/day	Highest acceleration value/day	SIB duration (waking hours)
Inactivity/day (mins)	-0.03	**0.6**	0.38						
Light Activity/day (mins)	-0.41	-0.2	-0.26	0.16					
Moderate Activity/day (mins)	-0.44	-0.26	-0.21	-0.16	0.47				
Lowest acceleration value/day	**-0.58**	**-0.59**	-0.21	-0.37	0.48	0.47			
Highest acceleration value/day	**-0.57**	-0.17	-0.26	-0.02	**0.66**	**0.93**	0.49		
SIB duration (waking hours)	0.46	0.28	-0.01	**0.55**	-0.25	**-0.53**	-0.36	-0.5	
Number of SIBs (waking hours)	-0.49	0.31	0.41	0.34	0.09	-0.35	0.48	-0.25	0.31
SIB duration (at least 15 mins, waking hrs)	**0.57**	0.57	-0.04	**0.53**	-0.3	**-0.52**	-0.49	-0.51	**0.99**
Estimated sleep duration	0.36	-0.26	**-0.52**	-0.21	-0.01	-0.01	-0.17	-0.04	0.13
SNS	-0.49	-0.38	-0.36	0.12	**-0.54**	**-0.55**	**-0.66**	**-0.67**	0.49
TUB Activity levels (weighted)	**0.73**	-0.45	-0.39	-0.37	0.5	0.47	**0.56**	**0.61**	-0.25
Activity with individuals (TUB)	0.19	**-0.86**	**-0.91**	-0.3	0.01	-0.05	-0.17	-0.13	0.21
QPR	-0.12	0.32	0.41	0.09	0.35	0.38	0.53	0.52	-0.43
BCIS	-0.28	0.45	-0.08	0.13	0.29	0.34	0.49	0.47	-0.4
PSP	0.52	-0.08	-0.07	-0.03	-0.41	-0.44	**-0.58**	**-0.57**	0.46
MAP	**0.64**	-0.51	-0.37	-0.03	-0.41	-0.44	**-0.58**	**-0.57**	0.46
EXP	0.2	-0.32	-0.45						
CAINS	0.54	-0.51	-0.49	-0.03	-0.41	-0.44	**-0.58**	**-0.57**	0.46
Participant MAS-A (TR)	**-0.95**	0.12		1	1	-1			1
Therapist MAS-A (TR)	0.15	0.12		1	1	-1			1
Participant MAS-A (RR)	-0.37	**0.76**	**0.76**	0.39	-0.3	-0.25	-0.27	-0.33	0.03
Therapist MAS-A (RR)	-0.78	-0.05							
Therapeutic alliance (PR)	-0.98								
Therapeutic alliance (TR)	-0.87								
Current activity level satisfaction	0.18	0.13	0						

Table A2.2 Contd.

	Number of SIBs (waking hours)	SIB duration (at least 15 mins, waking hrs)	Estimated sleep duration	SNS	TUB Activity levels (weighted)	Activity with individuals (TUB)	QPR	BCIS	PSP
SIB duration (at least 15 mins, waking hrs)	0.17								
Estimated sleep duration	-0.29	0.19							
SNS	-0.34	**0.59**	0.16						
TUB Activity levels (weighted)	-0.16	-0.27	0.35	-0.44					
Activity with individuals (TUB)	**-0.53**	0.29	**0.55**	0.3	0.56				
QPR	0.52	**-0.54**	-0.42	**-0.73**	0	-0.39			
BCIS	**0.53**	-0.51	-0.43	0.07	-0.36	0.42	0.22		
PSP	-0.49	**0.57**	0.36	-0.19	0.48	0.31	-0.02	-0.2	
MAP	-0.49	**0.57**	0.36	0.05	**0.63**	**0.63**	-0.31	-0.34	0.54
EXP				-0.33	0.57	0.56	0.29	-0.33	0.37
CAINS	-0.49	**0.57**	0.36	-0.14	**0.73**	**0.72**	-0.06	-0.34	0.56
Participant MAS-A (TR)	1	1	1	0.5	**-0.77**	-0.57	-0.04	0.29	-0.59
Therapist MAS-A (TR)	1	1	1	-0.51	-0.15	-0.15	0.49	-0.56	0.22
Participant MAS-A (RR)	0.38	0	-0.51	-0.17	-0.56	**-0.84**	0.38	0.45	-0.21
Therapist MAS-A (RR)				**0.93**	-0.73	-0.53	-0.53	**0.83**	-0.69
Therapeutic alliance (PR)				0.87	-0.94	-0.65	0.38	**1**	-0.12
Therapeutic alliance (TR)				0.98	**-1**	-0.87	0.06	0.92	-0.44
Current activity level satisfaction				0.23	-0.08	0.1	-0.15	0.47	0.37

Table A2.2 Contd.

	MAP	EXP	CAINS	Participant MAS-A (TR)	Therapist MAS-A (TR)	Participant MAS-A (RR)	Therapist MAS-A (RR)	Therapeutic alliance (PR)	Therapeutic alliance (TR)
EXP	0.36								
CAINS	**0.87**	**0.77**							
Participant MAS-A (TR)	**-0.86**	-0.21	0.7						
Therapist MAS-A (TR)	-0.24	-0.3	-0.27	-0.23					
Participant MAS-A (RR)	**-0.78**	-0.42	**-0.75**	**0.88**	-0.02				
Therapist MAS-A (RR)	-0.57	-0.48	-0.6	0.91	-0.94	0.51			
Therapeutic alliance (PR)	**-1**	-0.71	-0.91	1		0.76	0.74		
Therapeutic alliance (TR)	-0.92	-0.9	-1			0.5	0.92	0.94	
Current activity level satisfaction	0.43	-0.11	0.23	-0.61	-0.5	-0.2	0.37	0.94	**1**

Table A2.3: Probability values for sig. results in correlation matrix

	Case	Time	Treatment (Y/N)	Inactivity/day (mins)	Light Activity/day (mins)	Moderate Activity/day (mins)	Lowest acceleration value/day	Highest acceleration value/day	SIB duration (waking hours)
Inactivity/day (mins)		**0.02**							
Lowest acceleration value/day	**0.04**	**0.03**							
Highest acceleration value/day	**0.04**				**0.01**	**>0.01**			
SIB duration (waking hours)				**0.04**		**0.05**			
SIB duration (at least 15 mins, waking hrs)	**0.03**			**0.04**		**0.05**			**>0.01**
Estimated sleep duration			**0.05**						
SNS					**0.04**	**0.03**	**0.01**	**0.01**	
TUB Activity levels (weighted)	**>0.01**						**0.05**	**0.03**	
Activity with individuals (TUB)		**>0.01**	**>0.01**						
PSP							**0.04**	**0.04**	
MAP	**0.03**						**0.04**	**0.04**	
CAINS							**0.04**	**0.04**	
Participant MAS-A (TR)	**>0.01**								
Participant MAS-A (RR)		**>0.01**	**>0.01**						

Table A2.3 Contd.

	Number of SIBs (waking hours)	SIB duration (at least 15 mins, waking hrs)	Estimated sleep duration	SNS	TUB Activity levels (weighted)	Activity with individuals (TUB)	BCIS	MAP	EXP	CAINS	Participant MAS-A (TR)	Therapeutic alliance (TR)
SNS		**0.03**										
Activity with individuals (TUB)	**0.05**		**0.04**									
QPR		**0.05**		**>0.01**								
BCIS	**0.05**											
PSP		**0.03**										
MAP		**0.03**			**0.03**	**0.03**						
CAINS		**0.03**			**>0.01**	**>0.01**		**>0.01**	**>0.01**			
Participant MAS-A (TR)					**0.04**			**0.01**				
Participant MAS-A (RR)						**>0.01**		**>0.01**		**>0.01**	**>0.01**	
Therapist MAS-A (RR)				**>0.01**			**0.04**					
Therapeutic alliance (PR)							**0.04**	**0.05**				
Therapeutic alliance (TR)					**>0.01**							
Current activity level satisfaction												**>0.01**

Appendix 2.6

Fig. A1 Motivation and Pleasure Deficits

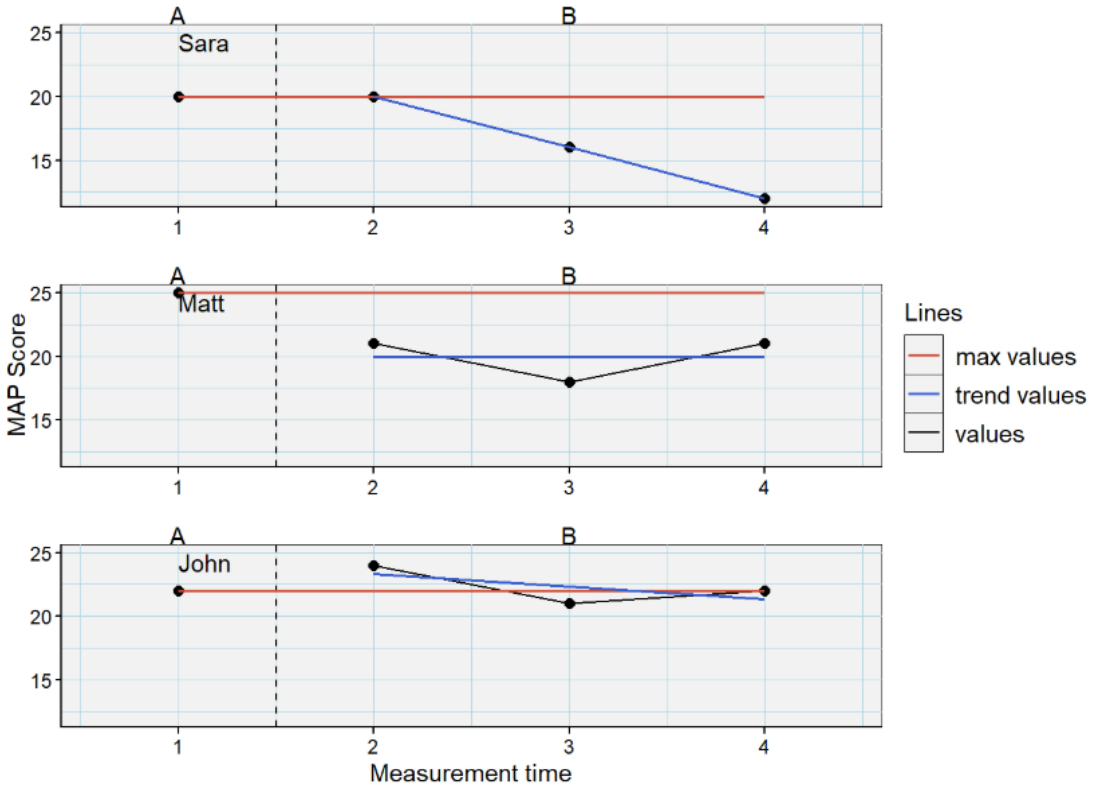


Fig. A2 Experiential deficits

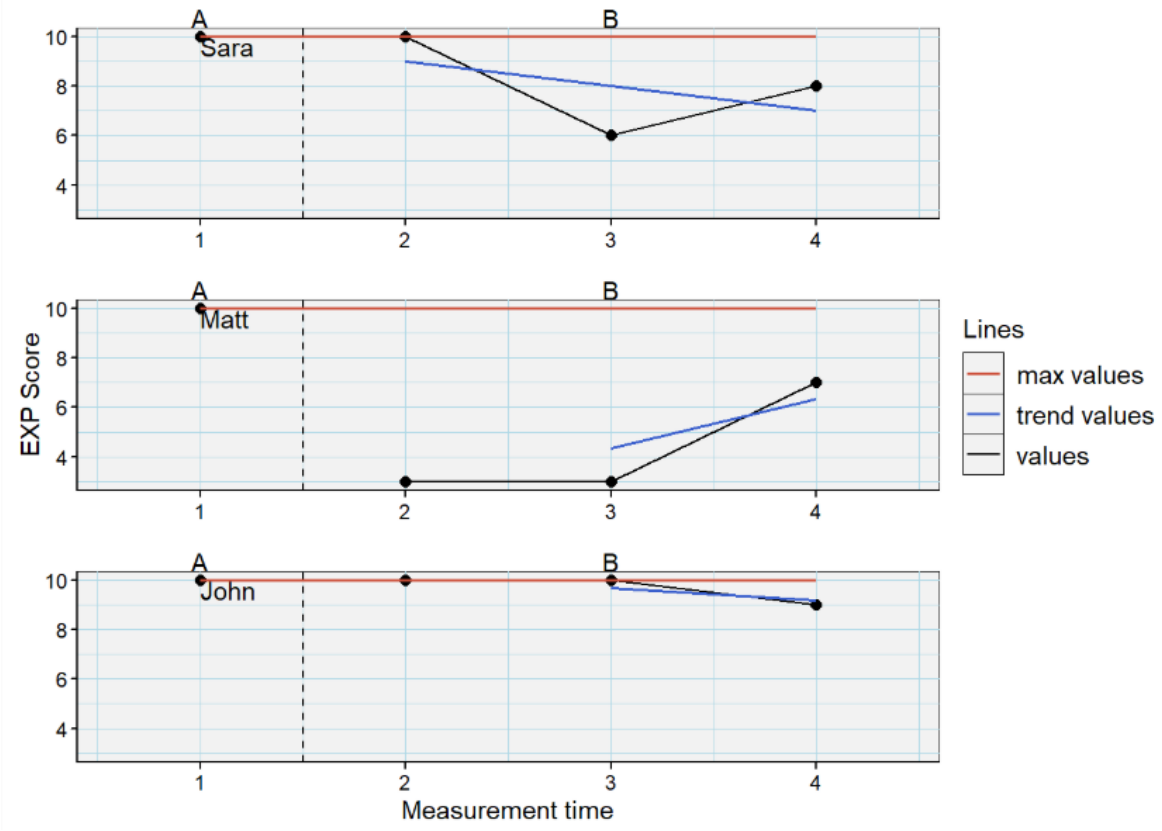


Fig. A3 Self-reported Negative Symptoms

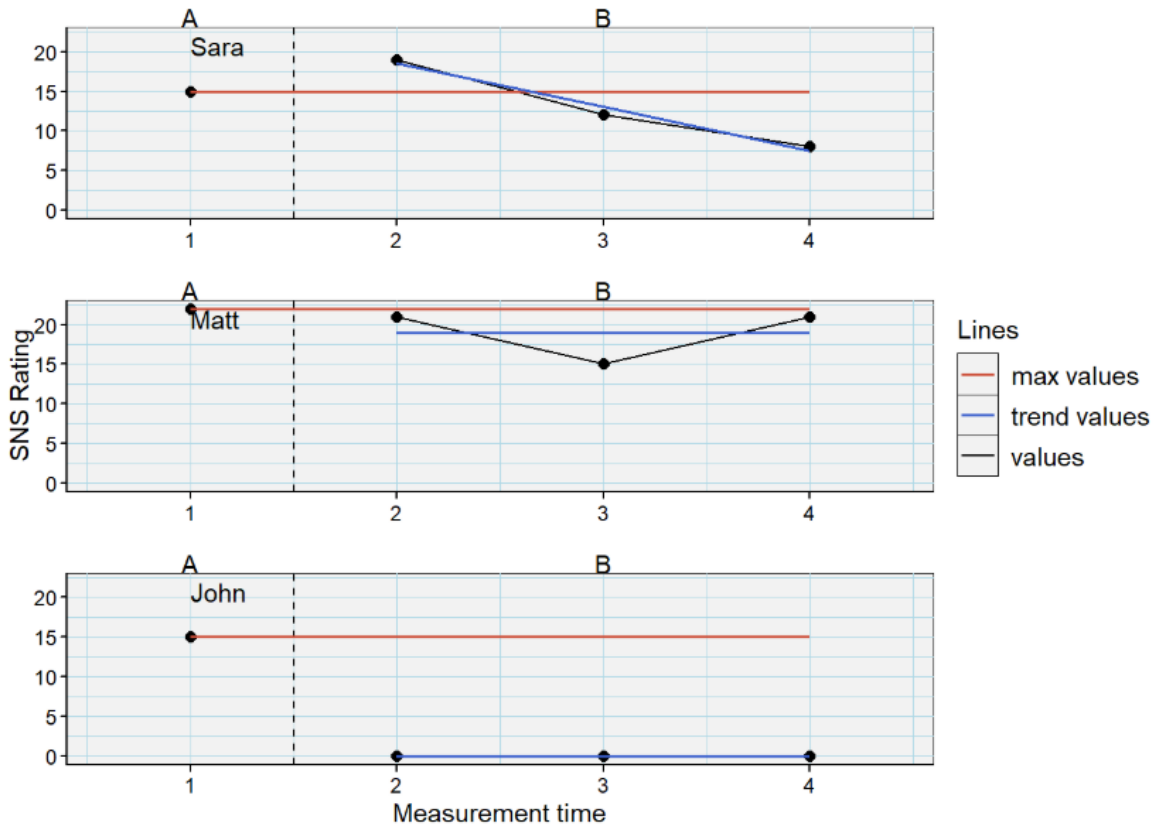


Fig. A4 Beck Cognitive Insight Scale

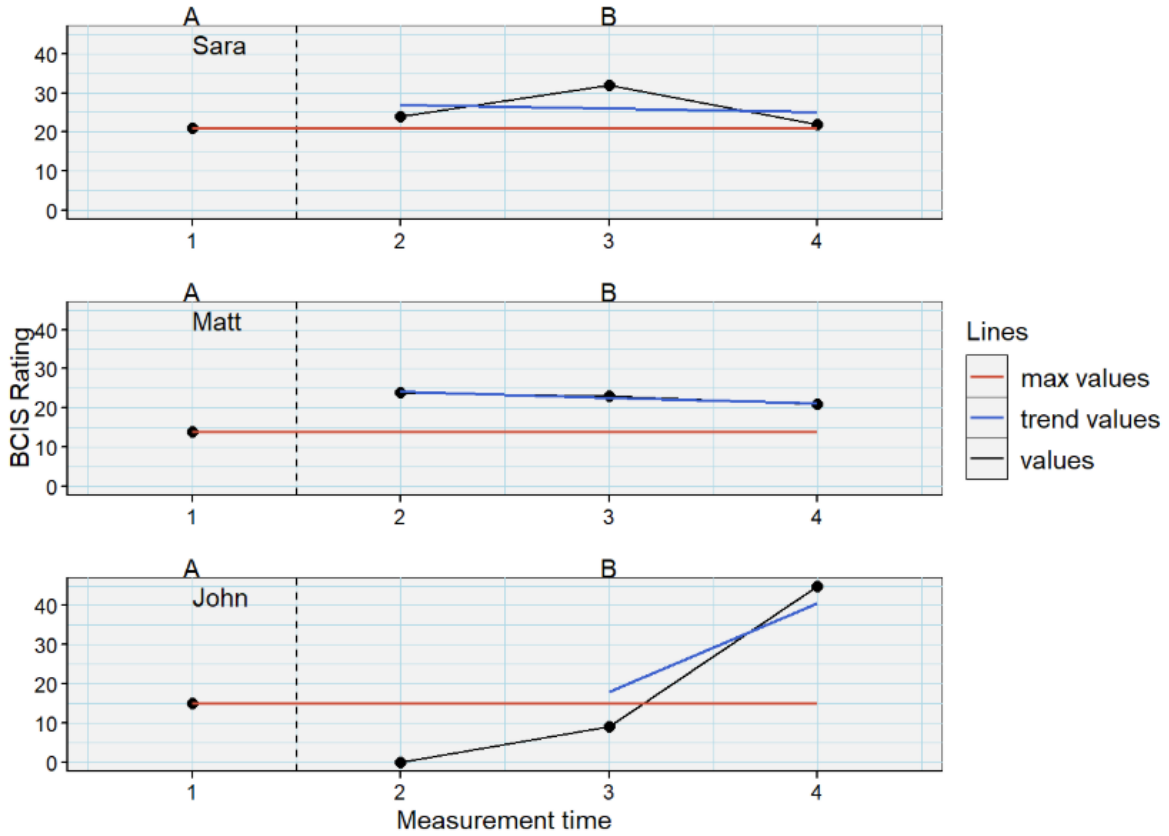


Fig. A5 Personal and Social Performance Scale

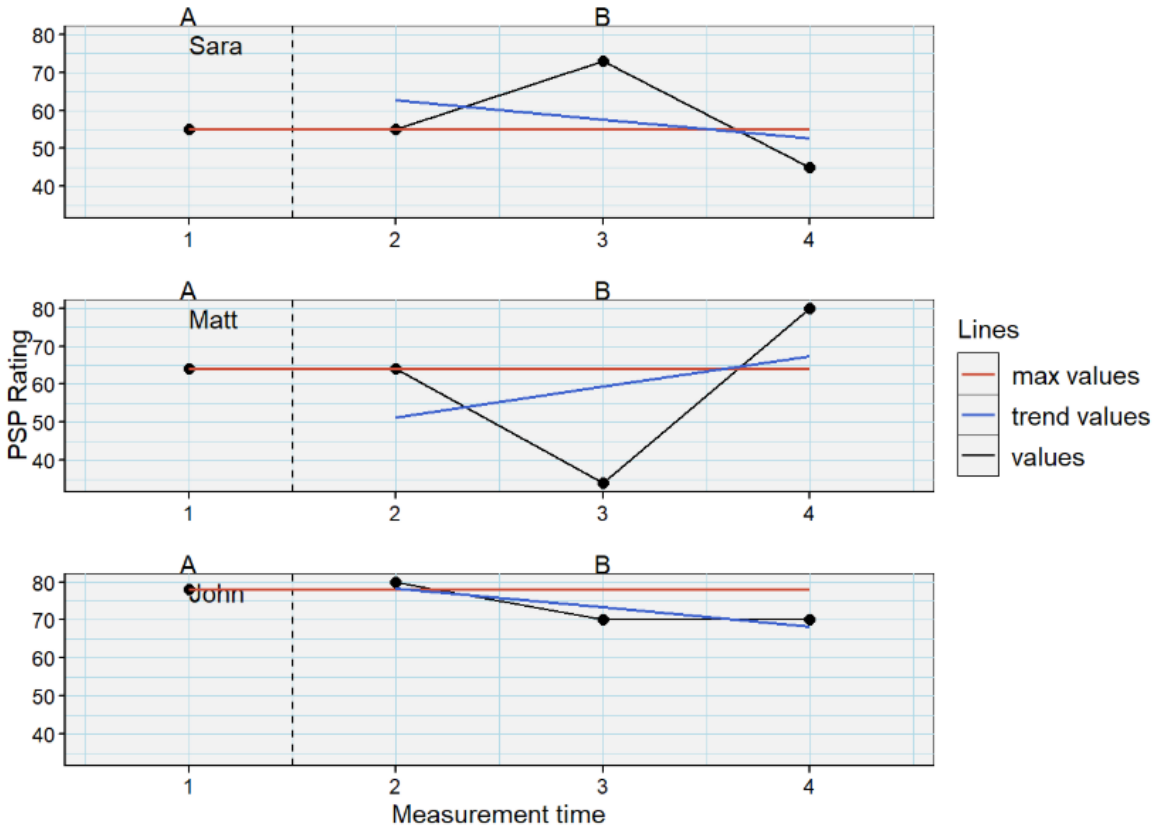


Fig. A6 Questionnaire about the Process of Recovery

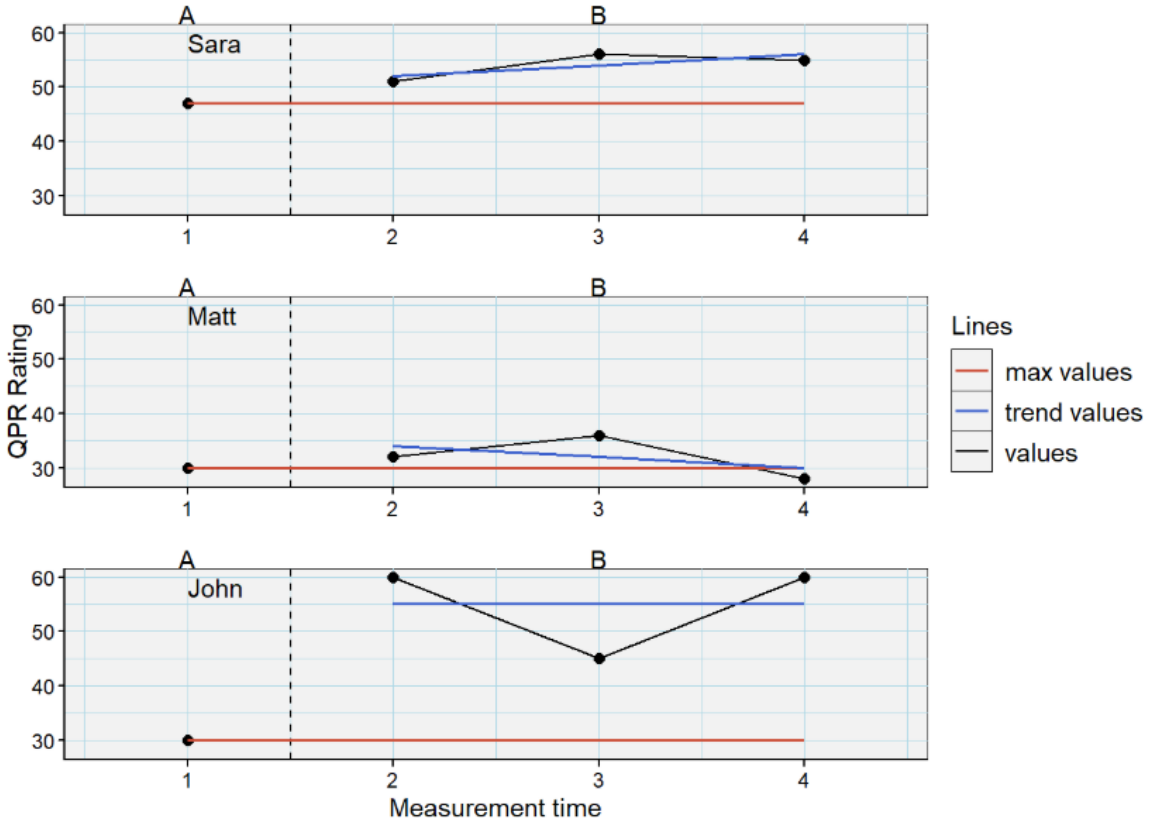


Fig. A7 Time Use Budget Level 0 activity

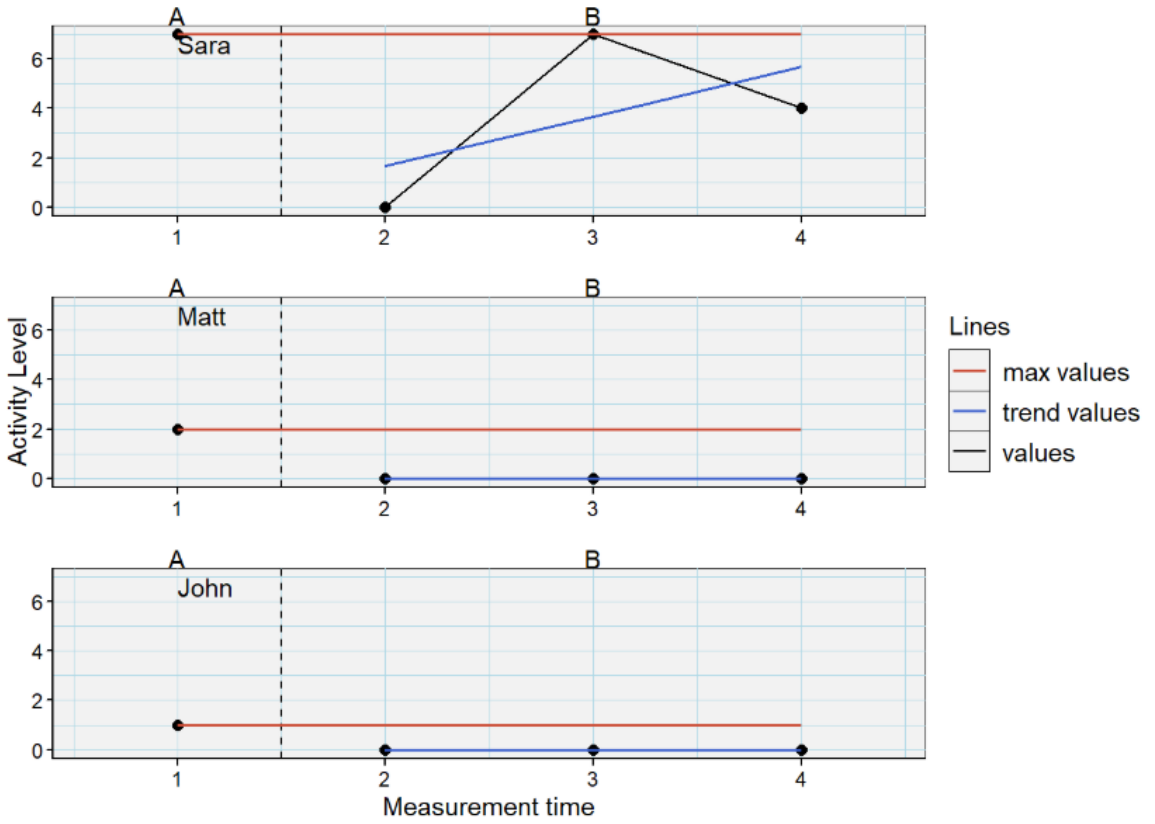


Fig. A8 Time Use Budget level one activity

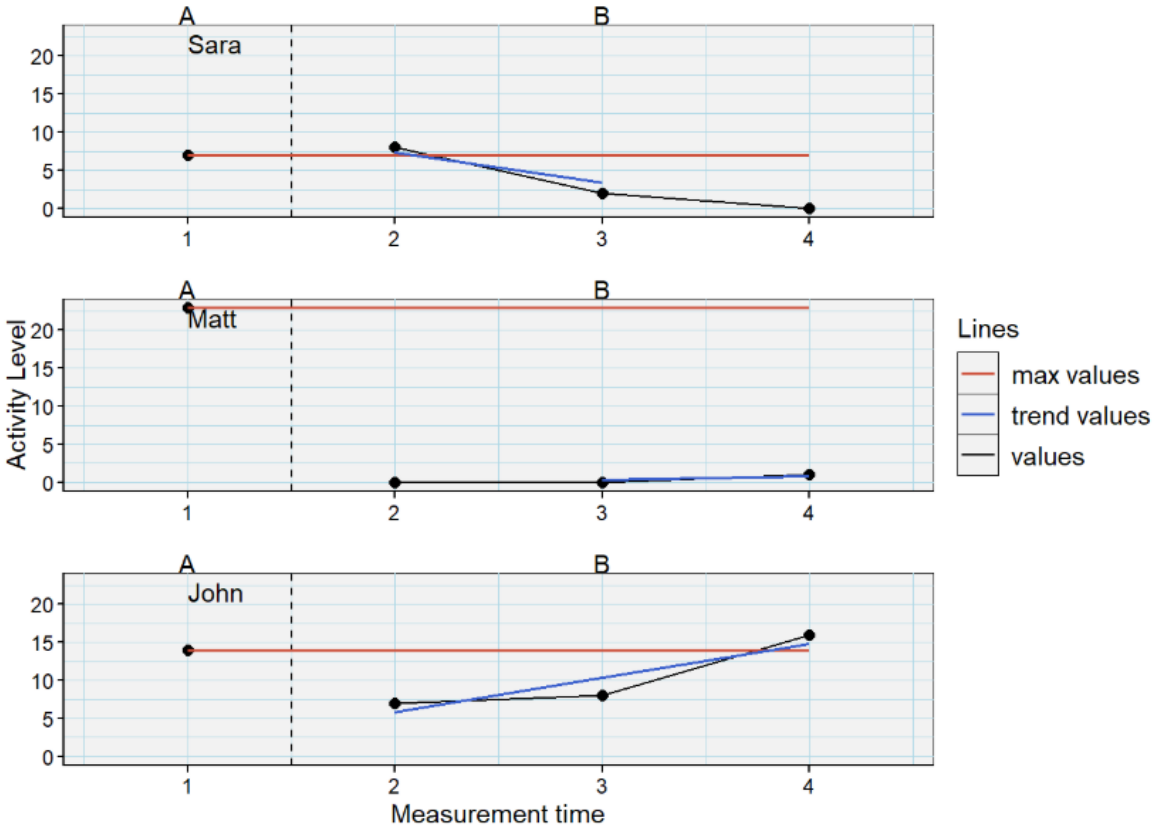


Fig. A9 Time Use Budget Level two activity

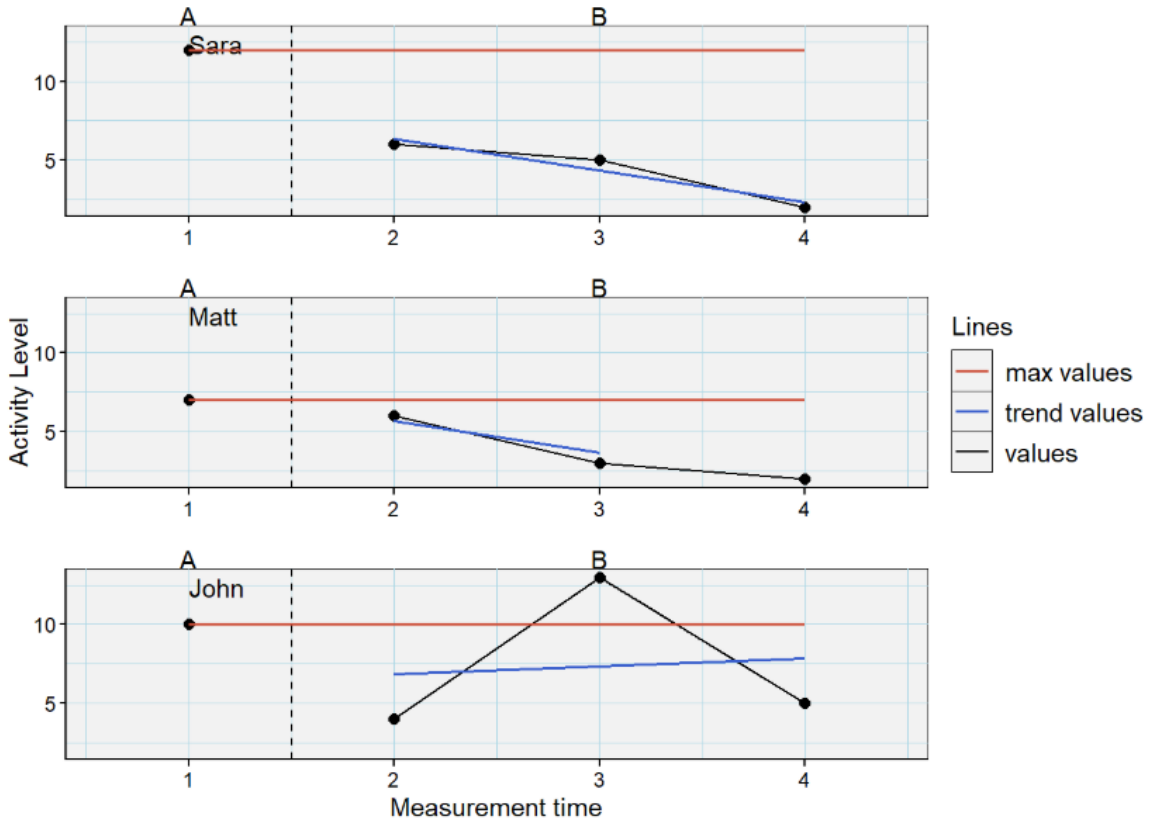


Fig. A10 Time Use Budget Level three activity

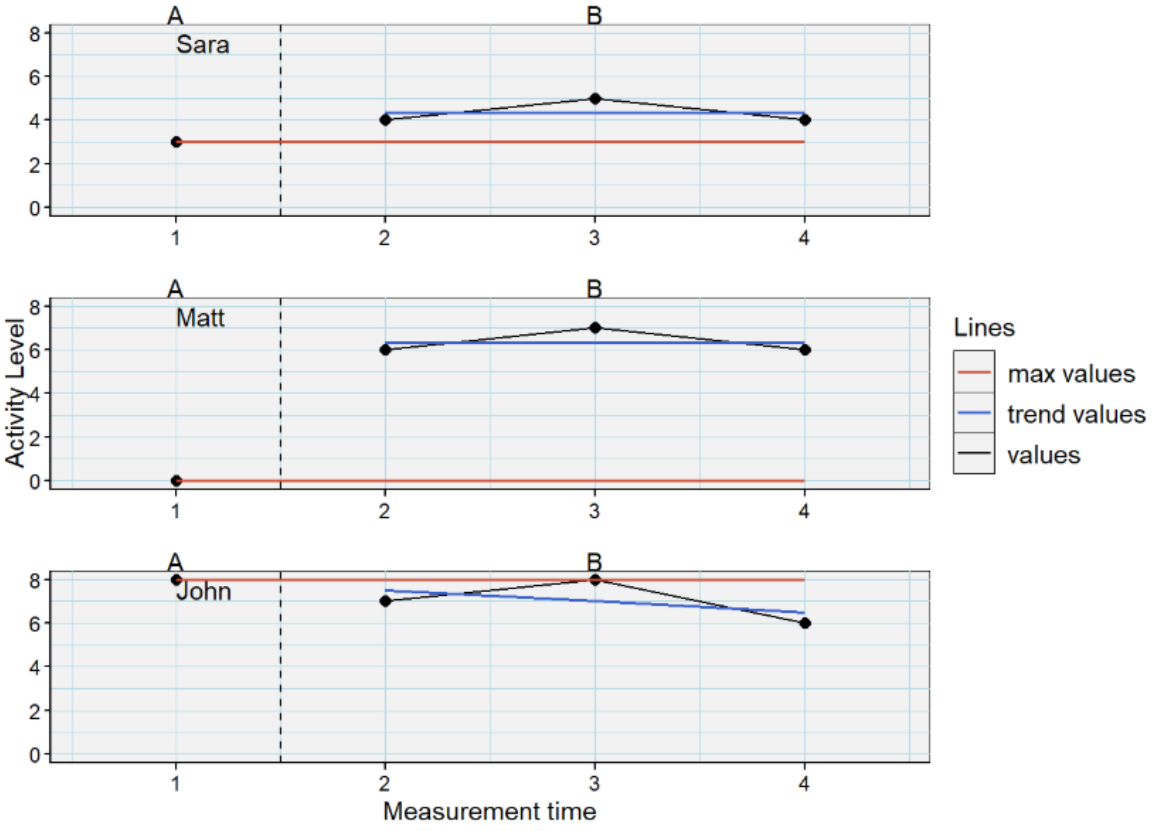


Fig. A11 Activity with individuals (Time Use Budget)

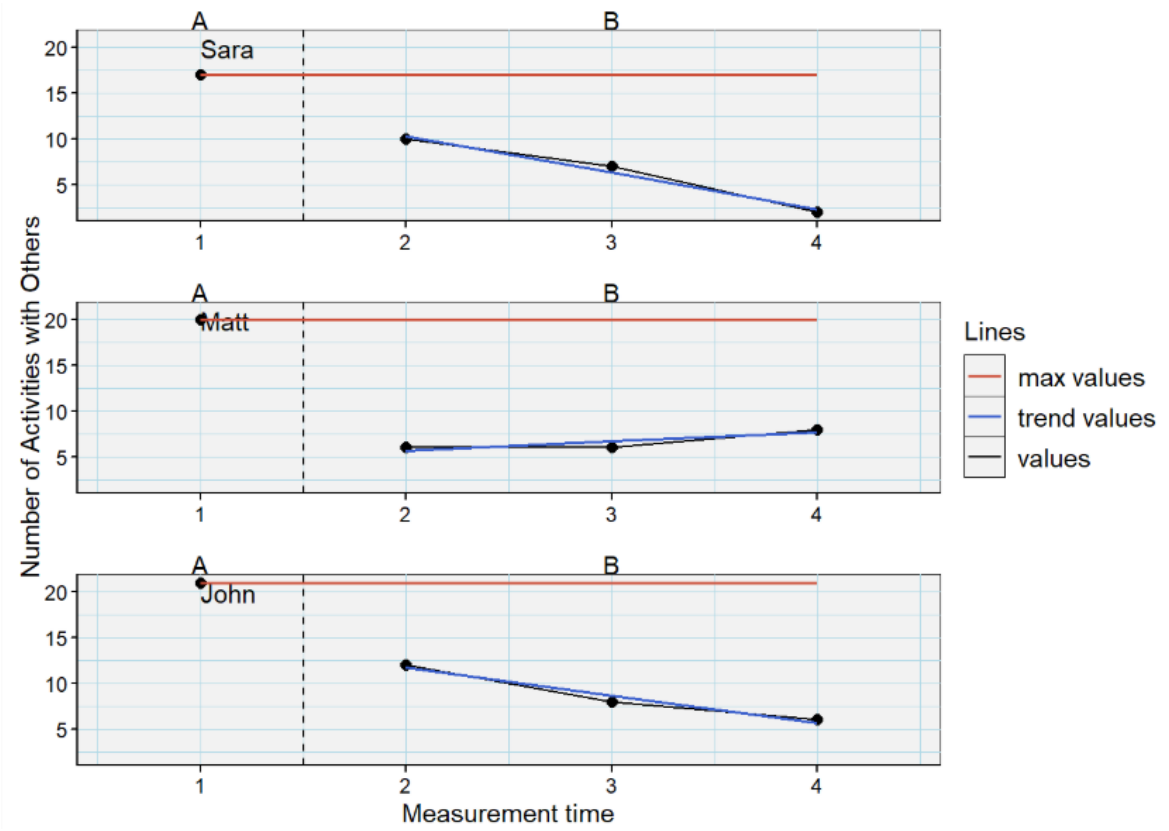


Fig. A12 Satisfaction with current activity levels

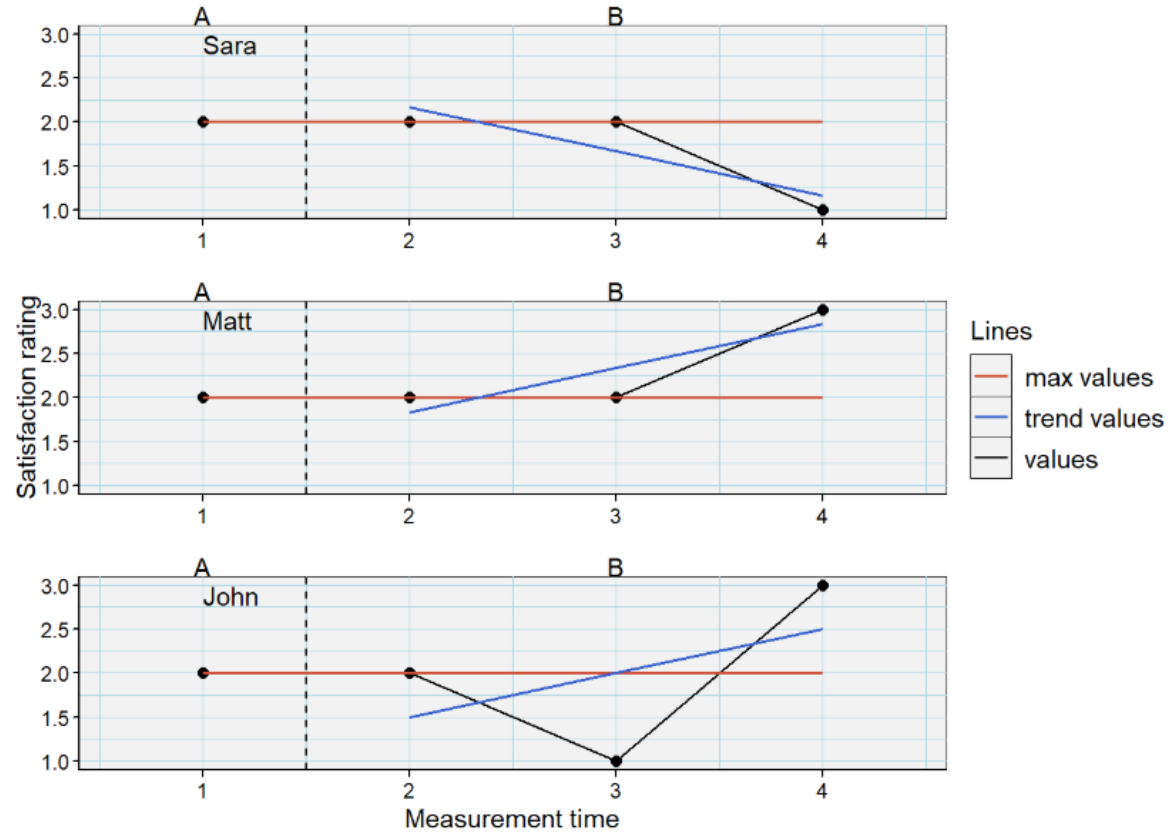


Fig. A13 Inactivity per day (minutes)

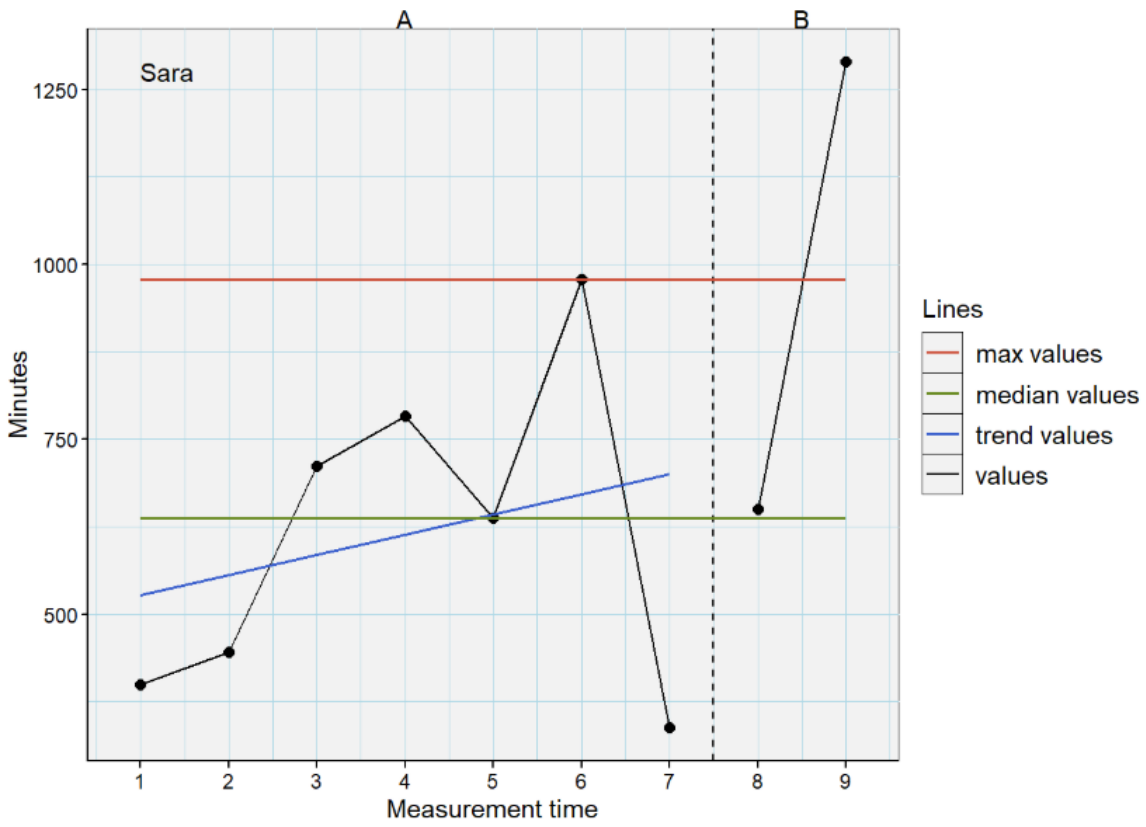


Fig. A14 Moderate activity per day (minutes)

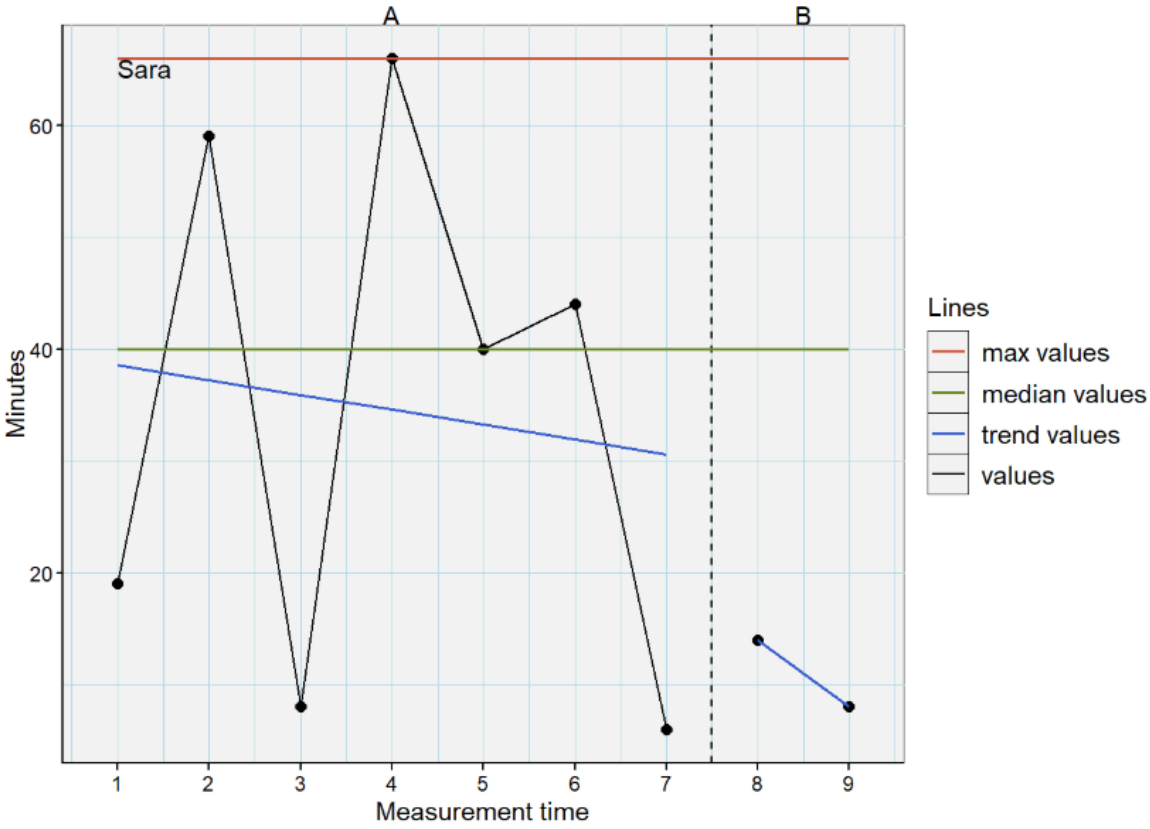


Fig. A15 Lowest acceleration value per day

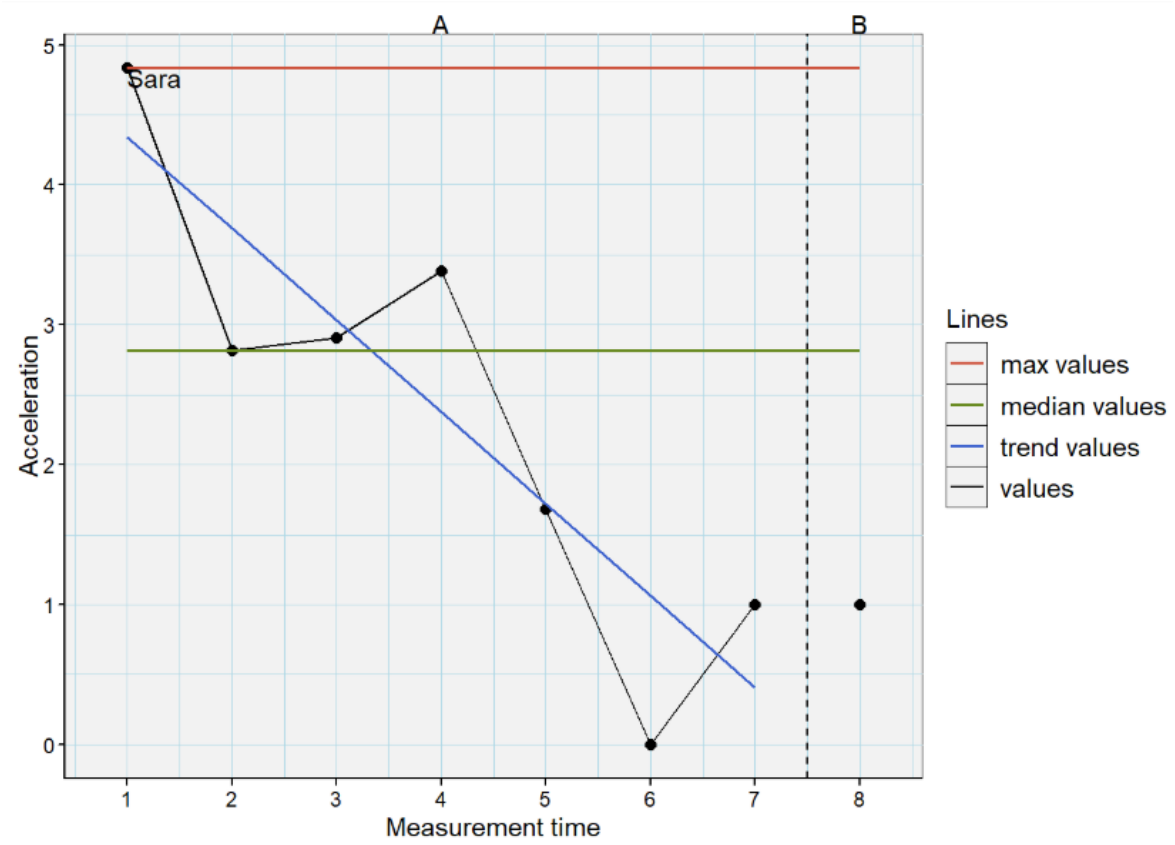


Fig. A16 Highest acceleration value per day

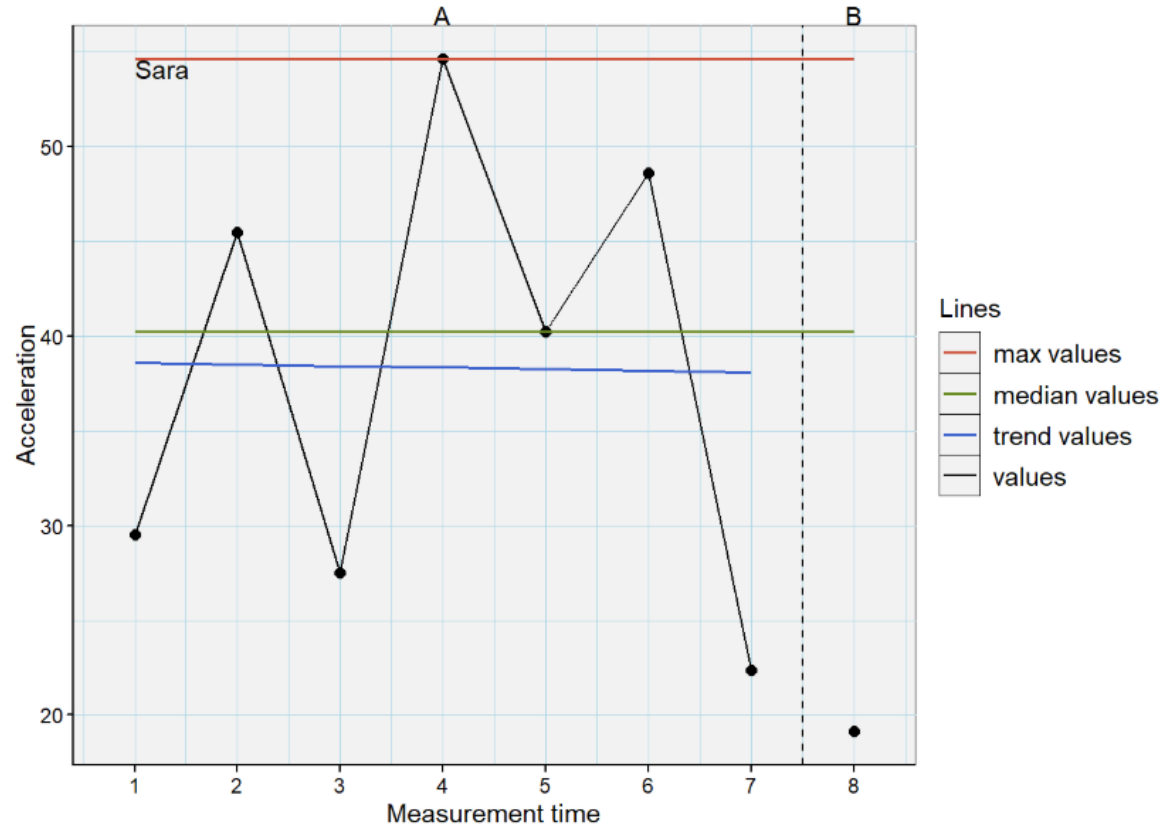


Fig. A17 Number of Sedentary Inactivity Bouts (waking hours)

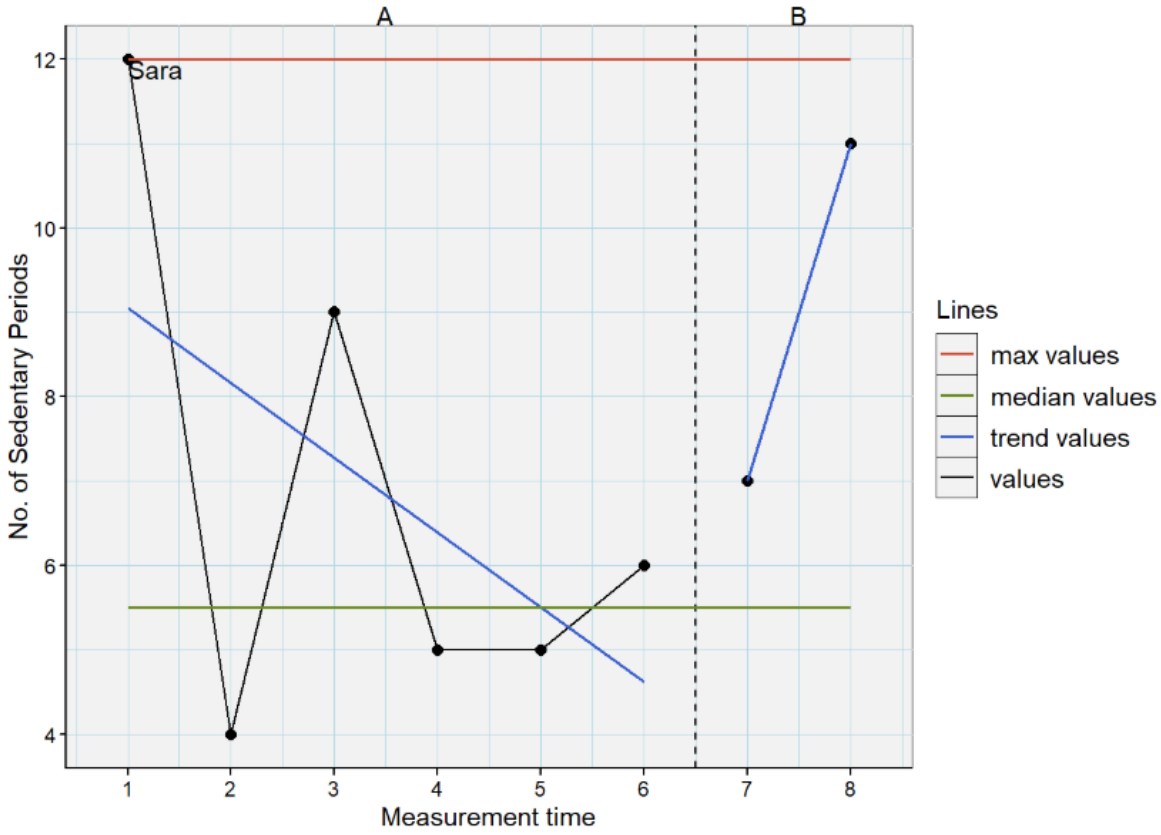


Fig. A18 Sedentary Inactivity Bout Duration (waking hours)

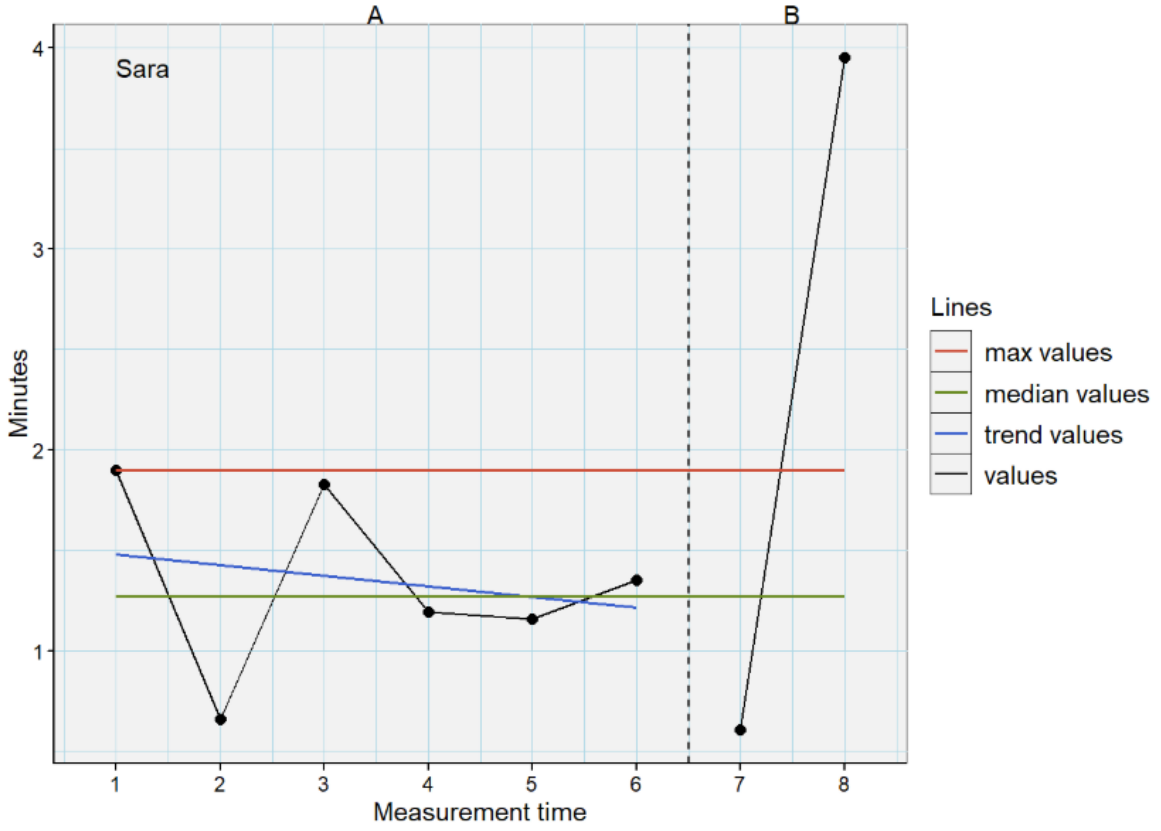


Fig. A19 Duration of Sedentary Inactivity Bouts per day (with periods of at least 15 minutes only)

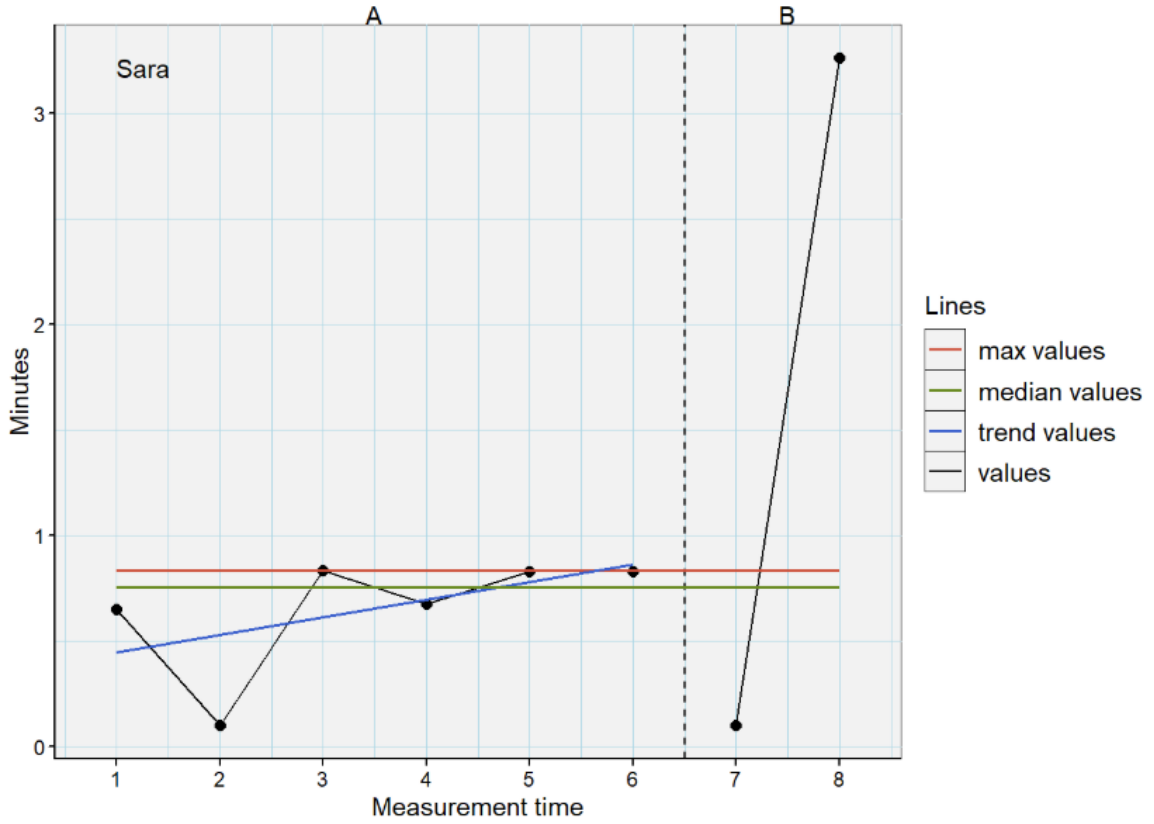
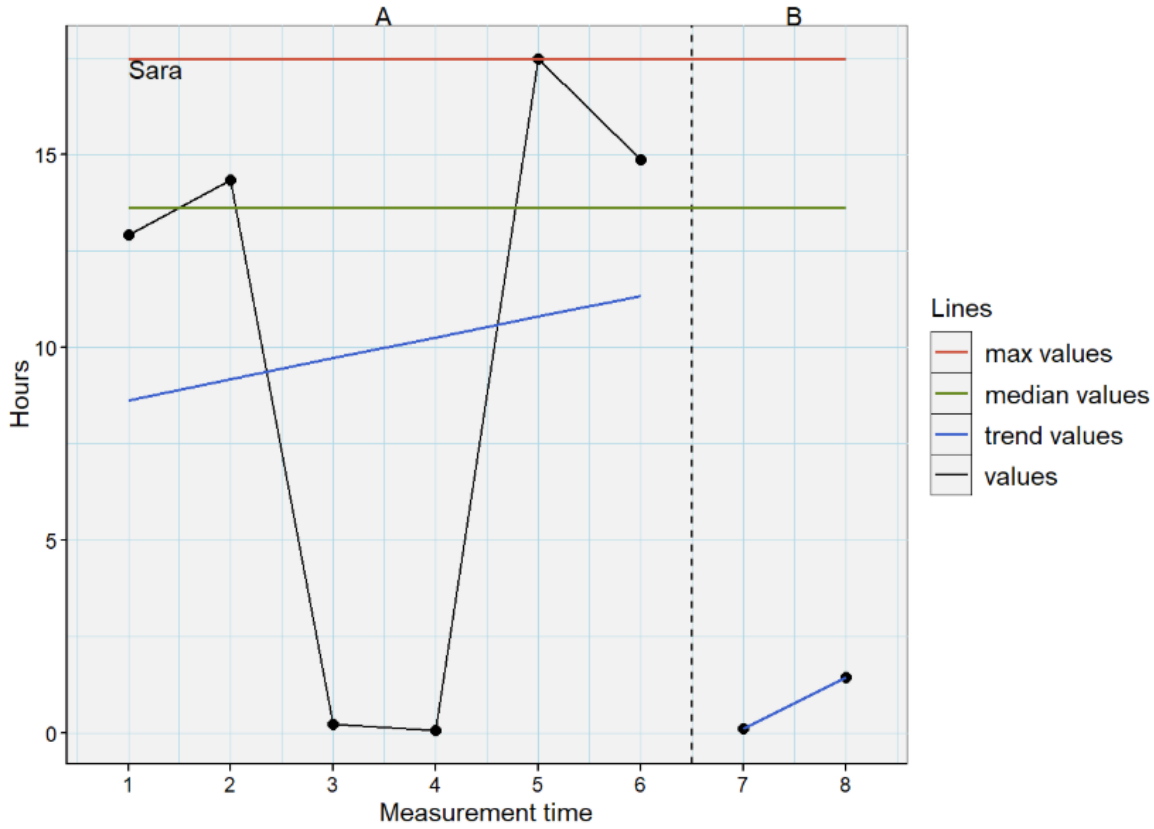


Fig. A20 Estimated sleep duration



Appendix 2.7

Table A2.4: Personal Questionnaire Rapid Scaling Technique data

Session number	John	Sara			Matt		
	Depression	Motivation for activity	Sleep	Eating	Fatigue	Sleepiness	Appetite
1	3						
2	4	4	5				
3	3	3	4				
4	2	3	4	4	2	3	1
5	2	3	4	4	3	4	2
6	3	2		5	3	4	4

Appendix 2.8

Table A2.5: Original versus imputed means and medians

	Original data		Imputed Data	
	Mean	Median	Mean(SE)	Median(SE)
Number of sedentary periods per day	6.21	6.00	6.13(0.14)	5.85(0.13)
Sustained inactivity bouts (day period)	2.32	1.51	2.39(0.05)	1.55(0.02)
Sustained inactivity bouts of minimum 15 minute (day period)	1.84	0.88	1.89(0.09)	0.93(0.00)
Duration of estimated sleep period	9.89	13.79	9.47(0.68)	13.62(0.12)
Minutes of inactivity per day	684.30	638.00	772.64(6118.67)	721.15(16221.82)
Light activity minutes per day	63.67	50.00	57.25(30.54)	48.05(47.21)
Moderate activity minutes per day	34.57	11.00	16.98(10.99)	11.80(31.54)
Average acceleration (least active 5 hours of the day)	1.20	0.00	0.68(0.04)	0(0)
Average acceleration (most active 5 hours of the day)	30.40	27.51	26.47(5.07)	22.73(8.62)

Appendix 2.9

Table A2.6: Regression output (non-significant results)

Variable	Phase (CI)	Total metacognition (researcher rated; CI)	Case
Sustained inactivity bouts (day period)	-2.13 (-5.58-1.33), p=0.20	0.21(-0.05-0.47), p=0.10	0.32(-0.54-1.18), p=0.44
Duration of estimated sleep period	-9.34 (-19.11-0.44)	0.79(-0.03-1.61)	1.63(-1.79-5.05)
Minutes of inactivity per day	32.26(-580.36-644.89)	15.31(-61.27-91.88)	31.59
Light activity minutes per day	-5.55(-59.87-48.76)	-1.05(-7.45-5.35)	9.22
Moderate activity minutes per day	0.92(-25.71-27.55)	-1.40(-4.68-1.87)	8.77
Average acceleration (least active 5 hours of the day)	0.27(-3.44-3.97)	-0.11(-0.54-0.31)	0.33
Average acceleration (most active 5 hours of the day)	-0.15(-21.20-20.91)	-1.00(-3.21-1.22)	5.17