cpe.psychopen.eu | ISSN 2625-3410 | September 2023 | Volume 5 | Issue 3 |

CLINICAL PSYCHOLOGY IN EUROPE

The Official Academic Journal of the
European Association of Clinical Psychology
and Psychological Treatment

PsychOpen GOLD

CLINICAL PSYCHOLOGY IN EUROPE

The Official Academic Journal of the
European Association of Clinical Psychology and Psychological Treatment

Editors-in-Chief

Winfried Rief
Division of Clinical Psychology and Psychological Treatment
Department of Psychology
Philipps-University of Marburg
Marburg, Germany
email: rief@uni-marburg.de

&

Cornelia Weise
Division of Clinical Psychology and Psychological Treatment
Department of Psychology
Philipps-University of Marburg
Marburg, Germany
email: weise@uni-marburg.de

Section Editors

Colette Hirsch
EACLIPT-Board | London, UK

Anton-Rupert Laireiter
Vienna, Austria

Tania Lincoln
Hamburg, Germany

Jolanda Meeuwissen
Utrecht, Netherlands

Omer Van den Bergh
Leuven, Belgium

Editorial Board

Gerhard Andersson
EACLIPT-Board | Linköping, Sweden

Claudi Bockting
EACLIPT-Board | Amsterdam, Netherlands

Cristina Botella
Castelló de la Plana, Spain

Per Carlbring
Stockholm, Sweden

Trudie Chalder
London, UK

Roman Cieślak
EACLIPT-Board | Warsaw, Poland

David Clark
Oxford, UK

Ioana Alina Cristea
Cluj-Napoca, Romania

Pim Cuijpers
Amsterdam, Netherlands

Daniel David
Cluj-Napoca, Romania

Céline Douilliez
EACLIPT-Board | Louvain-La-Neuve, Belgium

Anke Ehlers
Oxford, UK

Thomas Ehring
Munich, Germany

Giovanni Fava
Bologna, Italy

Jens Gaab
Basel, Switzerland

Martin Hautzinger
Tübingen, Germany

Dirk Hermans
Leuven, Belgium

Stefan Hofmann
Boston, USA

Emily Holmes
Stockholm, Sweden

Jutta Joormann
New Haven, USA

Maria Karekla
University of Cyprus, Cyprus

Andreas Maercker
EACLIPT-Board | Zurich, Switzerland

Robert Masten
Ljubljana, Slovenia

Lance McCracken
Uppsala, Sweden

Thomas Probst
Krems, Austria

Bernhard Strauß
Jena, Germany

Claus Vögele
Luxembourg, Luxembourg

PsychOpen GOLD

CLINICAL PSYCHOLOGY IN EUROPE

September 2023 | Volume 5 | Issue 3

Contents

Editorial

Growing (Up) in Times of Multiple Crises – A Call for Mental Health (Research) Action

Julia Asbrand, Tanja Michael, Hanna Christiansen, Gerhard Reese

Scientific Update and Overview

Research Into Evidence-Based Psychological Interventions Needs a Stronger Focus on Replicability

Helen Niemeyer, Christine Knaevelsrud, Robbie C. M. van Aert, Thomas Ehring

The current discussion surrounding replicability offers a unique opportunity to refine the criteria for replicability in research on psychological interventions. A research strategy to potentially increase replicability is proposed.

Research Articles

Loneliness Across the COVID-19 Pandemic: Risk Factors in Norwegian Young People

Mari Hysing, Keith J. Petrie, Allison G. Harvey, Kari-Jussie Lønning, Børge Sivertsen

Loneliness among students in 2022 is still high and not back to pre-pandemic levels, with online learning increasing the risk of student loneliness.

Worry Intervention in an Older Adult With a Persecutory Delusion: A Single Case Experimental Design

Poppy Brown, Anna Crabtree

A brief intervention focussed on reducing worry is successful both in reducing worry and paranoia in an older adult with a persecutory delusion.

Shame on Me? Love Me Tender! Inducing and Reducing Shame and Fear in Social Anxiety in an Analogous Sample

Jakob Fink-Lamotte, Jürgen Hoyer, Pauline Platter, Christian Stierle, Cornelia Exner

Uncover the potential of self-compassion interventions in addressing intense shame and their impact on psychological disorders like social anxiety.

Scheduled Support Versus Support on Demand in Internet-Delivered Cognitive Behavioral Therapy for Social Anxiety Disorder: Randomized Controlled Trial

Anton Käll, Cecilia Olsson Lynch, Kajsa Sundling, Tomas Furmark, Per Carlbring, Gerhard Andersson

An on-demand support solution produced similar symptom reductions compared to the standard, weekly therapist support for participants with social anxiety disorder.

Building an Early Warning System for Depression: Rationale, Objectives, and Methods of the WARN-D Study

Eiko I. Fried, Ricarda K. K. Proppert, Carlotta L. Rieble

Depression is a severe and debilitating disorder. To help prevent depression, the WARN-D study aims to build a personalized early warning system for depression.

...

PsychOpen GOLD is a publishing service by
Leibniz Institute for Psychology Information (ZPID), Germany.
www.leibniz-psychology.org

PsychOpen GOLD

CLINICAL PSYCHOLOGY IN EUROPE

September 2023 | Volume 5 | Issue 3

Systematic Review and Meta-analyses

Aetiological Understanding of Fibromyalgia, Irritable Bowel Syndrome, Chronic Fatigue Syndrome and Classificatory Analogues: A Systematic Umbrella Review

Maria Kleinstäuber, Andreas Schröder, Sarah Daehler, Karen Johanne Pallesen3 , Charlotte U. Rask, Mathias Sanyer, Omer Van den Bergh, Marie Weinreich Petersen, Judith G. M. Rosmalen

Research on aetiological factors in FSS and SSD is dominated by a focus on biological factors.

Clinical Psychology in Europe

PsychOpen GOLD

Editorial

Growing (Up) in Times of Multiple Crises – A Call for Mental Health (Research) Action

Julia Asbrand[1], Tanja Michael[2], Hanna Christiansen[3,4], Gerhard Reese[5]

[1] *Department of Psychology, Friedrich-Schiller-University Jena, Jena, Germany.* **[2]** *Department of Psychology, Saarland University, Saarbrucken, Germany.* **[3]** *Department of Psychology, University of Marburg, Marburg, Germany.* **[4]** *Deutsches Zentrum für Psychische Gesundheit (DZPG), Bochum/Marburg, Germany.* **[5]** *Department of Psychology, RPTU Kaiserslautern-Landau, Campus Landau, Landau, Germany.*

Clinical Psychology in Europe, 2023, Vol. 5(3), Article e12713, https://doi.org/10.32872/cpe.12713

Published (VoR): 2023-09-29

Corresponding Author: Julia Asbrand, Friedrich-Schiller-University Jena, Department of Psychology, Semmelweisstr. 12, 07743 Jena, Germany. E-mail: julia.asbrand@hu-berlin.de

The rock band *Fury in the slaughterhouse* sang that “every generation got its own disease”, however, remaining in that wording, the current generation of children and adolescents in Europe has to cope with several severe “diseases” at the same time: the war of Russia against Ukraine, the social and mental health consequences of the Covid-19-pandemic, the economic downturn, societal polarization, and last but not least, the twin crises of biodiversity loss and climate change.

Each of these crises have visible and measurable consequences, and some crises mutually reinforce each other. The climate crisis, for instance, has already brought irreversible damage to some societies and natural habitats around the world. According to the Intergovernmental Panel on Climate Change (IPCC, 2023), weather phenomena such as droughts, storms and floods will become more likely and more intense. It is estimated that people who are born today will experience heat waves up to seven times more often than people who are in their forties to sixties today (Thiery et al., 2021). Furthermore, climate change is linked to macroeconomic consequences negatively affecting the economic situations of states and individuals and is thus a driving force behind increasing poverty (e.g., Kotz et al., 2021). The climate crisis also accelerates biodiversity loss. Further, both environmental degradation as well as climate change undermine peace and increase the likelihood for conflict between groups, representing additional stressors for development both on an individual and a societal level (e.g., Palmer, 2022). Unsurprisingly, a recent meta-analysis shows that climate events are negatively correlated with mental health (Cuijpers et al., 2023), and a recent review demonstrates that the risk for mental health problems among young people is particularly high (Ma, Moore, & Cleary, 2022).

The Covid-19 pandemic decreased mental health in the general population, and younger age groups in particular (Santomauro et al., 2021). Furthermore, several studies show that the worsened mental health of young people remained up to two years after the onset of the pandemic (Hansen et al., 2023). A recent study among German adolescents also shows that both pandemic-related and climate-related distress are linked to more depression and anxiety symptoms and to reduced health-related quality of life. Distress related to the Russia-Ukraine war was associated with greater anxiety. Critically, these associations remained significant when controlling for important covariates (e.g., gender, distress caused by personal problems), showing that the crisis measures have incremental predictive value (Lass-Hennemann et al., 2023). However, self-efficacy and, though to a lesser extent, expressive flexibility were associated with better mental health. There are no studies yet examining how the war influences mental health of children and adolescents living in an area directly exposed to the war.

Despite the increased need to address mental health problems among young people (e.g. Deng et al., 2023), sufficient in-patient and outpatient mental health care systems are yet to be implemented. Actually, waiting times have doubled during the pandemic and low-threshold effective interventions are lacking (e.g. Overhage et al., 2023).

In sum, the current evidence suggests that global crises impact the mental health and healthy upbringing of young people. Therefore, policies should include interventions that help children and adolescents in particular to cope with the stress caused by the crises.

The Systemic Structure Underneath: What Can We (Not) Do?

However, it is not enough to only develop strategies to help individuals to cope better with stress or to increase mental health care capacities. Large scale crises, as the ones depicted above, have in common that they are usually a result of collective (in-)action and as such, these crises cannot be remedied by individual action alone: As an individual, I can neither address the pandemic, solve the climate crisis nor end a war all by myself. In other words: An individual can hardly experience self-efficacy when faced with these challenges.

These challenges can only be addressed by collective efforts, and these efforts must be implemented on various societal levels. For example, the multilevel model of societal change introduced by Geels and Schot (2007) represents a framework that helps to understand how and which levels of society need to be addressed to achieve societal change. In short, the model suggests that certain pressures such as climate change or resource scarcity open windows of opportunity within a political regime. Networks of innovators and groups with joint ideas can then use such a window to engender change within society. For example, the Fridays-for-Future movement did so, and changed the way climate change is treated in politics and society. Recent psychological models suggest the pathways through which collective and participatory efficacy beliefs can foster such collective actions (Hamann et al., 2023).

Building on these systemic considerations, clinical psychology and psychotherapy, both at the level of care and research, urgently need to move away from an exclusively individual approach to a consideration of the individual in the system, its structures, and their relevant life-environments such as schools or the work-place. How could this look like?

A Call for Health(y) Action

First of all, broader prevention structures focusing on systemic levels of mental health as well as self-efficacy (Lass-Hennemann et al., 2023) and adaptive coping are necessary (Mah et al., 2020). These need to be implemented and institutionalized within structural levels, and have to provide an outreach service – independent of youth's knowledge about health care structures or individual resources. Second, these prevention structures must also be flanked by measures that mitigate major risk factors for mental health and a negative trajectory when growing up. Specifically, these are all measures targeting financial and social injustices and inequity, i.e. installing appropriate climate protection measures, providing access to education and living above the poverty line is part of health care. Third, our research must do justice to the complex interplay between the individual and society, also in the field of mental health. This means that research approaches must be promoted in clinical psychology and psychotherapy that situate the individual in society and clarify the effects of society on the individual, e.g. which systemic structures limit treatment success? How do societal crises affect mental health?

We believe that the current age of multiple crises holds several challenges for mental health care and research, foremost the fact that we are affected by these crises as health care professionals and researchers. To adequately address these challenges, we need to expand (preventive) healthcare, intensify our research efforts and at the societal level to help young people grow up healthy.

Funding: The authors have no funding to report.

Acknowledgments: The authors have no additional (i.e., non-financial) support to report.

Competing Interests: The authors have declared that no competing interests exist.

References

Cuijpers, P., Miguel, C., Ciharova, M., Kumar, M., Brander, L., Kumar, P., & Karyotaki, E. (2023). Impact of climate events, pollution, and green spaces on mental health: An umbrella review of meta-analyses. *Psychological Medicine, 53*(3), 638–653. https://doi.org/10.1017/S0033291722003890

Deng, J., Zhou, F., Hou, W., Heybati, K., Lohit, S., Abbas, U., Silver, Z., Wong, C. Y., Chang, O., Huang, E., Zuo, Q. K., Moskalyk, M., Ramaraju, H. B., & Heybati, S. (2023). Prevalence of mental health symptoms in children and adolescents during the COVID-19 pandemic: A meta-analysis. *Annals of the New York Academy of Sciences, 1520*(1), 53–73. https://doi.org/10.1111/nyas.14947

Geels, F. W., & Schot, J. (2007). Typology of sociotechnical transition pathways. *Research Policy, 36*(3), 399–417. https://doi.org/10.1016/j.respol.2007.01.003

Hamann, K. R., Wullenkord, M. C., Reese, G., & van Zomeren, M. (2023). Believing that we can change our world for the better: A triple-A (agent-action-aim) framework of self-efficacy beliefs in the context of collective social and ecological aims. *Personality and Social Psychology Review.* Advance online publication. https://doi.org/10.1177/10888683231178056

Hansen, J., Galimov, A., Unger, J. B., Sussman, S. Y., & Hanewinkel, R. (2023). Mental health and physical complaints of German children and adolescents before and during the COVID-19 pandemic: A repeated cross-sectional study. *International Journal of Environmental Research and Public Health, 20*(5), Article 4478. https://doi.org/10.3390/ijerph20054478

Intergovernmental Panel on Climate Change. (2023). *Climate Change 2023: Synthesis Report.* Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (Eds.)]. https://doi.org/10.59327/IPCC/AR6-9789291691647

Kotz, M., Wenz, L., Stechemesser, A., Kalkuhl, M., & Levermann, A. (2021). Day-to-day temperature variability reduces economic growth. *Nature Climate Change, 11*, 319–325. https://doi.org/10.1038/s41558-020-00985-5

Lass-Hennemann, J., Sopp, M. R., Ruf, N., Equit, M., Schäfer, S. K., Wirth, B. E., & Michael, T. (2023). *Generation climate crisis, COVID-19, and Russia-Ukraine-War – Global crises and mental health in adolescents* [Preprint Version 1]. Research Square. https://doi.org/10.21203/rs.3.rs-3131485/v1

Ma, T., Moore, J., & Cleary, A. (2022). Climate change impacts on the mental health and wellbeing of young people: A scoping review of risk and protective factors. *Social Science & Medicine, 301*, Article 114888. https://doi.org/10.1016/j.socscimed.2022.114888

Mah, A. Y., Chapman, D. A., Markowitz, E. M., & Lickel, B. (2020). Coping with climate change: Three insights for research, intervention, and communication to promote adaptive coping to climate change. *Journal of Anxiety Disorders, 75*, Article 102282. https://doi.org/10.1016/j.janxdis.2020.102282

Overhage, L., Hailu, R., Busch, A. B., Mehrotra, A., Michelson, K. A., & Huskamp, H. A. (2023). Trends in acute care use for mental health conditions among youth during the COVID-19 pandemic. *JAMA Psychiatry, 80*(9), 924–932. https://doi.org/10.1001/jamapsychiatry.2023.2195

Palmer, L. (2022). Solving twin crises. *Nature Sustainability, 5*, 816–817. https://doi.org/10.1038/s41893-022-00973-x

Santomauro, D. F., Herrera, A. M. M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott, D. M., Abbafati, C., Adolph, C., Amlag, J. O., Aravkin, A. Y., Bang-Jensen, B. I., Bertolacci, G. J., Bloom, S. S., Castellano, R., Castro, E., Chakrabarti, S., Chattopadhyay, J., Cogen, R. M., Collins, J. K., . . . Ferrari, A. J. (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *Lancet, 398*(10312), 1700–1712. https://doi.org/10.1016/S0140-6736(21)02143-7

Thiery, W., Lange, S., Rogelj, J., Schleussner, C.-F., Gudmundsson, L., Seneviratne, S. I., Andrijevic, M., Frieler, K., Emanuel, K., Geiger, T., Bresch, D. N., Zhao, F., Willner, S. N., Büchner, M., Volkholz, J., Bauer, N., Chang, J., Ciais, P., Dury, M., . . . Wada, Y. (2021). Intergenerational inequities in exposure to climate extremes. *Science, 374*(6564), 158–160. https://doi.org/10.1126/science.abi7339

EACLIPT

Clinical Psychology in Europe (CPE) is the official journal of the European Association of Clinical Psychology and Psychological Treatment (EACLIPT).

leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.

Clinical Psychology in Europe

PsychOpen GOLD

Scientific Update and Overview

Research Into Evidence-Based Psychological Interventions Needs a Stronger Focus on Replicability

Helen Niemeyer[1], Christine Knaevelsrud[1], Robbie C. M. van Aert[2], Thomas Ehring[3]

[1] *Department of Clinical Psychological Intervention, Freie Universität Berlin, Berlin, Germany.* **[2]** *Department of Methodology and Statistics, Tilburg University, Tilburg, the Netherlands.* **[3]** *Department of Psychology, LMU Munich, Munich, Germany.*

Clinical Psychology in Europe, 2023, Vol. 5(3), Article e9997, https://doi.org/10.32872/cpe.9997

Received: 2022-07-29 • **Accepted:** 2023-07-17 • **Published (VoR):** 2023-09-29

Handling Editor: Cornelia Weise, Philipps-University of Marburg, Marburg, Germany

Corresponding Author: Helen Niemeyer, Division of Clinical Psychological Intervention, Department of Education and Psychology, Freie Universität Berlin, Schlossstr. 27, 12163 Berlin, Germany. Phone: 0049-30-838-54798. E-mail: helen.niemeyer@fu-berlin.de

Abstract

Background: It is a precondition for evidence-based practice that research is replicable in a wide variety of clinical settings. Current standards for identifying evidence-based psychological interventions and making recommendations for clinical practice in clinical guidelines include criteria that are relevant for replicability, but a better understanding as well refined definitions of replicability are needed enabling empirical research on this topic. Recent advances on this issue were made in the wider field of psychology and in other disciplines, which offers the opportunity to define and potentially increase replicability also in research on psychological interventions.

Method: This article proposes a research strategy for assessing, understanding, and improving replicability in research on psychological interventions.

Results/Conclusion: First, we establish a replication taxonomy ranging from direct to conceptual replication adapted to the field of research on clinical interventions, propose study characteristics that increase the trustworthiness of results, and define statistical criteria for successful replication with respect to the quantitative outcomes of the original and replication studies. Second, we propose how to establish such standards for future research, i.e., in order to design future replication studies for psychological interventions as well as to apply them when investigating which factors are causing the (non-)replicability of findings in the current literature.

Keywords

replicability, evidence-based interventions, criteria development

Highlights

- Refined replicability criteria used to identify empirically supported treatments are proposed.
- Concrete steps for refining replication in research on psychological interventions are proposed.
- A taxonomy of direct to conceptual replication adapted to research on interventions is provided.

Recent years have seen an increased focus on conceptual approaches to the replicability of research findings, and a growing number of empirical investigations on this issue, in the areas of psychology (Klein et al., 2014; Klein et al., 2018; Open Science Collaboration [OSC], 2015), economics (e.g., Camerer et al., 2016), epidemiology (e.g., Kaltiala-Heino, Työläjärvi, & Lindberg, 2019; Zisook et al., 2007) and medicine (Errington, Denis, Perfito, Iorns, & Nosek, 2021). Replicability refers to “the ability of a researcher to duplicate the results of a prior study if the same procedures are followed but new data are collected” (Bollen, Cacioppo, Kaplan, Kronsnick, & Olds, 2015; p. 3). Research related to psychological interventions has not paid the same level of attention to recent conceptual developments of replicability (Tackett et al., 2017) as seen in other fields. Yet the strong emphasis on providing evidence-based treatments in clinical psychology and psychiatry (e.g., Tolin et al., 2015) demands that clinical practice should be directly informed and guided by the best available empirical evidence on the efficacy of interventions, as typically collected in randomized controlled trials (RCTs). A precondition for evidence-based practice is that the research is replicable in a wide variety of clinical settings in order to demonstrate high external validity.

Low replicability in a research field may be partly due to so-called “hidden moderators” (Van Bavel, Mende-Siedlecki, Brady, & Reinero, 2016), which prevent the effect from being observed in a replication due to an (unobserved) moderator. Examples include characteristics of the clinical population to which the intervention is offered, treatment-related moderators, or differences in contextual variables. In other words, a study might be successfully replicated in a research outpatient clinic but not in a regular community clinic. Identifying hidden moderators is crucial in order to critically evaluate the generalizability of treatment effects to different clinical settings. “Direct” and “conceptual” are labels for replication studies depending on the similarity to the original study (LeBel et al., 2018; Zwaan, Etz, Lucas, & Donnellan, 2018). Direct replication studies allow to investigate the replicability of a study result, whereas conceptual replications serve to determine the generalizability. The relevance of replication categories has been shown in other fields, such as economics (Fiala, Neubauer, & Peters, 2022; Peters, Langbein,

& Roberts, 2018), where different replication rates were found depending on the definition of the replication studies. In order to define the similarity between original and replication study consensus on the most important characteristics is necessary. The ”constraints on generality” criteria (COG; Simons, Shoda, & Lindsay, 2017) help to explicitly determine the targeted population and the study procedures in order to define a direct replication as well as to identify hidden moderators in conceptual studies. A COG statement overcomes the ambiguity of classifying replications as direct or conceptual post hoc because it specifies the target populations for the original claim (Simons et al., 2017; Simons, Shoda, & Lindsay, 2018).

In addition, non-replicability of effects may also be caused by questionable research practices (QRPs; John, Loewenstein, & Prelec, 2012). QRPs comprise a range of activities that are not a research field´s best practices, such as flexibly analyzing data until the results are significant (called *p*-hacking; Whitt et al., 2022) or hypothesizing after the results are known (called HARKing; John et al., 2012). They cause an overrepresentation of statistically significant results in the literature. Performing multiple analyses in combination with selectively reporting statistically significant results increases the number of false-positive findings in the published literature (Forstmeier, Wagenmakers, & Parker, 2017; Simmons, Nelson, & Simonsohn, 2011) and biases effect size estimation. Other factors that may cause non-replicability are reporting errors or sampling error. Importantly, in a given case of non-replicability, more than one factor can be expected to be relevant (Nosek et al., 2022).

Closely related to replicability is reproducibility. Reproducibility is obtained when the reanalysis of the original data using the same procedures arrives at the same result (Maassen et al., 2020). This is also referred to as computational or analytic reproducibility (LeBel et al., 2018). Reproducibility in psychology was investigated by Artner and colleagues (2021) who found that 70% of the reported statistical results were reproducible. When comparing reproducibility rates across disciplines, it is important to note that the definitions of replicability and reproducibility differ across disciplines (Artner et al., 2021). To date, reproducibility attempts are highly uncommon in research on psychological interventions (see also, Sandve, Nekrutenko, Taylor, & Hovig, 2013).

Do Current Research Standards Pay Enough Attention to Replicability?

Current standards for investigating psychological interventions, identifying evidence-based interventions, and making recommendations for clinical practice in clinical guidelines include criteria that are relevant for the issue of replicability. For example, the criteria for empirically supported treatments (ESTs; David, Lynn, & Montgomery, 2018) were laid down by the American Psychological Association’s (APA) Division 12 in the early 1990s (Chambless & Hollon, 1998; for a recent revision, see Tolin et al., 2015).

According to these criteria, treatment effects must have been demonstrated in several independent studies, and a systematic evaluation of the methodological quality of studies as well as risk of bias needs to have been conducted, e.g., using the Cochrane risk-of-bias tool (ROB; Sterne et al., 2019) or the Grading of Recommendations, Assessment, Development and Evaluations (GRADE; Guyatt et al., 2008), consisting of six domains (e.g., risk of bias, [im-]precision of effect estimates). The need to critically assess study quality and the risk of bias has also led to the development of specific reporting standards for clinical trials, such as the Consolidated Standards of Reporting Trials (CONSORT; Schulz et al., 2010), and for reporting systematic reviews and meta-analyses, such as the Preferred Reporting Items for Systematic Reviews and Meta-Analyses" statement (PRISMA; Moher et al., 2015) or the "Meta-Analysis Reporting Standards" (MARS; American Psychological Association, 2020).

However, despite these important advances, the criteria used to identify ESTs and/or recommend clinical interventions for clinical guidelines currently have not yet been updated in line with the recent advances on replicability in the wider field of psychology and in other disciplines (Errington et al., 2021). Although the reporting standards and rating schemes address some of the variables that are relevant to assess (the lack of) replicability in studies on psychological interventions (i.e., pre-specification of the hypotheses and statistical methods, examining publication bias and heterogeneity), they neither include all of the relevant aspects nor do they make an explicit distinction between different types of replication (e.g., direct versus conceptual replications) or specify statistical criteria for a successful replication. A refinement of the criteria for replication in research on psychological interventions and specific suggestions for their application are therefore required. Moreover, an assessment of QRPs, reporting error and demands for pre-registration are currently not included in the quality assessment of clinical studies.

Currently there are only few investigations of the replicability of studies on psychological interventions. One exception is Sakaluk et al. (2019) who systematically examined the evidential value of treatments that have been classified as ESTs by standard criteria. They also applied Schimmack's replicability index (R-index, Schimmack, 2016), which focuses on statistical significance, and statistical power, as well as Bayesian meta-analysis. Results showed that statistical power and replicability estimates were low. Moreover, differences in the level of empirical support according to EST criteria did not parallel differences in indices of statistical power or replicability. Based on their analysis, the authors argued that higher methodological standards are necessary in research on psychological interventions, including sufficient statistical power and standards for reporting descriptive and inferential statistics.

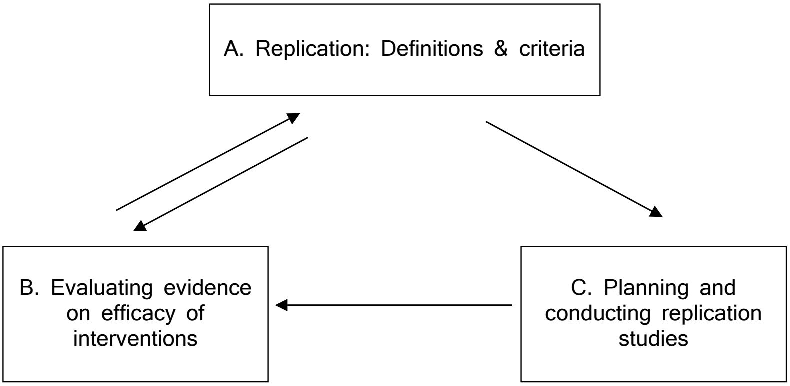
In line with Sakaluk and colleagues (2019) as well as with the recommendations developed in other areas of psychology and beyond (Ioannidis, 2008; Valentine, 2009), we suggest that there is a need to enhance the replicability of research into psychological interventions and therefore propose to refine the definition of and criteria for replicability in this field. To this aim, some of the developments and resources from other areas will be adopted and, if necessary, adapted to the specificities of research on psychological interventions, as well as the given criteria and definitions refined.

Proposing a Research Strategy for Assessing, Understanding, and Improving Replicability in Research Evaluating Psychological Interventions

To improve the current situation, we propose progress in three interrelated areas (A – C; see Figure 1). The concrete steps that need to be taken are described in the following sections.

Figure 1

A Strategy for Assessing, Understanding, and Improving Replicability in Research on Psychological Interventions



A. Replication: Definitions and Criteria

First, the definition of replication currently used in research on psychological interventions is refined, based on a taxonomy of different study design types of replication, study characteristics that increase the trustworthiness of results, and statistical criteria for (un-)successful replication. At a minimum, we suggest three aspects to be crucial:

1. Taxonomy of Replication

Refining replication in research on psychological interventions is a complex endeavor. The definition of replication as aiming to duplicate the results of an original study by applying the same procedures to a new sample (Bollen et al., 2015) provides no specific criteria as to what constitutes "the same procedure" with respect to the characteristics of an original study. Similarly, the EST criteria that treatment effects need to be demonstrated in several independent studies do not specify any details of the study designs of the required independent studies (Tolin et al., 2015).

Attempts to refine the concept of replication have been made in other areas of psychology and social sciences. We adopt the approach of LeBel et al. (2018) who provide a replication taxonomy ranging from direct to conceptual replication, depending on the degree of similarity between an original and a replication study according to several design facets, such as the operationalization of the independent and dependent variables, or investigator independence. To investigate the replicability of a treatment effect, direct replications are necessary. Conceptual replications cannot falsify the hypothesis of replicability, but can, on the other hand, help to evaluate the boundary conditions of treatment effects, the generalizability of intervention effects to different contexts, and/or the mechanisms of change underlying treatment effects. They can help to answer the question of whether (and which) hidden moderators are a cause of low replicability in “combination” with direct replications.

In order to define the characteristics that need to be identical for a study to qualify as direct replication, the constraints on generality criteria (COG; Simons et al., 2017) are applied. The COG criteria provide a general scheme for which characteristics of study participants (the targeted population), study material and procedures, and the temporal specificity of an effect are necessary to be kept the same for a replication study to be an exact replication. Principles for choosing variables for the COG should be known empirical or theoretical boundary conditions, conditions that are tied to the substance of the study, and factors that experts consider to be important.

The taxonomy suggested by LeBel and colleagues (2018) combined with the COG results in a continuum from direct to conceptual replication that can be pre-specified. The dimensions underlying the classification of replication types should include procedural details (e.g., diagnostic instruments, blinding of assessors, unconcealed allocation/risk of bias), statistical methods, contextual variables (e.g., cultural context), therapist-related factors (manual adherence), and researcher-related factors (e.g., allegiance, conflicts of interest), all of which are also potential moderator variables.

Consider, for instance, a case in which a newly developed intervention for depression is first tested against a waitlist condition (WL) and is found to be superior. A subsequent study replicates the initial study, but compares the same intervention to treatment as usual (TAU). A direct replication of the newly developed intervention for depression would need to consist of a second comparison to WL, whereas the use of a different control condition (or treatment delivery in a natural setting, or applying the intervention over the internet etc.) constitutes a conceptual replication that already tells us something about the generalizability of the intervention effects and the relative efficacy of the new treatment. As another example, we might consider a case in which a new 12-session treatment for panic disorder is favorably tested against WL. A subsequent study also compares this new treatment to WL but uses a protocol that involves only 10 sessions, is conducted in a different country, and examines a slightly older patient population; and this second study does not find the treatment to be efficacious. Is this a failed replication study? Due to the lack of clear criteria, we are not currently able to provide a definitive answer to this question. With so many changes at once, we will never know why it did not replicate. Therefore, we need the changes to be decided on and documented more specifically; ideally, replication studies should change on one dimension at a time, so that differences in effects can be clearly attributed.

Incentives for authors for the use of a COG statement integrated into the taxonomy by LeBel and colleagues (2018) could be a protection from overly broad claims, a higher likelihood of successful replications, and inspiring follow-up studies that built upon the findings. Editors and reviewers could request a COG statement. Incentives for editors could be to have an equivalent measure to evaluate all papers, and for reviewers to have a measure for quality control, whereas for readers it helps to learn about the generality of the claims of a study (Simons et al., 2017).

2. Study Characteristics That Increase the Trustworthiness of Results

Although some important methodological factors are included in current standards of study quality assessment, there is evidence that many intervention studies fall short of characteristics that increase the trustworthiness of results. Moreover, QRPs and publication bias distort the literature and limit the replicability of studies. In addition to the existing guidelines we propose to include the following issues:

- An assessment of reporting errors should be conducted. For consistency checks of *p* values, “statcheck” can be applied (Epskamp & Nuijten, 2016).
- Pre-registration should be mandatory. The study design and analysis plan need to be pre-specified and saved in a public registry or published prior to data collection. Pre-registration is a measure to enhance transparency, document timestamped decisions, helping to differentiate between confirmatory and exploratory analyses, and for reducing *p*-hacking and HARKing. Alternatively, registered reports (RRs) are a sensible publishing format that reduces QRPs and publication bias because in RRs the peer review is conducted prior to the data collection. This emphasizes the research question and the quality of methodology instead of the significance of the results (Chambers & Tzavella, 2022). Checklists for pre-registration and recommendations for RRs have been developed in the wider field of psychology to enhance the quality of reports and pre-registrations (Nuijten, Hartgerink, van Assen, Epskamp, & Wicherts, 2016). Developments in adjacent fields are ahead, such as in biomedical research where journals banded together to make registration mandatory (Siebert et al., 2020; ClinicalTrials.gov). Registered reports and replication reports are a promising format also for clinical psychological journals.

- A systematic assessment of whether the information provided in a pre-registration is sufficient should always be conducted and should be considered in the EST criteria or guidelines.
- It should be assessed whether the final study report matches the pre-registered plan. We do acknowledge, though, that this places an extra burden on reviewers, who need to spend more time reviewing a manuscript. To reduce this burden journals can invite specialized reviewers to specifically review open science aspects of the manuscript, such as whether the pre-registration matches the final study report or checking any shared materials.
- Open data and open materials should become standard to enhance transparency. Replication studies benefit to a large extent from open data and materials. However, it should be noted that open data and materials is not a prerequisite for replicating studies (Buzbas, Devezer, & Baumgaertner, 2023). If highly sensitive data present challenges to open data principles, restricted access to data, e.g. according to the different access categories of the German Psychological Association (DGPs[1]), is also a viable alternative. This is in line with the standards of the American Psychological Association (2020), which invites researchers to share their data. It should be motivated if data cannot be shared due to ethical or legal constraints, e.g. due to participant confidentiality or missing consent. Open material and sensitive material with restricted access can both be stored in repositories, such as the Open Science Framework (OSF; osf.io).

3. Criteria for Successful Replication

As described in the taxonomy of replication, exact versus conceptual replication studies provide different information in case of replication success or failure. For example, when a conceptual replication study shows a failure of replication, this might be the result of hidden moderators. However, criteria are necessary for determining when (both direct and conceptual) replication studies are a success or failure. This conceptual issue has also not been explicitly addressed in mental health research to date, i.e. what defines a successful replication with respect to the statistical outcome of both the original and the replication study. That is, in addition to the definition of the study design as direct or conceptual replication, we propose criteria for the comparison of the quantitative results of an original and a replication study and the assessment of the replication of the study results as successful or failure, which are currently missing in research on psychological interventions.

Recent large-scale replication studies have proposed and comparatively evaluated different criteria, such as statistical significance, i.e., a study is deemed to be replicated if both the original study and the replication are statistically (non-)significant, or the

---

1) https://zwpd.transmit.de/images/zwpd/dienstleistungen/ethikkommission/vorlage_opendata_v1.docx

direction of both effect estimates is the same (OSC, 2015). However, an application of criteria for (un)successful replication in research on psychological interventions is lacking (see also Nosek et al., 2022).

Given that multiple statistical options to determine replication success exist (OSC, 2015; Zwaan, Etz, Lucas, & Donnellan, 2018) and that there is no consensus for one particular method, we provide a short overview of the most relevant ones: Both original and replication studies are statistically (non-)significant, the direction of both effect estimates is the same, the original effect falls within the confidence interval of the replication, original and replication result are combined and significance is assessed (OSC, 2015), statistical consistency between the original study and replications is evaluated in multisite replication projects (Mathur & VanderWeele, 2020), the small telescopes approach (Simonsohn, 2015), sceptical p-value (Held, 2020), and replication Bayes factor (Ly, Etz, Marsman, & Wagenmakers, 2019). These criteria represent the currently most prominent options for evaluating replicability. Recently, a comparison of seven approaches (significance, small telescopes, classical and Bayesian meta-analysis, Bayes factor and replication Bayes factor, as well as skeptical p-value (Held, 2020) has been conducted (Muradchanian, Hoekstra, Kiers, & van Ravenzwaaij, 2021). According to the authors, Bayesian metrics as well as meta-analytic methods were found to perform slightly better than the other approaches in terms of true and false positives rates. That is, a positive replication result is observed when the underlying true effect is non-zero or when the true effect is practically zero under different levels of publication bias in a simulation study. When evaluating replicability in research on psychological interventions, we suggest applying multiple methods, all of which should be preregistered before conducting the study. Researchers should come to conclusions based on the results of all the methods, as they perform quite similarly. Moreover, applying more methods also provides more information.

All criteria presented in the three categories taxonomy of replication, study characteristics that increase the trustworthiness of results, and criteria for successful replication are provided in an info box (see Table 1). We exemplarily propose up to three specific criteria for each COG subdomain. This list is not exhaustive, because study designs and research foci differ considerably. We recommend that researchers adapt the COG specifically to the study designs that are utilized in their research domains.

Table 1

Info Box for Replication Studies in Clinical Psychology

Overall domains / Subdomains

1. Taxonomy of replication: Constraints on generality (COG)

Participants[a]
- Diagnoses
- Symptom severity
- Comorbidity

Materials / stimuli[a]
- Manual used
- Adherence to manual
- Therapist training / supervision

Procedure[a]
- Primary and secondary outcomes
- Type of assessment (e.g., clinician-based vs. self-rated)
- Type of allocation

Historical / temporal specificity[b]
- Changes in diagnostic criteria (e.g. in DSM)
- Common use of cellphones or internet access for app- and browser-based interventions / blended approaches

2. Study characteristics that increase the trustworthiness of results

Scales[c] for quality assessment used (according to study type)

Are reporting errors absent in the study?

Preregistration
- Is a study pre-registered or is it a registered report?
- Are there sufficient details in the pre-registration/registered report?
- Do the analyses in the pre-registration match those in the final study report?

3. Criteria for successful replication: Methods to consider

Are the data and study materials openly available?

Are both original and replication study statistically significant?

Are the effect sizes of both the original and replication study in the same direction?

Does the effect size of the original study lie in the CI of the replication?

Is the meta-analytic effect size of combining the original and replication study statistically significant?

Overall domains / Subdomains
Is the effect size of the original study consistent with the replications in a multisite replication project (Mathur & VanderWeele, 2020)?
Small telescopes approach (Simonsohn, 2015): Is the replication effect size not significantly smaller than an effect size that would have 33% statistical power based on the sample size of the original study?
Replication Bayes factor (Verhagen & Wagenmakers, 2014; Wagenmakers, Verhagen, & Ly, 2016: Is there more evidence that the effect size of the replication is a null effect compared to the effect observed in the original study?

Note. DSM = Diagnostic and Statistical Manual; CI = Confidence interval.
[a] The proposed specific criteria are exemplary and not exhaustive. [b] This category takes into account that norms and standards change over time, and studies should be evaluated according to the respective historical period.
[c] The quality assessment should be conducted according to the specific scale that is used.

B. Evaluating Evidence on Efficacy of Interventions

Beyond establishing standards for future research, it is also important to understand which factors are causing the (non-)replicability of findings in the current literature by systematically investigating moderators of treatment effects. Specifically, the relative contributions of the different variables outlined in Section A to replication success (outcome) are of interest, e.g. study quality, the type of replication design, and contextual variables. Pre-registration and a taxonomy of replication should also be systematically integrated into the classification of ESTs, clinical guidelines, and meta-analyses to enhance the transparency and methodological comparability. In addition, differences between preregistered/replicated studies and other studies should be studied.

Moderator analyses can best be addressed with meta-analytic methods. For example, the efficacy of some interventions may be highly dependent on context variables, e.g., successful replication may only be demonstrated in very direct replication designs and may have low generalizability to different contexts. Other interventions may be more context-independent, with effects being replicated even in less strict settings regarding patient or therapist characteristics or modes of treatment delivery. That is, the criteria for replication outlined above should be related to the evaluation of studies as ESTs and considered when summarizing studies in meta-analyses. Importantly, findings from this line of research can then be useful to further refine the replication concept and criteria (A). For example, if a particular therapist characteristic is not relevant for determining the replicability, it no longer needs to be taken into account when evaluating whether a study is a direct or conceptual replication.

Moderators can also include variables that are typically used to address meta-scientific questions, for example whether a study was pre-registered or provides open data. Thus, investigating pre-registration as moderator in meta-analyses against the background of replicability can shed light on whether pre-registered studies differ from non-pre-registered studies not only in terms of treatment efficacy and study quality, but also in the replicability of their results.

C. Planning and Conducting New Replication Studies

The new definitions and criteria (A) should be used to design future replication studies for psychological interventions in order to test the consistency of treatment effects by means of direct replication studies, as well as the generalizability of findings to varying contexts on the basis of an explicit taxonomy of replication. To guide future replication research, the taxonomy of different types of replication, including the relevant dimensions of similarity vs. dissimilarity of research design features and a COG statement, tailored to research on psychological interventions, should be applied. Researchers should start by directly replicating an original treatment effect in order to investigate whether the effect exists. Then, to examine the generalizability and detect hidden moderators, they should move on to conceptual replication studies, in which they modify important aspects of the study design (e.g., treatment manual used, characteristics of treatment delivery, definition of outcome, comparison condition, and contextual factors). Depending on how many and which variables in the COG are kept equal, the similarity of replication studies along the continuum from direct to conceptual replications should be varied. Thereby it can be determined in a direct replication whether an effect exists, and its boundary conditions and mechanisms can be identified in conceptual replications. Thus, the distinction between direct and conceptual replication studies will be helpful for assessing the heterogeneity of findings for a particular intervention. That is, conceptual replications will test whether the proposed constraints on generality are accurate, leading to a more refined understanding of the robustness of effects. A systematic program of research should evaluate how the size of an effect varies as a function of those constraints (Simons et al., 2018).

An important first step is to conduct an exact replication study to confirm the result of the original study. Second, in order to identify the most important hidden moderators assessed conceptual replications and also meta-analyses should be conducted, once a sufficient number of replication studies has been conducted where as rule-of-thumb can be used that 5 to 10 studies are needed per included moderator in a meta-analysis (van Houwelingen, Arends, & Stijnen, 2002). An agreed set of quality standards and criteria based on the COG concept that must be included in clinical trial reports should be established and constantly refined. The criteria and quality standards will inform future replication studies, and should also be taken into account by experts evaluating the current state of evidence of an intervention, e.g. when developing clinical guidelines or establishing EST.

In the long term, the adoption of COG statements will lead to a more cumulative understanding of the scope of the effects of psychological interventions.

Conclusion

The current gold standard in evidence-based psychological treatments can be criticized for not paying sufficient attention to replicability. The current discussion surrounding replicability and reproducibility (Ioannidis, 2012; Munafò et al., 2017) offers the opportunity to define and potentially increase replicability also in mental health research. The development of an explicit concept and taxonomy of replication will enable the classification of studies investigating clinical interventions with respect to their similarity with original studies and will aid in planning and conducting replication studies in the future. The criteria themselves need to be continuously updated based on advances in replicability research in other areas and informed by emerging evidence regarding (moderators of) replicability in mental health research.

However, also a number of limitations have to be noted. Even if an effect is true, it is possible to fail to replicate due to seemingly innocuous differences in the implementation of the study (i.e. due to “hidden moderators”). Small variations in studies are unavoidable and exact replication is strictly impossible. Baribault and colleagues (2018) suggest to randomize variables that may be moderators of an effect in replication studies in order to test the robustness and generalizability of an effect. They propose a random selection of potential moderators, that is characteristics of the design that are not supposed to make a difference. If characteristics do not affect the results, this means that the results are more generalizable and to alter minor things should not matter. This is suggested for experimental research, e.g. different implementations of the same stimulus could be used to study whether the results are robust. However, as a large number of studies is necessary for this approach, it is not applicable to RCTs on psychological interventions. Compared to research in social psychology, studies in research on psychological interventions are much more costly and time-consuming, which makes it more difficult to study replicability. The question of how much money and effort researchers should spend on studying replicability given that conducting such studies is expensive in clinical psychology is related to the decision when to move on to other research topics, because studying replicability means at the same time that less scientific progress with respect to new findings will be made. This demonstrates that not all recommendations from social psychology are applicable in clinical psychology.

Based on this we would like to invite the readers to engage in discussions about the concrete criteria and next steps that we proposed. Designing replication studies should be based on empirical evidence and on theoretical predictions (Simons et al., 2018) and considered to be a collective research enterprise.

Funding: This research received funding from the Berlin University Alliance, an excellence initiative of the German Research Foundation (312_OpenCall_3).

Acknowledgments: We would like to thank Sarah Mannion for language editing.

Competing Interests: The authors have declared that no competing interests exist.

References

Artner, R., Verliefde, T., Steegen, S., Gomes, S., Traets, F., Tuerlinckx, F., & Vanpaemel, W. (2021). The reproducibility of statistical results in psychological research: An investigation using unpublished raw data. *Psychological Methods, 26*(5), 527–546. https://doi.org/10.1037/met0000365

American Psychological Association. (2020). *Publication manual of the American Psychological Association: The official guide to APA style* (7th ed.). American Psychological Association.

Baribault, B., Donkin, C., Little, D. R., Trueblood, J. S., Oravecz, Z., van Ravenzwaaij, D., White, C. N., De Boeck, P., & Vandekerckhove, J. (2018). Metastudies for robust tests of theory. *Proceedings of the National Academy of Sciences of the United States of America, 115*(11), 2607–2612. https://doi.org/10.1073/pnas.1708285114

Bollen, K., Cacioppo, J. T., Kaplan, R. M., Kronsnick, J. A., & Olds, J. L. (2015). *Social, behavioral, and economic sciences perspectives on robust and reliable science*. National Science Foundation. https://www.nsf.gov/sbe/AC_Materials/SBE_Robust_and_Reliable_Research_Report.pdf

Buzbas, E. O., Devezer, B., & Baumgaertner, B. (2023). The logical structure of experiments lays the foundation for a theory of reproducibility. *Royal Society Open Science, 10*(3), Article 221042. https://doi.org/10.1098/rsos.221042

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., & Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science, 351*(6280), 1433–1436. https://doi.org/10.1126/science.aaf0918

Chambers, C. D., & Tzavella, L. (2022). The past, present and future of Registered Reports. *Nature Human Behaviour, 6*, 29–42. https://doi.org/10.1038/s41562-021-01193-7

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66*(1), 7–18. https://doi.org/10.1037/0022-006X.66.1.7

David, D., Lynn, S. J., & Montgomery, G. H. (2018). *Evidence-based psychotherapy: The state of the science and practice*. Wiley.

Epskamp, S., & Nuijten, M. B. (2016). *Statcheck: Extract statistics from articles and recompute* p *values* (R package Version 1.2.2) [Computer software]. https://CRAN.R-project.org/package=statcheck

Errington, T. M., Denis, A., Perfito, N., Iorns, E., & Nosek, B. A. (2021). Challenges for assessing replicability in preclinical cancer biology. *eLife, 10*, Article e67995. https://doi.org/10.7554/eLife.67995

Fiala, N., Neubauer, F., & Peters, J. (2022). *Do economists replicate?* (Ruhr Economic Papers, No. 939). http://hdl.handle.net/10419/250076

Forstmeier, W., Wagenmakers, E.-J., & Parker, T. H. (2017). Detecting and avoiding likely false-positive findings – A practical guide. *Biological Reviews of the Cambridge Philosophical Society, 92*(4), 1941–1968. https://doi.org/10.1111/brv.12315

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. (2008). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *British Medical Journal, 336*(7650), 924–926. https://doi.org/10.1136/bmj.39489.470347.AD

Held, L. (2020). A new standard for the analysis and design of replication studies. *Journal of the Royal Statistical Society Series A: Statistics in Society, 183*(2), 431–448. https://doi.org/10.1111/rssa.12493

Ioannidis, J. P. A. (2008). Why most discovered true associations are inflated. *Epidemiology, 19*(5), 640–648. https://doi.org/10.1097/EDE.0b013e31818131e7

Ioannidis, J. P. A. (2012). Why science is not necessarily self-correcting. *Perspectives on Psychological Science, 7*(6), 645–654. https://doi.org/10.1177/1745691612464056

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science, 23*(5), 524–532. https://doi.org/10.1177/0956797611430953

Kaltiala-Heino, R., Työläjärvi, M., & Lindberg, N. (2019). Gender dysphoria in adolescent population: A 5-year replication study. *Clinical Child Psychology and Psychiatry, 24*(2), 379–387. https://doi.org/10.1177/1359104519838593

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., . . . Nosek, B. A. (2014). Investigating variation in replicability. *Social Psychology, 45*(3), 142–152. https://doi.org/10.1027/1864-9335/a000178

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., Aveyard, M., Axt, J. R., Babalola, M. T., Bahník, S., Batra, R., Berkics, M., Bernstein, M. J., Berry, D. R., Bialobrzeska, O., Binan, E. D., Bocian, K., Brandt, M. B., Busching, R., . . . Nosek, B. A. (2018). Many Labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science, 1*(4), 443–490. https://doi.org/10.1177/2515245918810225

LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). A unified framework to quantify the credibility of scientific findings. *Advances in Methods and Practices in Psychological Science, 1*(3), 389–402. https://doi.org/10.1177/2515245918787489

Ly, A., Etz, A., Marsman, M., & Wagenmakers, E.-J. (2019). Replication Bayes factors from evidence updating. *Behavior Research Methods, 51*(6), 2498–2508. https://doi.org/10.3758/s13428-018-1092-x

Maassen, E., van Assen, M. A. L. M., Nuijten, M. B., Olsson-Collentine, A., & Wicherts, J. M. (2020). Reproducibility of individual effect sizes in meta-analyses in psychology. *PLoS ONE, 15*(5), Article e0233107. https://doi.org/10.1371/journal.pone.0233107

Mathur, M. B., & VanderWeele, T. J. (2020). New statistical metrics for multisite replication projects. *Journal of the Royal Statistical Society Series A: Statistics in Society, 183*(3), 1145–1166. https://doi.org/10.1111/rssa.12572

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., & the PRISMA-P Group. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews, 4*(1), Article 1. https://doi.org/10.1186/2046-4053-4-1

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour, 1*(1), Article 0021. https://doi.org/10.1038/s41562-016-0021

Muradchanian, J., Hoekstra, R., Kiers, H., & van Ravenzwaaij, D. (2021). How best to quantify replication success? A simulation study on the comparison of replication success metrics. *Royal Society Open Science, 8*(5), Article 201697. https://doi.org/10.1098/rsos.201697

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Struhl, M. K., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology, 73*(1), 719–748. https://doi.org/10.1146/annurev-psych-020821-114157

Nuijten, M. B., Hartgerink, C. H. J., van Assen, M. A. L. M., Epskamp, S., & Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods, 48*(4), 1205–1226. https://doi.org/10.3758/s13428-015-0664-2

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), Article aac4716. https://doi.org/10.1126/science.aac4716

Peters, J., Langbein, J., & Roberts, G. (2018). Generalization in the tropics – Development policy, randomized controlled trials, and external validity. *The World Bank Research Observer, 33*(1), 34–64. https://doi.org/10.1093/wbro/lkx005

Sakaluk, J. K., Williams, A. J., Kilshaw, R. E., & Rhyner, K. T. (2019). Evaluating the evidential value of empirically supported psychological treatments (ESTs): A meta-scientific review. *Journal of Abnormal Psychology, 128*(6), 500–509. https://doi.org/10.1037/abn0000421

Sandve, G. K., Nekrutenko, A., Taylor, J., & Hovig, E. (2013). Ten simple rules for reproducible computational research. *PLoS Computational Biology, 9*(10), Article e1003285. https://doi.org/10.1371/journal.pcbi.1003285

Schimmack, U. (2016). *The replicability-index: Quantifying statistical research integrity.* https://wordpress.com/post/replication-index.wordpress.com/920

Schulz, K. F., Altman, D. G., Moher, D., & the CONSORT Group. (2010). CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *Trials, 11*(1), Article 32. https://doi.org/10.1186/1745-6215-11-32

Siebert, M., Gaba, J. F., Caquelin, L., Gouraud, H., Dupuy, A., Moher, D., & Naudet, F. (2020). Data-sharing recommendations in biomedical journals and randomised controlled trials: An audit of journals following the ICMJE recommendations. *BMJ Open, 10*(5), Article e038887. https://doi.org/10.1136/bmjopen-2020-038887

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on Generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science, 12*(6), 1123–1128. https://doi.org/10.1177/1745691617708630

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2018). Constraints on generality statements are needed to define direct replication. *Behavioral and Brain Sciences, 41*, Article e148. https://doi.org/10.1017/S0140525X18000845

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science, 26*(5), 559–569. https://doi.org/10.1177/0956797614567341

Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H.-Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., . . . Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ, 366*, Article l4898. https://doi.org/10.1136/bmj.l4898

Tackett, J. L., Lilienfeld, S. O., Patrick, C. J., Johnson, S. L., Krueger, R. F., Miller, J. D., Oltmanns, T. F., & Shrout, P. E. (2017). It's time to broaden the replicability conversation: Thoughts for and from clinical psychological science. *Perspectives on Psychological Science, 12*(5), 742–756. https://doi.org/10.1177/1745691617690042

Tolin, D. F., McKay, D., Forman, E. M., Klonsky, E. D., & Thombs, B. D. (2015). Empirically supported treatment: Recommendations for a new model. *Clinical Psychology: Science and Practice, 22*(4), 317–338. https://doi.org/10.1111/cpsp.12122

Valentine, J. C. (2009). Judging the quality of primary research. In H. Cooper, L. V. Hedges, & J. D. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (Vol. 2, pp. 129-146). Russel Sage Foundation.

Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016). Contextual sensitivity in scientific reproducibility. *Proceedings of the National Academy of Sciences of the United States of America, 113*(23), 6454–6459. https://doi.org/10.1073/pnas.1521897113 van Houwelingen, H. C., Arends, L. R., & Stijnen, T. (2002). Advanced methods in meta-analysis: Multivariate approach and meta-regression. *Statistics in Medicine, 21*(4), 589–624. https://doi.org/10.1002/sim.1040

Verhagen, J., & Wagenmakers, E.-J. (2014). Bayesian tests to quantify the result of a replication attempt. *Journal of Experimental Psychology: General, 143*(4), 1457–1475. https://doi.org/10.1037/a0036731

Wagenmakers, E.-J., Verhagen, J., & Ly, A. (2016). How to quantify the evidence for the absence of a correlation. *Behavior Research Methods, 48*(2), 413–426. https://doi.org/10.3758/s13428-015-0593-0

Whitt, C. M., Miranda, J. F., & Tullett, A. M. (2022). History of replication failures in psychology. In W. O'Donohue, A. Masudo, & S. Lilienfeld (Eds.), *Avoiding questionable research practices in applied psychology* (pp. 73-97). Springer.

Zisook, S., Rush, A. J., Lesser, I., Wisniewski, S. R., Trivedi, M., Husain, M. M., Balasubramani, G. K., Alpert, J. E., & Fava, M. (2007). Preadult onset vs. adult onset of major depressive disorder: A replication study. *Acta Psychiatrica Scandinavica, 115*(3), 196–205. https://doi.org/10.1111/j.1600-0447.2006.00868.x

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences, 41*, Article e120. https://doi.org/10.1017/S0140525X17001972

EACLIPT

Clinical Psychology in Europe (CPE) is the official journal of the European Association of Clinical Psychology and Psychological Treatment (EACLIPT).

leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.

Clinical Psychology in Europe

PsychOpen GOLD

Research Articles

Check for updates

Loneliness Across the COVID-19 Pandemic: Risk Factors in Norwegian Young People

Mari Hysing[1], Keith J. Petrie[2], Allison G. Harvey[3], Kari-Jussie Lønning[4,5], Børge Sivertsen[6,7]

[1] *Department of Psychosocial Science, Faculty of Psychology, University of Bergen, Bergen, Norway.* **[2]** *Department of Psychological Medicine, University of Auckland, Auckland, New Zealand.* **[3]** *Department of Psychology, University of California, Berkeley, CA, USA.* **[4]** *Modum Bad Psychiatric Hospital, Vikersund, Norway.* **[5]** *The Student Welfare Organization in Oslo and Akershus (SiO), Oslo, Norway.* **[6]** *Department of Health Promotion, Norwegian Institute of Public Health, Bergen, Norway.* **[7]** *Department of Research & Innovation, Helse-Fonna HF, Haugesund, Norway.*

Clinical Psychology in Europe, 2023, Vol. 5(3), Article e10483, https://doi.org/10.32872/cpe.10483

Received: 2022-10-17 • **Accepted:** 2023-07-17 • **Published (VoR):** 2023-09-29

Handling Editor: Cornelia Weise, Philipps-University of Marburg, Marburg, Germany

Corresponding Author: Mari Hysing, Department of Psychosocial Science, Faculty of Psychology, University of Bergen, Post box 7807, 5020 Bergen, Norway. +47 55 58 86 98. E-mail: mari.hysing@uib.no

Abstract

Background: There is evidence of increasing levels of loneliness in Norwegian young people before the COVID-19 pandemic. It is not clear how the COVID-19 pandemic, and the associated necessary restrictions, impacted on these trends.

Aims: To examine how loneliness in young people changed across the pandemic, how loneliness relates to demographic characteristics and how different pandemic restrictions impacted loneliness.

Method: We analyzed data from three waves of a Norwegian national higher education student survey (the SHoT-study). Data was examined from 2018 from a total of 49,836 students, 2021 from 62,212 students, and from 2022 from 53,362 (response rates 31-35%). Loneliness was measured by "The Three-Item Loneliness Scale" (T-ILS).

Results: There was a sharp increase in loneliness from 2018 to 2021, and a reduction in levels of loneliness in 2022, although at increased levels compared to prior to the pandemic. Females consistently report higher levels of loneliness than males, with a larger difference during the peak of the pandemic. There were higher rates of loneliness in geographical regions with higher COVID rates and greater pandemic-related restrictions during 2021. Loneliness was lower among students reporting more days on campus in 2021 and for those with lectures on campus in 2022, both with dose-response associations.

Conclusions: Loneliness is a major public health problem among young adults in higher education. Loneliness increased during the pandemic and has decreased but is still not back to pre-pandemic levels. The results suggest the importance of open campuses and in-person lectures, for increased social connectedness among young people.

Keywords

loneliness, social isolation, mental health, young adult, COVID-19

Highlights

- Loneliness increased among university students from 2018 to 2021 and decreased from 2021 to 2022, but was still higher in 2022 than pre-pandemic levels.
- Loneliness was higher in areas with higher restriction levels in 2021.
- Spending time on campus was associated with lower levels of loneliness.
- Online learning was related to higher levels of loneliness.

Loneliness is often described as a perceived deficiency in social relationships and is associated with a number of negative psychological and physical health outcomes (Hawkley & Cacioppo, 2010). There is growing recognition of loneliness as a significant public health issue with negative effects comparable to risk factors such as physical inactivity, obesity and smoking (Holt-Lunstad et al., 2017; Holt-Lunstad et al., 2010) A recent review of the prevalence of loneliness prior to the COVID-19 pandemic indicated heterogenous but at substantial levels of loneliness in many countries (Surkalim et al., 2022).

There is also evidence that loneliness is increasing in young adults. A recent meta-analysis and systematic review of 345 studies of adults aged 18-29 who completed the UCLA Loneliness scale between 1976 and 2019 found loneliness levels increased linearly each year (Buecker et al., 2021). Consistent with this pattern, Hysing and colleagues (2020) highlighted an increase in loneliness among Norwegian fulltime students from 2014 to 2018 with an overall increase in students feeling lonely from 16% to 23%. The study also found males reported the greatest increase in loneliness over time. However, it is not known if this trend continued. Based on pre-pandemic studies, the gender differences in loneliness have been inconsistent. On the one hand, two meta-analyses concluded that males had higher levels of loneliness (Maes et al., 2019; Mahon et al., 2006). On the other hand, a higher level of loneliness has been observed among women, relative to men, among young adults (Wickens et al., 2021). We have previously found that the youngest and oldest students reported the highest levels of loneliness pre-pandemic (Hysing et al., 2020), and the youngest may be at an extra risk of loneliness during the pandemic since they may not have established social networks.

The COVID-19 pandemic in 2020 gave rise to strict social restrictions and government mandated lockdowns in most countries to combat the spread of the virus. In Norway there were both national restrictions, and regional restrictions during the pandemic based on COVID rates (Han et al., 2020). For university students, a range of COVID-19 preventive measures impacted their everyday life from social distancing restrictions in the population at large to closed campuses and restrictions on time on campus and reliance on online teaching (Han et al., 2020). There were both regional differences in restriction level, but also differences in the transition from online to campus-based teaching when the restrictions were lifted. For university students, these restrictions on social activities and reliance on online education may have set the scene for an even further increase in the rate of loneliness. This is confirmed by unprecedented high levels of loneliness reported among young adults during periods of pandemic restrictions (Horigian et al., 2021; Padmanabhanunni & Pretorius, 2021; Sigfridsson & Brandt, 2021). The UK COVID-19 Social Study found young adults were at greater risk of loneliness during the pandemic, compared to pre-pandemic. Also, being a student was an increased risk for loneliness (Bu et al., 2020). Similarly, we have previously found that mental health problems were more prevalent among students in areas with a higher level of restrictions for going onto campus and greater online learning (Sivertsen et al., 2022).

The aim of the present study is to assess changes from prior to the pandemic (2018) to a period of restrictions for students during the pandemic (2021) and after most of the restrictions were lifted (2022). Further, we will assess if loneliness levels differ across key sociodemographic groups. Given the contradictory findings, we do not have a specific hypothesis regarding gender differences in loneliness. However, we hypothesize that younger students will report higher loneliness levels over time. Further, loneliness is expected to be higher in areas with high restriction levels during the pandemic and with more online and off-campus learning.

Method

Procedure

The SHoT study (Students' Health and Wellbeing Study) is a large Norwegian survey of students in higher education, conducted by three large student welfare organizations. Five surveys have been completed since 2010. This report is based on the three latest waves, conducted in 2018, 2021 and 2022. The SHoT 2018 and the SHoT 2022 were both conducted between February and April. SHoT 2021 was a briefer version focusing specifically on the COVID19 pandemic. SHoT 2021 was conducted between March and April. All full-time Norwegian students pursuing higher education were invited to participate. For SHoT 2018, SHoT 2021 and SHoT 2022, 162,512, 181,828, and 169,572 students fulfilled the inclusion criteria, of whom 50,054 (response rate: 30.8%), 62,498 (response rate 34.4%) and 59,554 (response rate: 35.1%) students completed the online questionnaires, respectively. In 2018, only students aged 18 to 35 years were included, while the 2021- and 2022-studies also included students older than 35. To enable comparisons across the three time points, the current study included students aged 18 to 35 years, yielding final sample sizes of 49,836 (2018), and 62,212 (2021), and 53,362 (2022). Detailed information of the SHoT study has been described elsewhere (Sivertsen et al., 2019).

Data Collection and Pandemic Restrictions

In Norway, the national and regional restrictions triggered by the COVID-19 pandemic changed over time. During the 2021 data collection, there were both national and regional restrictions, and there was mainly online teaching for the students and closed campuses, with some exceptions. For the 2022 data collection, there was still an ongoing pandemic, but the national and regional restrictions had lifted in Norway just before the data collection started. Still, some restrictions were in place and a hybrid of live and online teaching was offered.

Statistical Analyses

IBM SPSS Statistics 28 for Windows (SPSS Inc., Chicago, IL) was used for all statistical analyses. Pearson's chi-squared tests were used to examine changes in the prevalence of loneliness (the three T-ILS items) for male and female students separately. The magnitude of gender differences was examined using Cohen's *h*, which is measure of distance between two proportions (and interpreted similarly to Cohen's *d*). Chi-squared tests were also used to examine the association between loneliness and age group, and levels of campus closure (SHoT 2021) and online lectures (SHoT 2022). Geographical differences in loneliness (T-ILS) in the SHoT 2021 were examined by computing Estimated Marginal Means, means adjusting for sociodemographic factors (age, sex, relationship status and ethnicity), and COVID-19 factors (# of tests, positive test, having been in quarantine). There was generally very little missing data on the included variables across all three waves, and the missing values were handled using listwise deletion.

Ethics

All procedures involving human subjects/patients were approved by the Regional Committee for Medical and Health Research Ethics in Western Norway (SHoT 2018: no. 2017/1176, SHoT 2021: no. 176205, and SHoT 2022: no. 326437, respectively). Electronic informed consent was obtained after complete description of the study to the participants. Following completion of the surveys, the participants had received detailed information about the findings.

Patient and Public Involvement

The planning and design of all three SHoT studies were initiated and governed by the three largest student welfare organizations, which included deciding inclusion and exclusion criteria, and selecting potential research questions and instruments. Students were not involved in the actual collection of data, although recruitment was conducted in close collaboration with all the student welfare organizations in Norway.

Instruments

Demographic and COVID-19-Related Information

In all three SHoT studies, the students provided data on their age, gender, relationship status (single versus married/partner/boyfriend/girlfriend) and the education attained by their parents. Indication of gender had three response options: “woman,” “man” and “other”. Ethnicity was coded as Norwegian if the student or his/her parents were born in Norway, and “other” for all other countries. Based on the geographical location of each educational institution, students were categorized according to Norway’s recent county reform, which now includes 10 counties.

In the SHoT 2021 study, all students were also asked how many days they had physically spent on campus during the last 14 days, due to COVID-19 restrictions. In 2022, respondents were asked how much of the teaching had been online since fall, 2021. They also reported if they had been tested for COVID-19, number of tests, positive test (confirmed by an established test), and whether they had been in quarantine (which typically entails 10 days of staying at home/avoiding social contact).

Loneliness

In all three SHoT studies, loneliness was assessed using an abbreviated version of the widely used UCLA Loneliness Scale, the “Three-Item Loneliness Scale (T-ILS)” (Hughes et al., 2004). The T-ILS items (lack of companionship, feeling left out, and isolation) were each rated along a 5-point scale (“never”, “seldom”, “sometimes”, “often”, and “very often”). The T-ILS has displayed satisfactory reliability and both concurrent and discriminant validity (Hughes et al., 2004). More information about loneliness in the SHoT study has been published elsewhere (Hysing et al., 2020). In addition, the SHoT 2022 study also included a single item assessing to what extent the student felt s/he had enough friends at their campus, with the response options “I have many friends”, “I have some friends”, “I have few friends”, and “I have no friends”. The Cronbach’s alphas of the T-ILS were 0.87 (2022), 0.84 (2021), and 0.88 (2018).

Results

Sample Characteristics

As detailed in Table 1, female students comprised approximately 2/3 of the participants in all surveys. This differs a little from the gender distribution in higher education in Norway (around 60% women). The age range is similar across studies (18-35) and the mean age was 23.1 in 2018, 24,1 in 2021 and 24,0 in 2022. About half of the participants in all three samples reported being single. Ethnicity across the three SHoT samples was also relatively stable, with 8-10% percent being immigrants, defined as either the student or their parents being born outside Norway.

Table 1

Sociodemographic and Clinical Characteristics of the Three SHoT Studies

Characteristics	**Men**		**Women**		**Total**	
SHoT 2018						
Age, mean (*SD*)	23.4	(3.0)	23.0	(3.0)	23.1	(3.0)
Gender, % (*n*)	30.9%	(15,399)	69.1%	(34,437)		
Single, % (*n*)	56.2%	(8617)	47.5%	(16,238)	49.9%	(24,855)
Ethnicity, % (*n*)						
Norwegian	91.8%	(14,137)	92.1%	(31,711)	92.0%	(45,848)
Non-Norwegian	8.2%	(1262)	7.9%	(2726)	8.0%	(3988)
T-ILS score, *M* (*SD*)	7.13	(3.06)	7.66	(3.05)	7.50	(3.06)
SHoT 2021						
Age, *M* (*SD*)	24.3	(5.0)	24.1	(5.2)	24.1	(5.2)
Gender, % (*n*)	34.2%	(21,405)	65.6%	(40,807)		
Single, % (*n*)	55.1%	(11,777)	48.5%	(19,756)	50,8%	(31,533)
Ethnicity, % (*n*)						
Norwegian	91,3%	(19,542)	91.4%	(37,305)	91.4%	(56,847)
Non-Norwegian	8.7%	(1,863)	8.6%	(3,502)	8.6%	(5,365)
COVID-19 positive	3.1%	(622)	2.8%	(1091)	2.9%	(1703)
T-ILS score, *M* (*SD*)	8.64	2.98)	9.41	(2.87)	9.15	(2.93)
SHoT 2022						
Age, *M* (*SD*)	24.3	(3.3)	23.8	(3.2)	24.0	(3.2)
Gender, % (*n*)	33.6%	(17,939)	66.4%	(35,423)		
Single, % (*n*)	44.7%	(8023)	51.2%	(18,142)	49.0%	(26,165)
Ethnicity, % (*n*)						
Norwegian	89.6%	(16,080)	89.6%	(31,741)	89.6%	(47,821)
Non-Norwegian	10.4%	(1859)	10.4%	(3682)	10.4%	(5541)
COVID-19 positive	48.6%	(9636)	47.9%	(18905)	48.1%	(28,541)
T-ILS score, *M* (*SD*)	7.60	3.05	8.23	2.95	8.02	3.00

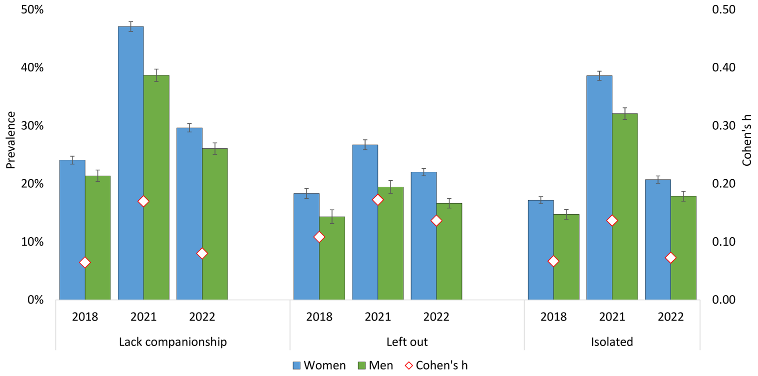
At the time of the SHoT 2021 data collection, 2.4% of the sample had tested positive for COVID-19, while 48.1% reported having tested positive by the time of SHoT 2022.

Changes in Loneliness From 2018 to 2022

There was a sharp increase in loneliness across all three T-ILS items from 2018 to 2021 (see Figure 1 for details). And while the prevalence of loneliness decreased from 2021 to 2022, the levels of loneliness were still higher in 2022 than before the pandemic in 2018. For example, 47.1% of female students reported “often” or “very often” lacking companionship during the pandemic in 2021, while the corresponding estimates before the pandemic (2018) and after pandemic restrictions were lifted (2022) was 24.1% and 29.6%, respectively. This trend was similar for male students too, but as detailed in Figure 1 (red diamonds indicating Cohen’s *h*), the gender differences showed that females reported more loneliness in 2021, compared to both 2018 and 2022.

Figure 1

Trend in Loneliness From 2018 to 2022 Among Female and Male Students in the SHoT Study



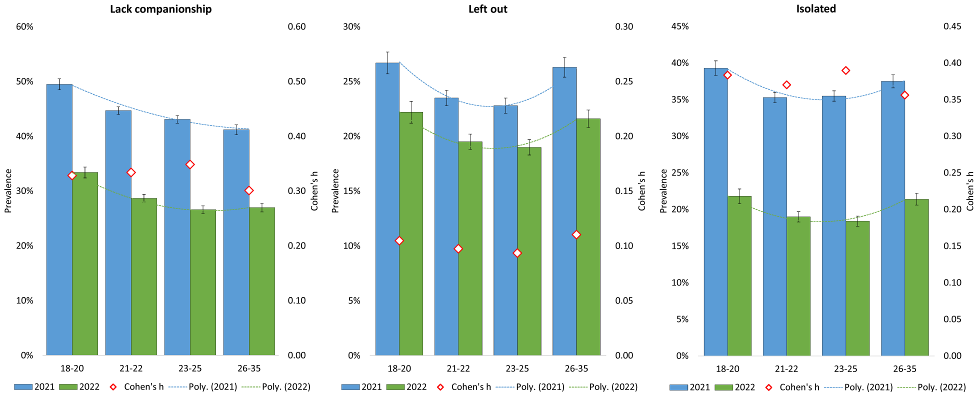
Note. Red diamonds represent gender differences expressed as Cohen’s *h*.

Age Differences in Loneliness

Figure 2 shows the prevalence of the three loneliness items across the different age groups in the SHoT 2021 and SHoT 2022 studies. As indicated by the dotted trend lines, there was a significant curvilinear relationship (all *p*s < .001) on feeling left out and isolated; both the youngest and oldest age-groups reported higher levels of feeling left out and feeling isolated (see Figure 2 for details). For the item on lacking companionship, the trend was more linear; the younger the student – the more they lacked companionship. The magnitude of differences between 2021 and 2022 was largest for feeling isolated and lacking companionship, with Cohen’s *h* effect sizes of around 0.4 and 0.3, respectively.

Figure 2

Loneliness and Age Group (in Men and Women Combined) in the SHoT 2021 and SHoT 2022 Studies



Note. Red diamonds represent differences between 2021 and 2022 expressed as Cohen's *h* (with 95% confidence intervals).

Geographical Differences in 2021

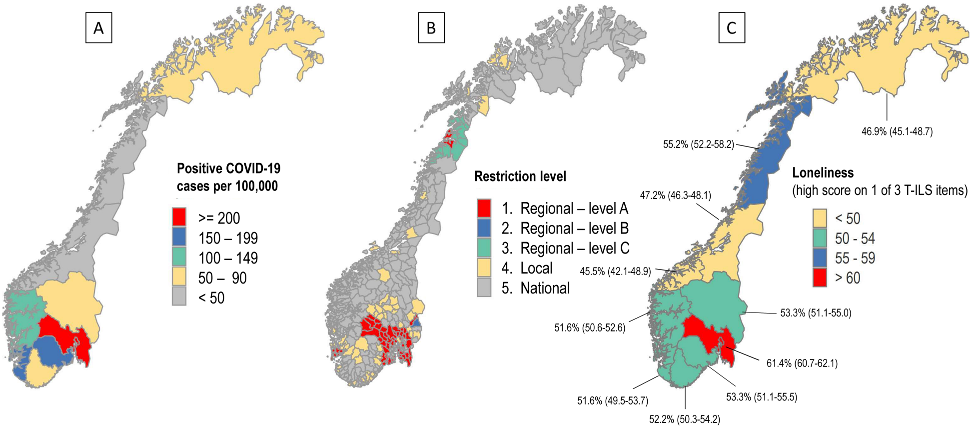
There were large geographical differences in COVID-19 cases, as displayed in Figure 3; Panel A. In March 2021, the South-Eastern region surrounding the capital of Oslo and parts of Northern Norway had substantially more COVID-19 cases compared with other areas in Norway. As displayed in Figure 3; Panel B, there were also large geographical variations in terms of imposed COVID-19-related restrictions in March 2021. As expected, the strictest measures (marked in red) followed the same geographical distribution as the COVID-19 cases. Although the SHoT waves in 2018 and 2022 found no geographical differences in loneliness (data not shown), the 2021 survey revealed significant geographical differences in adjusted levels of loneliness during the data collection in March 2021. As displayed in Figure 3; Panel C, students studying at an institution in the South-Eastern region (marked in red) and parts of Northern Norway (marked in orange), reported significantly more loneliness compared with other geographical regions, after adjusting for sociodemographic-related and COVID-19-related factors.

Loneliness and Campus Closure in 2021

Figure 4 displays the association between loneliness and campus closure in the SHoT 2021 study. There was a significant negative dose–response association between all three T-ILS items and days spent on campus. Students spending 7+ days on campus during the last 2 weeks, reported significantly less loneliness during this period, compared with students who were not permitted on campus, after adjusting for sociodemographic and COVID-19-related factors. The trend was similar for both male and female students.

Figure 3

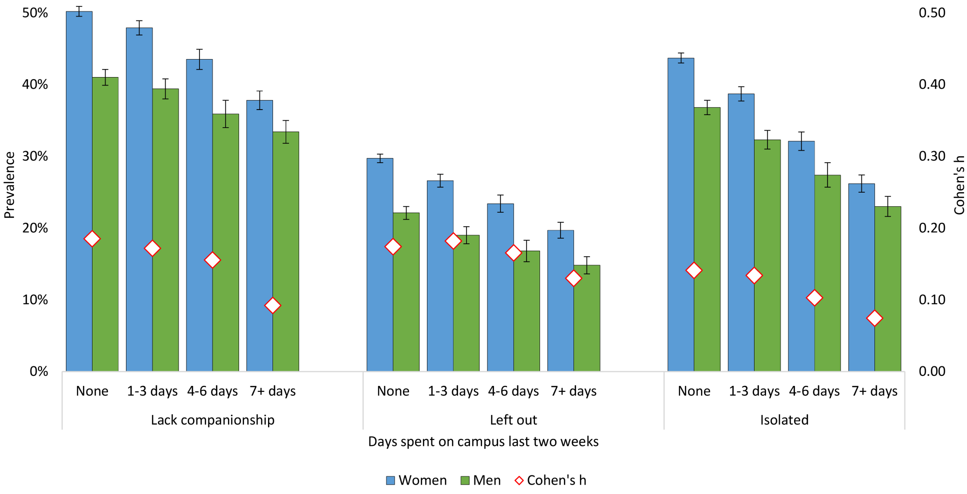
Geographical Differences in Number of Positive COVID-19 Cases (Panel A), COVID-Related Restrictions (Panel B) and Loneliness Prevalence (With 95% Confidence Intervals) in the SHoT 2021 Study (T-ILS; Panel C)



Note. Data for all three figures are based on the situation in March (only) 2021. Sources: A–B: The Norwegian Institute of Public Health. $ Estimated loneliness prevalence (any of the three T-ILS items “often” or “very often”), adjusting for sociodemographic and COVID-19 factors (# of tests, positive test, quarantine).

Figure 4

Loneliness by Campus Closure Due to COVID-19 in the SHoT 2021 Study



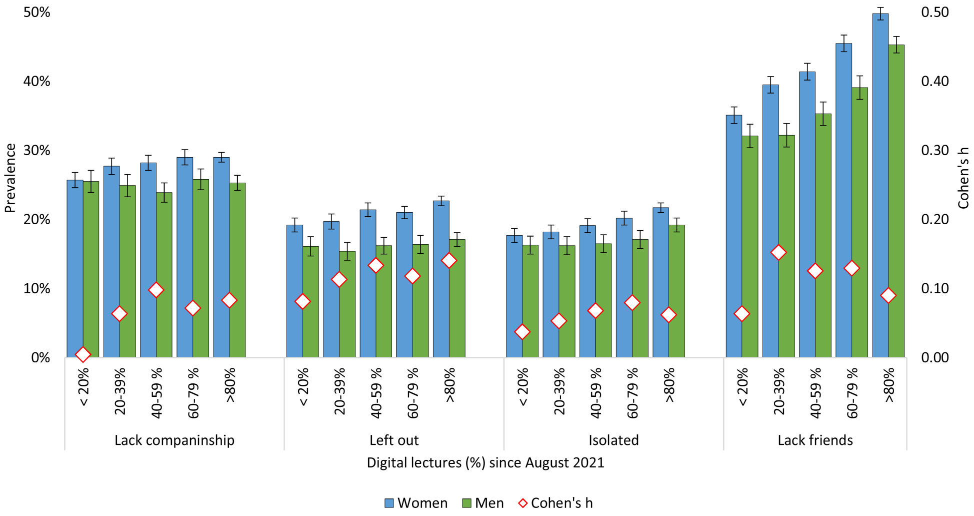
Note. Red diamonds represent gender differences expressed as Cohen’s *h*.

Loneliness and Remote Learning in 2022

As displayed in Figure 5, there was also a significant dose–response association between all loneliness items and the use of online lectures in 2022. Students who had their physical classes replaced by online lectures in 80-100% of the time since August 2021, reported significantly more loneliness compared to students who had more in person teaching. This graded association was present for all T-ILS items, but was especially strong for the item assessing to what extent students lacked friends at their place of study. For example, among female students who had predominantly remote learning, 49.8% reported having “no” or “few” friends, compared to 35.1% among those who had less than 20% of online lectures. The trend was similar for male students.

Figure 5

Loneliness and Lack of Friends by Degree of Digital Lectures in the SHoT 2022 Study



Note. Red diamonds represent gender differences expressed as Cohen’s *h*.

Discussion

The study showed a significant increase in loneliness during 2021 compared to the 2018 pre-pandemic SHoT survey. The level of loneliness was highest in regions with high COVID-19 associated restrictions in 2021 and among those students who did more remote learning, as well as among the youngest students. The 2022 SHoT study showed loneliness reduced significantly from 2021 but was still higher relative to pre-pandemic levels. Females consistently showed higher levels of loneliness, relative to men, with these gender differences increasing during the pandemic. Prior research has suggested that higher levels of loneliness in females may be due to a greater sensitivity of females to interpersonal relationships starting at adolescence (Maes et al., 2019). Further, the increased rate of mental health problems for women during the pandemic was partly explained by loneliness, underscoring the adverse consequences of loneliness (Dotsikas et al., 2023)

Together, these results confirm the higher levels of loneliness during the COVID-19 pandemic experienced by young adults that have reported in previous studies (Horigian et al., 2021; Padmanabhanunni & Pretorius, 2021; Sigfridsson & Brandt, 2021). Similar levels have also been found in the general population during the pandemic (Ernst et al., 2022). The increase in loneliness, with a twofold increase in some items that comprise the loneliness measure, confirms young adults are at high-risk group for loneliness. The rise in off campus online lectures seems to have been particularly difficult for female students, who reported feeling more socially isolated and lonely than students who were less affected by campus restrictions.

The results confirm a trend of loneliness as an increasing public health concern (Lim et al., 2020), given that the level of loneliness has shown a gradual increase from 2010 until 2022, in addition to the time limited peak during the pandemic (Hysing et al., 2020). Still, a meta-analysis has shown lower rates of loneliness among northern European countries compared to other geographical regions, and thus this may indicate that loneliness rates are even higher in other countries (Surkalim et al., 2022). The high rate of loneliness is especially worrisome given it is an established risk factor for both mental and physical health problems in this age group (Christiansen et al., 2021). Consistent with this finding, the increase in depression symptoms observed in a study of young adults may be due to this the rise in loneliness (Horigian et al., 2021). Although beyond the scope of the present study, future studies should investigate how loneliness is associated with later mental and physical health in young adults.

The unprecedented high levels of loneliness among young adults in higher education seem to be driven largely by the restriction levels which impacted on the formation of normal friendship patterns. This is in line with previous studies which have found higher levels of loneliness among students during lockdown periods in comparison to times with less restrictions (Macalli et al., 2022) The regional differences in loneliness related to the impact may be accounted for by a range of restrictions, both restrictions directly related to being a student, such as campus lock downs and online teaching, but also on more general restrictions on social contact. Still, the dose-response associations between days on campus in 2021, and similarly to online teaching in 2022, raises the potential importance of live face-to-face instruction on student's loneliness. This may be especially important to establish social relationships in the class and student group, which is indicated by the strong association between the proportion of online teaching and having friends at the study site.

The strength of the present study is that the surveys are similar in inclusion and recruitment across the three data collections and have identical measures of loneliness. The results should be interpreted in light of some limitations. The attrition rate is high across the health surveys with no information about non-participants other than age and gender. We cannot exclude the possibility of selective attrition among those with health problems. Also, it is possible that students who a particularly lonely may not participate. If so, the results might have underestimated the true level of loneliness in the population. The loneliness measure is a well-validated and commonly used assessment of loneliness, but it is an indirect measure and does not ask the participants directly if they fell lonely as has been done in some previous studies (Wickens et al., 2021) Further, the data collection has been done at set time points, and more frequent assessments could have given more detailed information about stability and changes in loneliness across the pandemic and restriction levels. Regarding the reported rates of COVID-19, these are uncertain and could be an underestimation due to the lack of testing and confirmation of COVID-19.

The current results confirm the adverse public health consequences of the COVID-19 pandemic and related pandemic restrictions. In Norway, higher education was one of the domains with high levels of restrictions (Helsingen et al., 2020). When governments and health officials are making decisions regarding restrictions, the results underscore the need to consider the adverse psychological consequences of restrictions in addition to direct health impact. The study helps to identify high risk groups and predictors of loneliness that could inform policy and interventions to reduce harm in these groups.

Further, the results of the present study confirm loneliness as a major public health concern among young adults in higher education and interventions in these settings may be needed. The youngest students were at higher risk, and this indicates the importance of supporting young adults in establishing a social network during the transition to university- and college life. There are available and effective interventions to reduce loneliness, however, they have mainly been tested in high risk groups and with individual or group based approaches (Eccles & Qualter, 2021). There are still relatively few interventions to reduce loneliness among young adults (Hawkley et al., 2022). Identifying predictors of loneliness among young adults may also give insights into how we can reduce loneliness by systemic changes. For instance, finding the right balance between online teaching and physical presence for students may be areas that need to be considered both in response to future pandemic restrictions and when planning for teaching in higher education post pandemic. At present, higher learning institutions are redesigning their teaching to find the balance between in-person and digital presence, and preventing loneliness and establishing social relationships is an important aspect to consider.

Funding: SHoT 2022 was supported by the Norwegian Ministry of Education and Research (grant number: n/a) and the Norwegian Ministry of Health and Care Services (grant number: n/a).

Acknowledgments: The authors have no additional (i.e., non-financial) support to report.

Competing Interests: None of the authors have any competing interest.

References

Bu, F., Steptoe, A., & Fancourt, D. (2020). Who is lonely in lockdown? Cross-cohort analyses of predictors of loneliness before and during the COVID-19 pandemic. *Public Health, 186*, 31–34. https://doi.org/10.1016/j.puhe.2020.06.036

Buecker, S., Mund, M., Chwastek, S., Sostmann, M., & Luhmann, M. (2021). Is loneliness in emerging adults increasing over time? A preregistered cross-temporal meta-analysis and systematic review. *Psychological Bulletin, 147*(8), 787–805. https://doi.org/10.1037/bul0000332

Christiansen, J., Qualter, P., Friis, K., Pedersen, S., Lund, R., Andersen, C., Bekker-Jeppesen, M., & Lasgaard, M. (2021). Associations of loneliness and social isolation with physical and mental health among adolescents and young adults. *Perspectives in Public Health, 141*(4), 226–236. https://doi.org/10.1177/17579139211016077

Dotsikas, K., Crosby, L., McMunn, A., Osborn, D., Walters, K., & Dykxhoorn, J. (2023). The gender dimensions of mental health during the Covid-19 pandemic: A path analysis. *PLoS One, 18*(5), Article e0283514. https://doi.org/10.1371/journal.pone.0283514

Eccles, A. M., & Qualter, P. (2021). Alleviating loneliness in young people–A meta-analysis of interventions. *Child and Adolescent Mental Health, 26*(1), 17–33. https://doi.org/10.1111/camh.12389

Ernst, M., Niederer, D., Werner, A. M., Czaja, S. J., Mikton, C., Ong, A. D., Rosen, T., Brähler, E., & Beutel, M. E. (2022). Loneliness before and during the COVID-19 pandemic: A systematic review with meta-analysis. *The American Psychologist, 77*(5), 660–677. https://doi.org/10.1037/amp0001005

Han, E., Tan, M. M. J., Turk, E., Sridhar, D., Leung, G. M., Shibuya, K., Asgari, N., Oh, J., García-Basteiro, A. L., & Hanefeld, J. (2020). Lessons learnt from easing COVID-19 restrictions: An analysis of countries and regions in Asia Pacific and Europe. *Lancet, 396*(10261), 1525–1534. https://doi.org/10.1016/S0140-6736(20)32007-9

Hawkley, L. C., Buecker, S., Kaiser, T., & Luhmann, M. (2022). Loneliness from young adulthood to old age: Explaining age differences in loneliness. *International Journal of Behavioral Development, 46*(1), 39–49. https://doi.org/10.1177/0165025420971048

Hawkley, L. C., & Cacioppo, J. T. (2010). Loneliness matters: A theoretical and empirical review of consequences and mechanisms. *Annals of Behavioral Medicine, 40*(2), 218–227. https://doi.org/10.1007/s12160-010-9210-8

Helsingen, L. M., Refsum, E., Gjøstein, D. K., Løberg, M., Bretthauer, M., Kalager, M., & Emilsson, L. (2020). The COVID-19 pandemic in Norway and Sweden – Threats, trust, and impact on daily life: A comparative survey. *BMC Public Health, 20*(1), Article 1597. https://doi.org/10.1186/s12889-020-09615-3

Holt-Lunstad, J., Robles, T. F., & Sbarra, D. A. (2017). Advancing social connection as a public health priority in the United States. *The American Psychologist, 72*(6), 517–530. https://doi.org/10.1037/amp0000103

Holt-Lunstad, J., Smith, T. B., & Layton, J. B. (2010). Social relationships and mortality risk: A meta-analytic review. *PLoS Medicine, 7*(7), Article e1000316. https://doi.org/10.1371/journal.pmed.1000316

Horigian, V. E., Schmidt, R. D., & Feaster, D. J. (2021). Loneliness, mental health, and substance use among US young adults during COVID-19. *Journal of Psychoactive Drugs, 53*(1), 1–9. https://doi.org/10.1080/02791072.2020.1836435

Hughes, M. E., Waite, L. J., Hawkley, L. C., & Cacioppo, J. T. (2004). A short scale for measuring loneliness in large surveys – Results from two population-based studies. *Research on Aging, 26*(6), 655–672. https://doi.org/10.1177/0164027504268574

Hysing, M., Petrie, K. J., Bøe, T., Lønning, K. J., & Sivertsen, B. (2020). Only the lonely: A study of loneliness among university students in Norway. *Clinical Psychology in Europe, 2*(1), Article e2781. https://doi.org/10.32872/cpe.v2i1.2781

Lim, M. H., Eres, R., & Vasan, S. (2020). Understanding loneliness in the twenty-first century: An update on correlates, risk factors, and potential solutions. *Social Psychiatry and Psychiatric Epidemiology, 55*(7), 793–810. https://doi.org/10.1007/s00127-020-01889-7

Macalli, M., Kinouani, S., Texier, N., Schück, S., & Tzourio, C. (2022). Contribution of perceived loneliness to suicidal thoughts among French university students during the COVID-19 pandemic. *Scientific Reports, 12*(1), Article 16833. https://doi.org/10.1038/s41598-022-21288-z

Maes, M., Nelemans, S. A., Danneel, S., Fernandez-Castilla, B., Van den Noortgate, W., Goossens, L., & Vanhalst, J. (2019). Loneliness and social anxiety across childhood and adolescence: Multilevel meta-analyses of cross-sectional and longitudinal associations. *Developmental Psychology, 55*(7), 1548–1565. https://doi.org/10.1037/dev0000719

Mahon, N. E., Yarcheski, A., Yarcheski, T. J., Cannella, B. L., & Hanks, M. M. (2006). A meta-analytic study of predictors for loneliness during adolescence. *Nursing Research, 55*(6), 446–447. https://doi.org/10.1097/00006199-200611000-00009

Padmanabhanunni, A., & Pretorius, T. B. (2021). The unbearable loneliness of COVID-19: COVID-19-related correlates of loneliness in South Africa in young adults. *Psychiatry Research, 296*, Article 113658. https://doi.org/10.1016/j.psychres.2020.113658

Sigfridsson, A., & Brandt, H. (2021). *Loneliness among a sample of Swedish university students during the COVID-19 pandemic* [Master's thesis, Umeå University, Umeå, Sweden). https://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-183318

Sivertsen, B., Knapstad, M., Petrie, K. J., O'Connor, R., Lønning, K. J., & Hysing, M. (2022). Changes in mental health problems and suicidal behaviour in students and their associations with

COVID-19-related restrictions in Norway: A national repeated cross-sectional analysis. *BMJ Open, 12*(2), Article e057492. https://doi.org/10.1136/bmjopen-2021-057492

Sivertsen, B., Råkil, H., Munkvik, E., & Lønning, K. J. (2019). Cohort profile: The SHoT-study, a national health and well-being survey of Norwegian university students. *BMJ Open, 9*(1), Article e025200. https://doi.org/10.1136/bmjopen-2018-025200

Surkalim, D. L., Luo, M., Eres, R., Gebel, K., van Buskirk, J., Bauman, A., & Ding, D. (2022). The prevalence of loneliness across 113 countries: Systematic review and meta-analysis. *BMJ, 376*, Article e067068. https://doi.org/10.1136/bmj-2021-067068

Wickens, C. M., McDonald, A. J., Elton-Marshall, T., Wells, S., Nigatu, Y. T., Jankowicz, D., & Hamilton, H. A. (2021). Loneliness in the COVID-19 pandemic: Associations with age, gender and their interaction. *Journal of Psychiatric Research, 136*, 103–108. https://doi.org/10.1016/j.jpsychires.2021.01.047

EACLIPT

Clinical Psychology in Europe (CPE) is the official journal of the European Association of Clinical Psychology and Psychological Treatment (EACLIPT).

leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.

Clinical Psychology in Europe

PsychOpen GOLD

Research Articles

Check for updates

Worry Intervention in an Older Adult With a Persecutory Delusion: A Single Case Experimental Design

Poppy Brown [1], Anna Crabtree [2]

[1] *Oxford Institute for Clinical Psychology Training and Research, Oxford Health NHS Foundation trust and University of Oxford, Warneford Hospital, Oxford, United Kingdom.* **[2]** *Royal Holloway University of London, Egham Hill, Egham, United Kingdom.*

Clinical Psychology in Europe, 2023, Vol. 5(3), Article e11173, https://doi.org/10.32872/cpe.11173

Received: 2023-01-21 • **Accepted:** 2023-07-01 • **Published (VoR):** 2023-09-29

Handling Editor: Tania Lincoln, University of Hamburg, Hamburg, Germany

Corresponding Author: Poppy Brown, Oxford Institute for Clinical Psychology Training and Research, Oxford Health NHS Foundation trust and University of Oxford, Warneford Hospital Warneford Lane, Oxford, UK, OX3 7JX. Phone: +44 1865 225145. E-mail: poppy.brown@hmc.ox.ac.uk

Supplementary Materials: Materials [see Index of Supplementary Materials]

Abstract

Background: This report presents the single case of Jack, a 67-year-old referred to our Older Adult Community Mental Health Team (OA CMHT) for his distressing persecutory delusion and high levels of worry. Jack also reported learning difficulties and autistic traits, although neither were formally diagnosed.

Method: Ten sessions of worry intervention taken from The Feeling Safe Programme worry module were used to reduce Jack's time spent worrying and increase his engagement in meaningful activity. Weekly face-to-face sessions were held, with Jack's brother acting as a co-therapist. Adaptations to the intervention were made based on Jack's learning preferences. An AB single case experimental design was adopted to compare Jack's scores on measures of worry, paranoia and delusional conviction, and wellbeing and daily functioning before and after intervention.

Results: Results demonstrate the worry intervention improved Jack's scores on all measures to a clinically significant degree.

Conclusions: This is the first known report of applying the worry intervention to an older adult. The results show the intervention can be of considerable benefit in terms of reducing worry and paranoia, in the context of both older age and suspected neurodiversity.

Keywords

worry, paranoia, persecutory delusion, older adult, single case experimental design

Highlights

- The evidence base for treating persecutory delusions through CBT is growing, but we do not know if these treatments are acceptable to, and effective with, older adults.
- This is the first known use of the worry module from The Feeling Safe Programme with an older adult.
- Significant reductions in paranoia and worry and improvements in wellbeing and functioning were seen.
- The intervention is brief and can be adapted to the learning needs of the individual.

Overview of the Literature

A persecutory delusion is a severe form of paranoia, where an individual holds a distressing belief with high conviction (above 50% certainty) about being at risk of harm from others (Freeman, 2016). Persecutory delusions are one of the most common symptoms of psychosis and can have a severe impact on an individual's life and wellbeing. Half of those experiencing persecutory delusions report levels of psychological wellbeing in the lowest 2% of the general population (Freeman et al., 2014). Levels of anxious avoidance are often comparable to what is seen in agoraphobia, and levels of worry are comparable to generalised anxiety disorder (Freeman, Taylor, et al., 2019).

A significant number of older adults are likely to be living with psychosis, including persecutory delusions, likely due to a combination of these symptoms being persistent over many years from first onset, and an estimated 2% of individuals experiencing a first episode of psychosis after the age of 65 (excluding psychotic experiences in the context of dementia) (Mitford et al., 2010; Vasiliadis et al., 2022). Despite this, there is limited research looking at persecutory delusions in older adults and no separate NICE guidance for treating psychosis in older adults. Older adults also do not have the same access to Assertive Outreach, Crisis, and Early Intervention in Psychosis (EIP) teams as working age adults. EIP services, for example, typically only accept referrals of patients aged 15-65, despite guidance stating that these services do not exclude individuals based on their age (Royal College of Psychiatrists, 2018).

NICE guidance for treating psychosis in adults recommends oral anti-psychotic medication in conjunction with psychological intervention – family intervention or CBT. A recent meta-analysis concluded that CBT was more effective for both hallucinations and delusions when compared with any control (Turner et al., 2020), and that the evidence base for its effectiveness is robust. Furthermore, a number of randomised controlled trials tailoring aspects of CBT to treat persecutory delusions have shown very positive effects; for example, The Worry Intervention Trial (Freeman et al., 2015). Compared with standard care, this eight-week worry intervention significantly reduced levels of worry and paranoia in 150 patients with persecutory delusions, adding to evidence that worry is a causal factor in the development and persistence of persecutory delusions that can be successfully ameliorated through intervention. The Worry Intervention now forms one of six modules within the Feeling Safe Programme, a modular psychological therapy for persecutory delusions that targets mechanisms (such as worry), that are known to cause and maintain paranoia (Freeman, Emsley, et al., 2021). The Feeling Safe Programme as a whole has demonstrated effect size improvements in delusions far above any previous intervention (Cohen's $d = 1.2$ versus Cohen's $d = 0.3$ for generic CBT for psychosis; Bighelli et al., 2018; van der Gaag et al., 2014) and training on this intervention is now being delivered to clinicians across England. In the recent trial of the intervention, patients typically completed two-to-three of the six modules in total, with the worry intervention most commonly being delivered first.

Given our ageing population, the number of older adults living with psychosis is likely to increase. This could have major clinical, social, and economic implications (Mitford et al., 2010). Evidence suggests older adults typically have more positive help-seeking attitudes than younger adults, meaning their low use of mental health services may be better explained by lack of service provision and/or ageism within healthcare in some cases (Mackenzie et al., 2008). To reduce ageism, provision must be based on need and appropriateness, rather than age. It is therefore important to assess whether current treatments used with adults of working age are also acceptable to, and effective, with older adults, with consideration given to the unique difficulties that older adults may more commonly face. This may include poor physical health, bereavement, and changes in roles e.g. from carer to being cared for, given there is evidence to suggest these factors can negatively impact mental health and therefore need to be considered in older adult's formulations (Laidlaw et al., 2004, 2016). Only one older adult was recruited in the Feeling Safe Trial, and older adults were excluded entirely from the Worry Intervention Trial. This report therefore presents the first known case of applying the worry intervention to an older adult with a persecutory delusion.

Introduction to the Case

Jack (pseudonym) was referred to our OA CMHT from his GP. He was temporarily living with his brother Mo, having previously lived with his late mother and been her main carer. Jack had struggled with psychotic experiences including voice hearing and delusional thinking for many years, with his symptoms managed through anti-psychotic medication. Prior to our input, Jack had never been offered psychological therapy. During the previous year he began to experience an exacerbation of his symptoms. After an assessment with the CMHT, he was referred to Psychological Therapies while awaiting care-coordinator allocation.

Assessment in the CMHT had raised concerns about a possible Alzheimer's diagnosis due to a low score on the Montreal Cognitive Assessment. However, upon our assessment, no memory difficulties were evident nor did Jack report any recent changes in his memory, cognition, or adaptive functioning. Further cognitive assessment was therefore not carried out. In our assessment, both Jack and Mo raised the possibility of Jack being neurodivergent and having a learning difficulty, although this had not been identified in his childhood. They reported a long history of Jack struggling with social interaction long before he had any psychotic experiences, as well as difficulty with abstracting and generalizing information, both of which can be characteristic of autism (American Psychiatric Association, 2013). They also noted he had always taken considerable time to process information, often needing things to be phrased more simply and clearly. Assessment of a possible autism spectrum condition was unlikely to be possible in the absence of an accurate neurodevelopmental history and was outside the scope of the current intervention and clinical need. Moreover, his difficulties were not severe enough to warrant being treated within a specialist intellectual disability service. However, it was considered that these learning difficulties, potentially in addition to his high levels of paranoia and anxiety, may have been what led to the question of a dementia being raised. These hypotheses, including the potential for identifying cognitive change or decline, were held as part of the formulation and the intervention was adapted as required.

Jack gave informed consent to be seen by a trainee clinical psychologist, and both Mo and Jack gave consent to record sessions, and for this case to be published and included in an anonymized report written for the clinician's university. Although Jack reported often feeling unhappy, he at no point expressed any thoughts of life not being worth living or wanting to hurt himself or others. Jack reported no physical health conditions or concerns. His psychiatric medication included sertraline and risperidone 500mcg.

Assessment

The authors gathered assessment information from past clinical notes and through two sessions with Jack. At assessment Jack described severe anxiety and worry regarding being arrested, which he was certain was due to happen imminently. He spoke about some work he had done for a contractor several years ago and reported worries about being called to trial for tax evasion. He was convinced that although he was innocent of any crime, the police would be able to imprison him. He described hearing a number of nasty voices linked to these worries, including a policewoman who would say she was going to arrest Jack, put him in jail, then ensure he would be homeless and bankrupt upon release, and this woman's husband, also a police officer, who would threaten to beat him up.

Jack described himself as a 'natural born worrier'. He felt a lot of responsibility as a child after losing his father at a young age and needing to help care for Mo, his younger brother. The exacerbation of his worries and development of psychotic experiences occurred after some challenging life events, including being defrauded by an employee of his bank in 2005, for which Jack blamed himself, and experiencing bullying by a previous supervisor at work who was both physically and verbally aggressive to him. Mo provided corroboration and further detail on these incidents, which he believed had triggered Jack's current delusion. Given evidence that the content of delusions is often based on real past experiences of harm or victimization (Freeman, 2016) this seemed a plausible hypothesis.

To manage his worries Jack typically remained at home, always checking around for police if he did go out. He also avoided talking to anyone apart from Mo, for fear of people reporting him to the police. Jack struggled to sleep at night due to preoccupation with worry, often napping during the day as a result. Day-to-day, Jack spent time watching TV and sitting in the garden. Despite describing himself as a 'natural born worrier', Jack's worries did not appear to generalize to anything other than his concerns about the police.

Goals

Jack described wanting to be able to worry less and to feel safer when out and about. Although he described finding it very challenging to meet new people, he felt he would like to try a new hobby such as a woodworking course if he were able to escape his worries. We discussed whether attending a local Men's Sheds group (a community group where older men come together to share and learn new skills) could be a useful goal to set, and Jack agreed this would be suitable to work towards.

Outcome Measures

Table 1 displays the outcome measures and time points they were completed.

Design

An AB design was followed. A three-week baseline period (A) was established before and during assessment, prior to intervention. The intervention phase (B) comprised weekly CBT sessions with measures completed at the start of each session. This design allowed inferences to be made regarding the impact of therapy on specified outcomes. Jack found completing questionnaires each week challenging and the assistance he needed to complete them could take considerable session time. Therefore, only service compulsory measures and worry analogue scales were completed weekly, with other measures completed just once at the start and end of treatment.

Table 1

Outcome Measures

Construct	Measure	When completed
	Worry	
Dunn Worry Questionnaire (DWQ)	The DWQ (Freeman et al., 2020) is a ten item measure of general worry developed as an improvement to the Penn State Worry Questionnaire (Meyer et al., 1990). Scores range from zero to 40, with higher scores reflecting higher levels of worry. A score of 21 and above indicates clinically significant levels of worry.	Once at baseline (Phase A) and once at end of treatment (Phase B)
Visual analogue scales (VAS)	Two VAS were completed: How worried have you been about other people this week on a scale of 0 (not worried at all) to 10 (worried all the time)?', and 'How distressed have you been about your worries about other people this week from 0 (not worried at all) to 10 (worried all the time)?'. These scales are recommended for weekly use when using the Worry Intervention.	At the start of each intervention session (i.e. throughout Phase B)
	Paranoia	
Revised Green et al Paranoid Thoughts Scale (R-GPS)	The R-GPTS-B (Freeman, Lister, et al., 2019) comprises eight items measuring ideas of persecution and shows excellent psychometric properties. Scores range from zero to 40, with scores of above 11 reflecting clinically significant paranoia, and scores of 18-27 the likely presence of a persecutory delusion. Although the measure has not been specifically validated for use with older adults, the measure does show invariance between age groups.	Once at baseline (Phase A) and once at end of treatment (Phase B)
Visual analogue scale (VAS)	An analogue scale ranging from 0 (don't believe it at all) to 100% (believe it totally) was administered to measure conviction in Jack' persecutory delusion that he would be unjustly arrested for tax evasion. This is a scale within the Psychotic Symptoms Rating Scale (PSYRATS) that is commonly been used as an outcome measure of delusional conviction (e.g. Freeman, Lister, et al., 2019)	Once at baseline (Phase A) and once at end of treatment (Phase B)
	Wellbeing and functioning	
Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-10)	The CORE-10 (Barkham et al., 2005) measures wellbeing, functioning, problems/symptoms, and risk. The measure was initially developed for use in adult services, but has been validated for use in older adult populations (Barkham et al., 2005). Scores range from zero to 40, with higher scores depicting more severe difficulties. Presence of clinically significant symptoms (caseness) is defined as a score of 10 or above (Barkham et al., 2005).	Every week throughout baseline (Phase A) and intervention (Phase B)
Work and Social Adjustment Scale (WSAS)	The WSAS measures the impact of mental health difficulties on day-to-day functioning (Mundt et al., 2002). The scale has five items covering work, home management, social leisure, private leisure, and relationships. Scores range from 0 to 40 with higher scores indicating greater impairment. The scale demonstrates good internal consistency, reliability, convergent and criterion validity. Caseness is defined as a score of 10 or above (IAPT, 2011). This measure is used routinely in older adult services, although no known validation of the scale within older adult populations has been reported.	Every week throughout baseline (Phase A) and intervention (Phase B)

Cognitive Behavioural Formulation

A shared understanding of Jack's difficulties was built using Freeman's cognitive model of paranoia (Freeman, 2016). This uses a "vicious flower" formulation to understand why paranoia is maintained. One of the key mechanisms within this model is worry, hence why this model was chosen. Two mini cycles that were to be the focus of the intervention were discussed and drawn out together with Jack (Appendix A, Supplementary Materials), and the clinician also developed a separate more complete formulation (Appendix B, Supplementary Materials). The first mini cycle shows how Jack's worried thoughts made him feel anxious in his body, e.g., his heart would race. Consequently, he worried more, taking the anxiety to be a sign of something being wrong. The second mini cycle shows how Jack's feelings of unsafety led him to worry, in turn making him feel even more unsafe, because his worries always focused on worst case scenarios. Jack also noted his worry meant he slept badly, did limited meaningful activity, and avoided engaging with others meaning his social network was small. These factors were included in the wider formulation.

A number of other variables were discussed with Jack that were thought to contribute to his feeling unsafe that also form part of the cognitive model of paranoia and were added to the clinician's formulation. For instance, Jack's experience of hearing nasty voices understandably made him feel unsafe. Jack also described some negative beliefs about himself that the clinician considered as important developmental factors in Jack's presentation. He wondered whether he had a 'weak mind', possibly due to stigmatising cohort beliefs about psychosis held among some older adults (Farrer et al., 2008). Jack also reported often feeling different to others – not uncommon among older adults with neurodiversity and who struggle with social interaction (Hickey et al., 2018) and possibly exacerbated by his experience of bullying – a common feeling of self-vulnerability that paranoia can build upon (Freeman, 2016). Moreover, in the past year Jack had moved from being a carer for his mother for which he felt much pride, to being cared for by his younger brother. This transition in role investments, a concept within Laidlaw's formulation for older adults (Laidlaw et al., 2003), may have contributed to Jack viewing himself more negatively.

To manage his worries about being unsafe Jack used safety behaviours of avoiding places where he felt the police might be more likely to catch him and checking around for police cars. Additionally, Jack showed evidence of sometimes jumping to conclusions when considering evidence for his beliefs, often struggling to consider alternative explanations. For example, upon hearing a siren he would tend to assume that the police must be coming to arrest him, and not consider alternative explanations.
Based on this formulation, testable hypotheses were developed.

1. The worry intervention will reduce Jack's levels of worry as measured by the DWQ and weekly VAS.

2. Given worry is a maintenance factor for paranoia (Freeman et al., 2015), Jack's paranoia as measured by the R-GPTS-B and delusional conviction will also reduce.
3. Improving Jack's worry will allow him to engage in more meaningful daily activity and experience better wellbeing, evidenced by improved scores on the CORE-10 and WSAS.

Intervention

Treatment comprised ten 60–75-minute sessions face to face over three months. Generic CBT for psychosis was considered as an option initially but given Jack's high levels of distressing worry and the demonstrated effectiveness of worry intervention for reducing both worry and paranoia, it was decided a worry intervention would be tried initially. These options and the recommendation were explained to Jack in layperson terms, who agreed with and consented to the plan given reducing his worry was something he most wanted help with. The worry module of the Feeling Safe Programme was therefore followed. This intervention is typically six-eight sessions, but content was paced more slowly to account for Jack's learning preferences. Frequent feedback was elicited to ensure sessions were clear, helpful, and well-paced, and short session summaries were written as Jack found these easier to review than the full intervention module booklets.

The intervention began with worry psychoeducation and a diary to identify Jack's time spent worrying. This showed he spent up to 15 hours a day worrying and not engaged in any other form of activity. The diary also identified Jack's triggers and 'peak times' of worry. Sitting in his living room unoccupied was a clear trigger, and peak times were first thing in the morning and last thing at night. To build motivation to reduce worry, in session two we identified Jack's positive and negative beliefs about worry. Jack thought worrying helped him to organise his mind and prepare for bad things happening, but he also felt strongly that worrying made him feel distressed and anxious. Overall, therefore, Jack was strongly in favour of reducing his worrying.

The concepts of worry periods and worry postponement were therefore introduced in Session 3. These techniques aim to postpone worry until a designated time and place, allowing an outlet for worry that is time limited and controlled. Outside of worry periods the aim is to stay occupied with meaningful activity to help keep worry away. Given he spent so many hours each morning lying in bed worrying, Jack felt his worry period needed to be early in the day, else he would not be able to keep postponing his worry. He chose the location as a spare room he normally didn't use. We began a list of enjoyable activities for Jack to engage in outside of worry periods. These included a puzzle book, history podcasts, and helping Mo prepare meals. Given the importance of structure as a tool for reducing boredom and inactivity among older adults (Baumann, 2013), to do lists and timetables were co-created with Jack to support him to increase his activity.

Jack initially found the concept of worry periods difficult. Given we were aiming to reduce worry, he felt allowing himself to worry at all would make it escalate and impossible to control. We tested this meta-worry in a behavioural experiment, where Jack compared his worry on days with and without worry periods. Although his beliefs about the worry becoming uncontrollable did not come true, Jack also did not find the worry periods helped him to worry less outside of the periods. He felt he did not actually need an outlet for his worries, with activity engagement being the most helpful tool for reducing worry. Eliminating worry periods is ultimately the desired outcome by the end of a patient's recovery from worry and given Jack's worry was already improving considerably, we agreed he would continue without using worry periods.

Three sessions then focussed on new exercises for letting go of worry. These included getting active, connecting with others, and using positive imagery. Jack practised these between sessions, with one task being to visit Men In Sheds, helping Jack to try to meet one of his goals for therapy. Finally, we ended with two review sessions where a therapy blueprint was created.

Jack engaged extremely well in therapy. With Jack's agreement, Mo was present for the first six sessions so that he understood the treatment tools and tasks and could support Jack with them between sessions when required. Jack then attended sessions 7-9 alone so he could practise retaining and implementing the information without assistance, with Mo returning for the final review session.

Throughout, the clinician aimed to instil hope in Jack for successful recovery, keeping in mind the stigmatising cohort beliefs that many older adults today can hold about psychosis. Validation and empathy were given for how distressing Jack's worries were, and curiosity was shown regarding the evidence for his delusional beliefs whenever Jack raised this. These elements helped to form a strong therapeutic relationship.

Results

Figure 1 shows Jack's scores on the standardised worry and paranoia measures pre-and-post-treatment. Pre-intervention, Jack was experiencing clinically significant levels of worry (a score of 28, where above 21 discriminates clinical severity), and paranoia (a score of 19, where above 11 discriminates clinical severity). Post-intervention, Jack's worry reduced hugely to a score of just 6, and his paranoia 9, both scores falling below clinical cut-offs. Scores on analogue scales of worry (Figure 2) also showed this, and an equally large drop was seen in Jack's delusional conviction, which fell from 100% at baseline to 25% at end of intervention.

Figure 1

Pre-and-Post-Intervention Scores for Worry and Paranoia

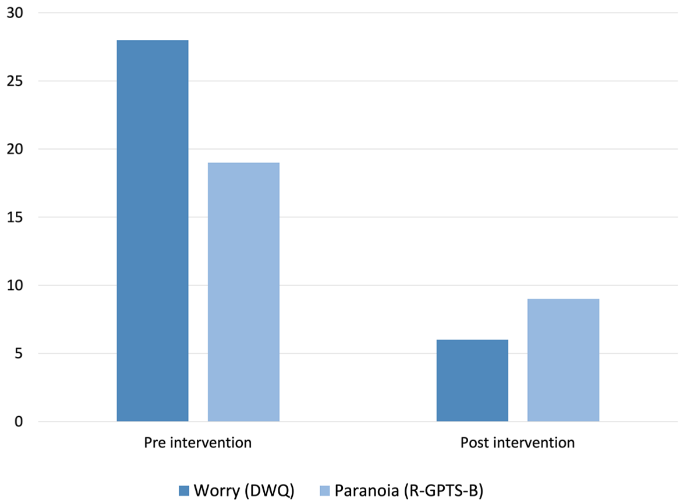


Figure 2

VAS Scores for Worry and Associated Distress During Phase B

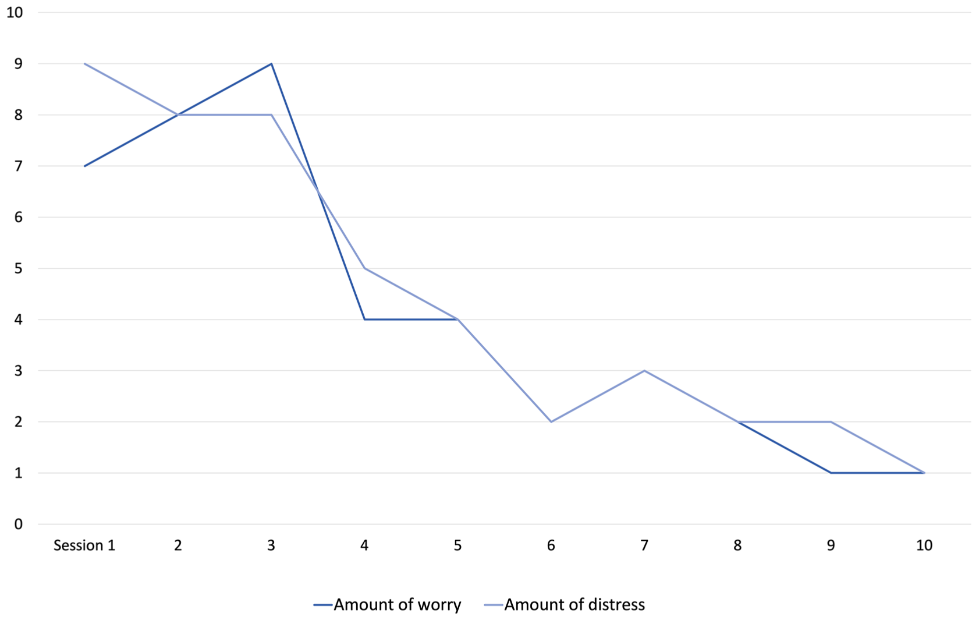
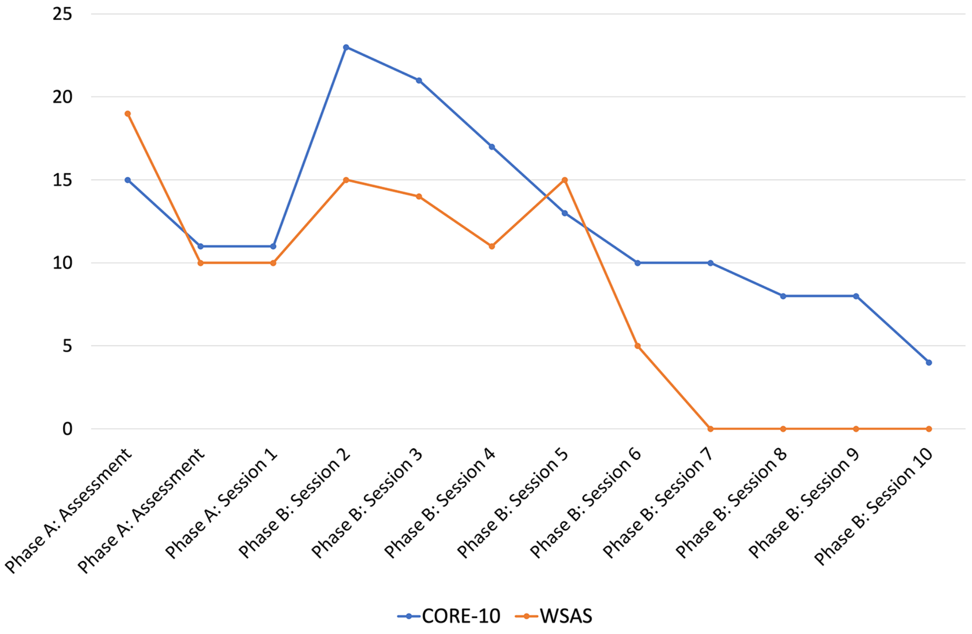


Figure 3 shows Jack's scores on the CORE-10 and WSAS measures across baseline and intervention. Jack completed these before each session began, meaning the scores for Treatment Session 1 were still part of the Phase A. At first baseline, Jack's scores on both measures were indicative of 'moderate' difficulties. Neither baseline remained entirely stable, however, with the scores reducing to the 'mild range' for the second and third baseline measures. Scores on both measures extended into the 'moderate-severe' range early on in treatment, gradually reducing until scores were very low (CORE-10) or indeed zero (WSAS) by end of treatment.

Figure 3

Changes in Wellbeing and Functioning Over Phases A and B



Discussion

This report describes the use of a worry intervention in an older adult with a persecutory delusion. All three hypotheses were supported: By the end of treatment, Jack's worry considerably reduced, as did his paranoia and delusional conviction, and his wellbeing and daily functioning improved, all to a clinically significant level. Although not formally measured, Jack also reported no longer hearing his nasty voices by the end of treatment.

Although there were large improvements overall, it was discussed in therapy why some of the measures initially increased in score (meaning a worsening of symptoms).

At assessment, Jack was clear that he was worrying a lot and wanted to reduce this, but it was not until we began a worry diary that he realised just how much time each day he spent worrying and how much of an impact this was having on his daily life and wellbeing. He therefore reflected that his earlier scores had perhaps been an understatement of his difficulties.

Conversely, there was a striking reduction in Jack's worry on the VAS after Session 3. This occurred after introducing the concepts of worry periods and worry postponement, and therefore when Jack began building more activity into his day. While this did not immediately translate into similar improvements on the wellbeing and daily functioning measures, this was perhaps due to how hard Jack found he had to work at postponing and reducing his worry. Alternatively, there may simply be a higher margin of error in Jack's CORE and WSAS scores as compared to the VAS because Jack found these measures difficult to complete. Within the WSAS Jack struggled to separate out the extent to which he had completed daily tasks, with the extent to which worry had impacted his ability to complete daily tasks. It therefore took some time for the clinician to find accurate question phrasing that allowed Jack to understand what was being asked and therefore respond accurately.

There are a number of threats to validity in this report. The baseline period was brief, and to reduce the burden on Jack the non-routine measures were only completed once at baseline. Ideally, the R-GPTS and DWQ would have been measured throughout baseline, but given Jack required support to complete them this was not possible within available session time. Additionally, the WSAS was not stable between the first baseline measure and latter two, with a smaller but still notable lack of stability also evident in the CORE. As discussed, Jack initially found these measures confusing to complete. While the intervention could have been delayed, achieving a longer, more stable baseline, this was considered unethical given there was availability to see Jack immediately. These design limitations mean caution is warranted with interpretating the results; it is possible Jack could have experienced natural recovery without the intervention. However, the extent of improvement was significant, and tallied closely with different stages of intervention, which does support the conclusion that the intervention was the primary cause of improvement.

Post-intervention Jack reported feeling proud of his achievements and confident for the future. Some concern about being arrested remained, but he acknowledged he could not be helped directly with this, and what had been most helpful was reducing his time spent worrying about it. Jack was therefore discharged from the Psychology team and CMHT. Due to limited capacity the CMHT had not been able to offer care-coordination during therapy, and Jack now felt he no longer needed any.

Overall, this report shows a brief intervention on worry led to large reductions in paranoia in an individual with a persecutory delusion. This supports the cognitive model of paranoia, where worry is a contributory causal factor in paranoia's maintenance that can be targeted therapeutically (Freeman, 2016). Moreover, the report is a first step to showing this intervention can successfully be applied with an older adult, including where there is possible neurodiversity. It will be useful to test this further, including with those in their 70s, 80s, and 90s.

Therapist's Reflections on the Case and its Clinical Implications

When discussing this case in a multi-disciplinary meeting, the therapist (PB) was met with surprise by some colleagues who had expected treatment to primarily focus on reality testing Jack's delusion and helping him consider alternative explanations. The successful results of this single case provided further clear and helpful evidence to the team on the importance of intervening on factors around an individual's delusion, rather than always focussing directly on the delusion itself.

Having Mo as a co-therapist was also extremely helpful in this case, which provided learning for our service in terms of making more use of family members. Mo was very well engaged in the sessions, and able to motivate and remind Jack of homework tasks during the weeks that he may otherwise have forgotten. As a trainee clinician only working two days a week in the service and thus limited in capacity to do check-in phone calls in between sessions, this was particularly useful. Given his longer experience of communicating with Jack, Mo was also sometimes able to rephrase questions or explanations in a way that was more understandable to Jack. However, it was clear that their relationship was very respectful, kind, and stable. Mo was careful not to speak or act on behalf of Jack, but equally Jack was able to look to Mo for support when needed. It was reflected in supervision how things might have been different had their relationship been more challenging, and how the therapist might have needed to step in more frequently to manage this, potentially sectioning of parts of session to be conducted with only Jack.

We are also aware that in some ways this case felt somewhat different to other older adult cases. Perhaps most notably, Jack was not struggling with any comorbid physical health difficulties, a variable which often needs a lot of attention in older adult work. While it would have been easy to treat Jack the same as we might treat an adult of working age, mid-way through therapy it was helpful to step back and apply a Laidlaw formulation to his case and consider the potential impact of role investments and cohort beliefs in particular, even though these were not explicitly discussed during therapy.

Funding: The research received no specific funding from any agency.

Acknowledgments: The authors offer grateful thanks to Dr Rebecca Dow and Dr Ceara Moore for their advice and comments on this single case, and particularly thankful to Jack and Mo for their time and consent to write up this report.

Competing Interests: The authors have declared that no competing interests exist.

Ethics Statement: All names and any other identifiable information have been changed in order to preserve confidentiality. The client and their co-therapist gave full, informed, consent for a report to be written regarding their therapy sessions.

Related Versions: The work was completed as part of the first author’s Doctorate in Clinical Psychology at the University of Oxford.

Twitter Accounts: @_PoppyBrown

Supplementary Materials

The Supplementary Materials contain the following items (for access see Brown & Crabtree, 2023):

- Appendix A: Two mini formulations created with Jack.
- Appendix B: An enhanced formulation created by the clinician to guide intervention but not shared with Jack.

Index of Supplementary Materials

Brown, P., & Crabtree, A. (2023). *Supplementary materials to "Worry intervention in an older adult with a persecutory delusion: A single case experimental design"* [Online appendices]. PsychOpen GOLD. https://doi.org/10.23668/psycharchives.13202

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Publishing. https://doi.org/10.1176/appi.books.9780890425596

Barkham, M., Culverwell, A., Spindler, K., & Twigg, E. (2005). The CORE-OM in an older adult population: Psychometric status, acceptability, and feasibility. *Aging & Mental Health, 9*(3), 235–245. https://doi.org/10.1080/13607860500090052

Baumann, S. L. (2013). Feeling bored: A Parse research method study with older adults. *Nursing Science Quarterly, 26*(1), 42–52. https://doi.org/10.1177/0894318412467071

Bighelli, I., Salanti, G., Huhn, M., Schneider-Thoma, J., Krause, M., Reitmeir, C., Wallis, S., Schwermann, F., Pitschel-Walz, G., Barbui, C., Furukawa, T. A., & Leucht, S. (2018).

Psychological interventions to reduce positive symptoms in schizophrenia: Systematic review and network meta-analysis. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA), 17*(3), 316–329. https://doi.org/10.1002/wps.20577

Farrer, L., Leach, L., Griffiths, K. M., Christensen, H., & Jorm, A. F. (2008). Age differences in mental health literacy. *BMC Public Health, 8*, Article 125. https://doi.org/10.1186/1471-2458-8-125

Freeman, D. (2016). Persecutory delusions: A cognitive perspective on understanding and treatment. *The Lancet: Psychiatry, 3*(7), 685–692. https://doi.org/10.1016/S2215-0366(16)00066-3

Freeman, D., Bird, J. C., Loe, B. S., Kingdon, D., Startup, H., Clark, D. M., Ehlers, A., Černis, E., Wingham, G., Evans, N., Lister, R., Pugh, K., Cordwell, J., & Dunn, G. (2020). The Dunn Worry Questionnaire and the Paranoia Worries Questionnaire: New assessments of worry. *Psychological Medicine, 50*(5), 771–780. https://doi.org/10.1017/S0033291719000588

Freeman, D., Dunn, G., Startup, H., Pugh, K., Cordwell, J., Mander, H., Černis, E., Wingham, G., Shirvell, K., & Kingdon, D. (2015). Effects of cognitive behaviour therapy for worry on persecutory delusions in patients with psychosis (WIT): A parallel, single-blind, randomised controlled trial with a mediation analysis. *The Lancet: Psychiatry, 2*(4), 305–313. https://doi.org/10.1016/S2215-0366(15)00039-5

Freeman, D., Emsley, R., Diamond, R., Collett, N., Bold, E., Chadwick, E., Isham, L., Bird, J. C., Edwards, D., Kingdon, D., Fitzpatrick, R., Kabir, T., Waite, F., Carr, L., Causier, C., Černis, E., Kirkham, M., Lambe, S., Lister, R., . . . Twivy, E. (2021). Comparison of a theoretically driven cognitive therapy (the Feeling Safe Programme) with befriending for the treatment of persistent persecutory delusions: A parallel, single-blind, randomised controlled trial. *The Lancet: Psychiatry, 8*(8), 696–707. https://doi.org/10.1016/S2215-0366(21)00158-9

Freeman, D., Lister, R., Waite, F., Yu, L. M., Slater, M., Dunn, G., & Clark, D. (2019). Automated psychological therapy using virtual reality (VR) for patients with persecutory delusions: Study protocol for a single-blind parallel-group randomised controlled trial (THRIVE). *Trials, 20*(1), Article 87. https://doi.org/10.1186/s13063-019-3198-6

Freeman, D., Startup, H., Dunn, G., Wingham, G., Černis, E., Evans, N., Lister, R., Pugh, K., Cordwell, J., & Kingdon, D. (2014). Persecutory delusions and psychological well-being. *Social Psychiatry and Psychiatric Epidemiology, 49*(7), 1045–1050. https://doi.org/10.1007/s00127-013-0803-y

Freeman, D., Taylor, K. M., Molodynski, A., & Waite, F. (2019). Treatable clinical intervention targets for patients with schizophrenia. *Schizophrenia Research, 211*, 44–50. https://doi.org/10.1016/j.schres.2019.07.016

Hickey, A., Crabtree, J., & Stott, J. (2018). 'Suddenly the first fifty years of my life made sense': Experiences of older people with autism. *Autism, 22*(3), 357–367. https://doi.org/10.1177/1362361316680914

IAPT. (2011). *The IAPT Data Handbook: Guidance on recording and monitoring outcomes to support local evidence-based practice* (Version 2.0.1.). https://webarchive.nationalarchives.gov.uk/20160302160058/http:/www.iapt.nhs.uk/silo/files/iapt-data-handbook-v2.pdf

Laidlaw, K., Kishita, N., & Chellingsworth, M. (2016). *A clinician's guide to: CBT with older people.* University of East Anglia.

Laidlaw, K., Thompson, L. W., Dick-Siskin, L., & Gallagher-Thompson, D. (2003). *Cognitive behaviour therapy with older people.* https://doi.org/10.1002/9780470713402

Laidlaw, K., Thompson, L. W., & Gallagher-Thompson, D. (2004). Comprehensive conceptualization of cognitive behaviour therapy for late life depression. *Behavioural and Cognitive Psychotherapy, 32*(4), 389–399. https://doi.org/10.1017/S1352465804001584

Mackenzie, C. S., Scott, T., Mather, A., & Sareen, J. (2008). Older adults help-seeking attitudes and treatment beliefs concerning mental health problems. *The American Journal of Geriatric Psychiatry, 16*(12), 1010–1019. https://doi.org/10.1097/JGP.0b013e31818cd3be

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy, 28*(6), 487–495. https://doi.org/10.1016/0005-7967(90)90135-6

Mitford, E., Reay, R., McCabe, K., Paxton, R., & Turkington, D. (2010). Ageism in first episode psychosis. *International Journal of Geriatric Psychiatry, 25*(11), 1112–1118. https://doi.org/10.1002/gps.2437

Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. M. (2002). The Work and Social Adjustment Scale: A simple measure of impairment in functioning. *British Journal of Psychiatry, 180*(5), 461–464. https://doi.org/10.1192/bjp.180.5.461

Royal College of Psychiatrists. (2018, June). *Standards for Early Intervention in Psychosis Services – 1st Edition* (A. Chandra, E. Patterson, & S. Hodge, Eds.). https://www.rcpsych.ac.uk

Turner, D. T., Burger, S., Smit, F., Valmaggia, L. R., & van der Gaag, M. (2020). What constitutes sufficient evidence for case formulation-driven CBT for psychosis? Cumulative meta-analysis of the effect on hallucinations and delusions. *Schizophrenia Bulletin, 46*(5), 1072–1085. https://doi.org/10.1093/schbul/sbaa045

van der Gaag, M., Valmaggia, L. R., & Smit, F. (2014). The effects of individually tailored formulation-based cognitive behavioural therapy in auditory hallucinations and delusions: A meta-analysis. *Schizophrenia Research, 156*(1), 30–37. https://doi.org/10.1016/j.schres.2014.03.016

Vasiliadis, H. M., Pitrou, I., Lamoureux-Lamarche, C., Grenier, S., Nguyen, P. V. Q., & Hudon, C. (2022). Factors associated with late-life psychosis in primary care older adults without a diagnosis of dementia. *Social Psychiatry and Psychiatric Epidemiology, 57*(3), 505–518. https://doi.org/10.1007/s00127-021-02132-7

Clinical Psychology in Europe

PsychOpen GOLD

Research Articles

Shame on Me? Love Me Tender! Inducing and Reducing Shame and Fear in Social Anxiety in an Analogous Sample

Jakob Fink-Lamotte [1,2], Jürgen Hoyer [3], Pauline Platter [2], Christian Stierle [4], Cornelia Exner [2]

[1] *Clinical Psychology, University of Potsdam, Potsdam, Germany.* **[2]** *Clinical Psychology and Psychotherapy, University of Leipzig, Leipzig, Germany.* **[3]** *Clinical Psychology and Psychotherapy, Technische Universität Dresden, Dresden, Germany.* **[4]** *Hochschule Fresenius für Wirtschaft und Medien, Hamburg, Germany.*

Clinical Psychology in Europe, 2023, Vol. 5(3), Article e7895, https://doi.org/10.32872/cpe.7895

Received: 2021-12-01 • **Accepted:** 2023-06-28 • **Published (VoR):** 2023-09-29

Handling Editor: Winfried Rief, Philipps-University of Marburg, Marburg, Germany

Corresponding Author: Jakob Fink-Lamotte, University of Potsdam, Clinical Psychology, Karl-Liebknecht-Str 24/25, 14476 Potsdam, Germany. Phone: +49-331-9772115. E-mail: jakob.fink-lamotte@uni-potsdam.de

Supplementary Materials: Data, Materials [see Index of Supplementary Materials]

Abstract

Background: Shame is considered an important factor in the development and maintenance of many psychological disorders, e.g., social anxiety disorder, and an interesting target point for therapeutic intervention.

Method: In the present experimental study, we used an online-adopted Autobiographical Emotional Memory Task (AEMT) to induce shame and tested different micro-interventions (self-compassion, cognitive reappraisal, and a control intervention) with respect to their potential to reduce shame intensity. One-hundred-and-fifteen healthy subjects participated in the study and completed a series of self-report questionnaires on self-compassion, shame, and social anxiety.

Results: The experimental shame induction was well accepted and successful (with significantly heightened feelings of shame); there were no study drop-outs. There was a significant time*condition interaction, which was due the self-compassion-based intervention resulting in a significantly larger reduction of shame than the control condition (counting fishes). In addition, the main effect of the factor experimental condition was further moderated (enhanced) by trait social anxiety and trait self-compassion.

Conclusion: The findings demonstrate the usefulness of online-adopted AEMT for the experimental induction of shame. They suggest that especially self-compassion interventions can be beneficial in alleviating intense shame experiences, which is in accordance with self-compassion theory. Overall, the results are promising in the context of experimental shame research and its potential clinical impacts call for further replication.

Keywords

social anxiety, shame, shame induction, self-compassion, reappraisal

Highlights

- We targeted shame by testing micro-interventions to reduce shame intensity.
- Shame was successfully induced using an online Autobiographical Emotional Memory Task in an experimental design.
- Self-compassion is a significant intervention to reduce shame.
- The results have promising clinical implications as well as for future research.

Shame can broadly be understood as a global devaluation of the self and is characterized by a critical, judgmental, and condemning self-verbalization (self-directed private speech; (Lewis, 1971). As shame motivates people to view themselves critically, they behave in a more reserved and detached manner in social situations. Fessler (2004) argues that the psychological function of this behavior (as a “defense mechanism”) might be to protect us from the rejection of others. As a state, shame feels like being unmasked, judged, and humiliated in a specific situation (Tangney et al., 2005), while as a trait shame comprises the tendency to experience these feelings in a variety of different (social) situations.

When a strong desire for positive reactions from others is combined with a high level of insecurity about it, people might feel exaggerated shame (Schuster et al., 2021). This ambivalence of desire for recognition and interactional insecurity leads to constant self-critical monitoring, which can be an underlying mechanism of psychological disorders. Shame associated excessive self-attention and adopting an observer perspective (self-as-object) are central factors of Clark and Wells (1995) psychopathological model of social anxiety disorder (SAD). Subsequently, exaggerated shame is thought to be a particularly important maintaining factor for SAD (Gilbert & Miles, 2000; Hedman et al., 2013), although it further plays a crucial role in the development and maintenance of a variety of psychopathological disorders e.g. depression (for review: Kim et al., 2011), eating disorders (Nechita et al., 2021), post-traumatic stress disorder (Saraiya & Lopez-Castro, 2016). In order to avoid experiencing such shame and the rejection of others, people who suffer from these disorders avoid social situations to varying extents.

SAD is not only a highly prevalent but also a highly debilitating disorder (Fehm et al., 2005; Kessler, 2003). Several studies show a significant positive correlation between shame proneness and SAD (Fergus et al., 2010; Gilbert & Miles, 2000; Hedman et al., 2013; Schuster et al., 2021; Swee et al., 2021), although research on interventions specifically focusing on reducing (or preventing) exaggerated shame in SAD is scarce. Furthermore, there is a lack of experimental studies on the modification of shame to isolate theoretically important change processes. This is supported by a review of Goffnett et al. (2020), which only includes one study investigating interventions to change shame in the context of SAD.

Nonetheless, the review also showed the promising effect of psychotherapeutic interventions with a significant reduction in shame in a post-test in 89% across a variety of contextual aspects (PTSD, body image, borderline personality disorder, etc.). Most of these studies used interventions based on cognitive-behavioral therapy (CBT) and mindfulness, while four of them applied compassion-focused interventions. Compassion-focused therapy (CFT; Gilbert, 2010) is not only a promising approach for treating SAD, for example (Blackie & Kocovski, 2018; Goldin & Gross, 2010; Koszycki et al., 2016). Self-compassion is a central construct of CFT, which can be understood as a friendly and understanding self-perspective in difficult situations characterized by an understanding that suffering is an inevitable part of human nature, while accepting it in a mindful manner (Neff, 2003b). Nonetheless, self-compassion is more than simply friendliness; rather, it is about awareness of pain that may be present and having the intention to try to alleviate it (Gilbert, 2010). Studies have shown that patients with SAD have lower self-compassion than healthy individuals (Werner et al., 2012) and an intervention based on self-compassion can effectively reduce shame (see review of Goffnett et al. (2020). This preliminary evidence suggests that CFT might be especially efficacious for the treatment of exaggerated shame, quite in accordance with the underlying theory: whereas shame is associated with a global negative devaluation of the self, self-compassion clearly counteracts this tendency as it promotes a loving relationship with the self. While shame involves a severe and judgmental emotional relationship with the self, self-compassion teaches an empathic approach.

However, as common factors in psychotherapy might mask the effects of specific interventions, head-to-head comparison studies disentangling the most effective components of psychotherapy can only be successful when based on extremely large patient samples. As Mulder et al. (2017) suggest, studies based on online-based interventions that aim at transdiagnostic processes (such as shame) are very promising. Thus, it is hypothesized that process-level variance can be more accurately elucidated by holding therapist variance constant. Further, Hofmann and Hayes (2019) suggested a paradigm switch to a process-based therapy approach where moderators and mediators of clinical change are at the center of clinical research. We would like to add the notion that experimental studies that use micro-interventions and isolate theoretically important change processes like *trait social anxiety*, *trait self-compassion* and *trait shame* could also help to transcend the common factors problem (*Do psychotherapies work primarily through the specific factors described in treatment manuals, or common factors such as therapeutic relationship, expectations, confronting problems, mastery, and attribution of the outcome*?).

To test the specific effectiveness of CFT for shame in the context of different levels of social anxiety symptoms, we compared its effects with those of another established evidence-based emotion-regulation condition, cognitive reappraisal (REAP), which is one of the best-evaluated emotion-regulation strategies (Ochsner & Gross, 2007). Gross and Thompson (2007) defined reappraisal as changing “a situation’s meaning in a way that alters its emotional impact” (p. 20). There is evidence that reappraisal can be helpful in reducing symptoms of social anxiety (for a review, see Dryman & Heimberg, 2018). This study thus aimed to test whether a self-compassion micro-intervention (COMP) is superior in reducing shame in subjects with different levels of SA symptoms compared to a REAP intervention and a control micro-intervention (CONT).

In the present study, the *Autobiographical Emotional Memory Task* (AEMT) (Mills & D’Mello, 2014; Prkachin et al., 1999) was used, as a method that has been proven to successfully induce shame (de Hooge et al., 2010; Friis et al., 2017; Houazene et al., 2021; Keng & Tan, 2017). In the AEMT, participants are instructed to remember a recent embarrassing social situation and focus on the associated emotions and feelings associated. To induce shame in an online experiment, we modified the AEMT by including more detailed audio instructions. Therefore, a further aim of the present study was to first generate data on the validity of the online version of the AEMT and subsequently to test for the shame-specificity of the AEMT. We define the manipulation check as successful when a) state shame is efficiently induced in all three micro-intervention conditions and b) the increase of state shame is more pronounced compared to state fear.

Due to theoretical assumptions on the specific effects of self-compassion for shame, we expect that COMP will reduce shame more effectively than the REAP and the CONT (H 1). We also expect that the experimental induction of shame should lead to a higher level of shame in subjects with higher rather than lower levels of social anxiety (H 2.1). We thus expect that REAP and COMP result in a stronger reduction of shame and fear compared to CONT in subjects with lower compared to higher levels of social anxiety (H 2.2: interaction of condition and anxiety group).

Material and Method

Participants

The participants were recruited using a University of Leipzig internal database. As compensation for their participation, they either took part in a lottery (five vouchers worth 10 €) or received course credit. One-hundred-and-forty-four non-clinical subjects volunteered to participate and all provided written informed consent. Respondents had to actively tick whether the following inclusion and exclusion criteria applied.

Inclusion criteria: good language skills in German, aged between 18 and 65 years, being in a quite environment and having the ability to listen to audio files.

Exclusion criteria: being pregnant, suffering from a severe mental disorder other than social phobia or a severe health impairment, or a neurological disease (e.g. traumatic brain injury, falls with unconsciousness, neurodegenerative diseases, strokes, tic disorders), psychotropic substance abuse (except coffee and nicotine) or benzodiazepine or neuroleptic medication.

Subjects could not continue the experiment if they denied presence of one of the inclusion criteria or agreed with the presence of one of the exclusion criteria

At the end of the experiment, all subjects were asked if “something unusual” happened during the experiment. Based on the responses, n = 12 subjects were excluded due to self-reported distraction, n = 2 were excluded due to self-reported technical problems, n = 2 due to unreasonably long experiment durations, n = 3 because of a more than 2 *SD* variance in trait questionnaires and n = 10 subjects were excluded because they stated that they had not carried out the micro-interventions at all, had dropped out beforehand, or had not mentioned anything at all concerning the interventions whereby it was not ensured that these subjects heard the intervention at all. In total, one-hundred-and-fifteen subjects were included in the statistical analysis, of whom n = 39 had been randomly assigned to COMP, n = 37 to REAP, and n = 39 to the CONT.

Measures

Trait Shame: Tangney´s Test of Self-Conscious Affect

The level of shame proneness was assessed using the German version (Rüsch et al., 2007) of *Tangney´s Test of Self-Conscious Affect* (TOSCA-3; Tangney et al., 2000). The TOSCA-3, presenting 11 scenes (“You have broken an object at work and then hide it.”) with four reactions (e.g. “You would think about resigning.”) rated from 1, “not likely", to 5, “very likely", has been reported to have a high internal consistency (Cronbach’s α > .77, in this study α = .66). The TOSCA-3 results in sum-scores for shame-proneness between 11 (low shame proneness) and 55 (high shame proneness).

Social Anxiety: Social Interaction Anxiety Scale

The severity of social anxiety was assessed using the German version (Stangier et al., 1999) of the *Social Interaction Anxiety Scale* (SIAS; Mattick & Clarke, 1998). The 20-item scale (e.g. “I have difficulty making eye contact with others”), rated from 0, “not applicable at all", to 4, “very much applicable", has a high internal consistency (patients with SAD; *N* = 66; α = .86; healthy controls; *N* = 50; α = .90, in this study α = .94). The sum-scores ranging from 0 to 80. In the study of Stangier et al. (1999) the social anxiety group showed a mean sum-score of 40.8 (*SD* = 16.6), the non-clinical group showed a mean sum-score of 12.5 (*SD* = 5.7).

Trait Self-Compassion: Self-Compassion Scale

The *Self-Compassion Scale* (SCS; Neff, 2003a; German version SCS-D; Hupfeld & Ruffieux, 2011) was applied to measure the trait of self-compassion. The 26-item scale (e.g. “I disapprove and condemn my own faults and weaknesses.”), rated from 1, “almost never", to 5, “almost always", has been reported to have a high internal consistency (Cronbach’s α = .91, in this study α = .89). The SCS results in mean-scores for self-compassion trait between 1 (low) and 5 (high).

State-Trait Anxiety Inventory for State Anxiety

The short version of the *State-Trait Anxiety Inventory for State Anxiety* (STAI-SKD; Englert et al., 2011) was applied to measure the level of state fear. The 5-item German translation (e.g. “I am nervous”), rated from 1, “not at all", to 5, “very much", has a high internal consistency (Cronbach’s α > .84, in this study α = .9). The STAI-SKD results in mean-scores for state anxiety between 1 (low) and 5 (high).

State Shame and Guilt Scale

The state variable shame was assessed in questionnaire format using self-assessment via the five shame items (e.g. “I feel small and insignificant”), rated from 0, “not applicable at all", to 4, “very much applicable", of the *State Shame and Guilt Scale* (SSGS; Marschall et al., 1994). An example item for shame is “I want to sink into the ground and disappear.” The SSGS results in mean-scores for state shame between 1 (low) and 5 (high) and has a high internal consistency (this study: Cronbach’s α = .93).

Experimental Design

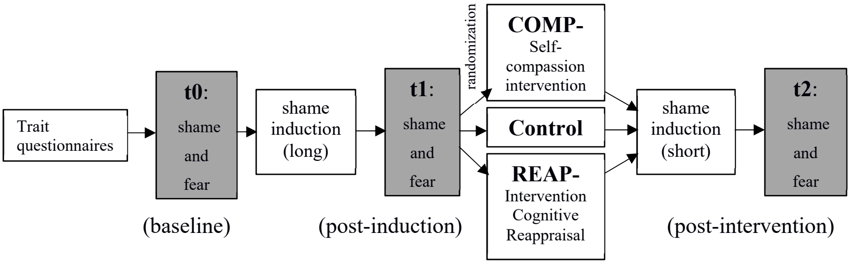
The influence of self-compassion vs. cognitive reappraisal vs. control on shame and fear was tested in an online experiment using Unipark with a mixed subject design. While the differences between the COMP, REAP, and CONT were analyzed by a between-subject design, time was assessed in a within-subject design. Therefore, three data points (baseline, t0; post-induction, t1; post-intervention, t2) were recorded for each participant. In the study documentation, we reported how we determined our sample size, all data exclusion (if any), and all manipulations and measures conducted.

Procedure

The data was collected between June and November 2020. On average, one trial lasted 42.3 minutes (*SD* = 16.14 minutes) and all experimental conditions took the same time, $F(2,109) = .104$, $p = .901$, $d = .002$. The participants were first informed of the details of the study and the test subjects’ written consent to participate was obtained. The participants then had to fill out the questionnaires listed above and state fear and shame were assessed with SSGS and STAI-SKD (t0). Thereafter, shame was induced using auditory instructions transmitted via headphones (see section "Shame Induction"). In the next step, state fear and shame were assessed again (t1) before participants received their randomization result and performed one of the three experimental conditions (COMP vs. REAP vs. CONT, see sections "Experimental Conditions"). Audio instructions were given over headphones. Then, shame was induced again, followed by the third assessment of state fear and shame (t2). Finally, questions regarding the usability and effectiveness of the manipulation ended the trial (see section "Manipulation Check"), followed by a debriefing. The schema of the experimental trial is shown in Figure 1.

Figure 1

Illustration of the Experimental Procedure



Shame Induction

Shame was induced using an auditive *Autobiographical Emotional Memory Task* (Mills & D'Mello, 2014; Prkachin et al., 1999). The participants were instructed to remember a humiliating social situation and to focus on the emotions and feelings associated with it. If they could not think of such a situation, they were given another auditive instruction (ICD-10 SAD diagnostic criteria: e.g., focusing attention towards oneself, feeling physiological exacerbation). Overall, the manipulation took six minutes. To isolate the intervention's core effect from mere time-effects, all participants received a 3-minute short version of the AEMT after the intervention for reasons of comparison (Appendix A, Supplementary Materials).

Experimental Condition: Self-Compassion Intervention (COMP)

In the self-compassion intervention (COMP), the subjects received auditory training to enhance self-compassion (Desmond, 2017; Gilbert, 2013; Neff, 2003b). In different sections, the participants were guided through an imagination exercise to increase mindfulness and acceptance, to feel human connectedness, and to build self-friendliness and wisdom. The training was adapted from a previously used intervention by Fink-Lamotte et al. (2022) to the context of social anxiety and the duration of the procedure was 8 minutes in total. A transcript of the instructions is attached in Appendix B, Supplementary Materials.

Experimental Condition: Cognitive Reappraisal Intervention (REAP)

In the cognitive reappraisal intervention (REAP), the subjects received an audio instruction to reevaluate maladaptive cognitions in social situations based on a previously used intervention by Fink et al. (2018). In the context of a guided imagination exercise, the intervention aimed to reflect on factual knowledge, decatastrophize, and strengthen self-efficacy as well as appraise an alternative and more positive and empowering perception of the social situation. The duration of the procedure was 6:49 minutes in total and a transcript of the instructions is attached in Appendix C, Supplementary Materials.

Control Condition: Counting Fishes (CONT)

In the control condition, the participants had to watch a video of an aquarium with moving fishes. They were instructed to count the number of times a yellow fish swam in and out of the picture. The duration of the procedure was 6:30 minutes in total and the experiment was adapted from Fink et al. (2018) and Fink and Exner (2019).

Manipulation Check

To check if the manipulation induced shame and/or anxiety, the participants received a four-item questionnaire. Similarly, four items assessed the subjective evaluation of the effectiveness of the instructions provided. In addition, we invited participants to describe their personal experiences and strategies during the intervention (see Appendix D, Supplementary Materials, for all materials concerning the manipulation-check). After the intervention, the participants were also asked to name specific aspects of the intervention that they perceived as helpful or hindering (Appendix E, Supplementary Materials).

Statistical Analysis

The software R (R Development Core Team, 2020) and JASP (JASP Team, 2020) were used for the statistical analysis. The statistical investigations were tested at the $\alpha = .05$ (two-tailed) level of significance. To test the manipulation check, two repeated ANOVAs investigating the effects of *condition* (COMP/REAP/CONT) and *anxiety group*, dividing the sample by a median split into low and high socially anxious groups, between t0 (before induction) and t1 (after induction), with t0 as a covariate, were calculated. These ANOVAs were run for both dependent variables *shame and fear*, specifically testing the within-subject factor *time* for induction (t0/t1) and intervention (t1/t2), and were followed by Bonferroni-corrected post-hoc tests for significant main effects and interactions. We are aware, that the Media-Split increases the probability of type I errors (Maxwell & Delaney, 1993), so we additionally ran two General Linear Mixed Model

(GLMM), with the continuous variable Trait Anxiety (sensitive analysis), which led to comparable results.

Furthermore, two ANCOVAs investigating the effects of *condition* and *anxiety group* for the difference t1 (after induction) – t2 (after intervention) as dependent variable for shame and fear, with t1 as a covariate, were calculated. The addition of t1 (t0) as a covariate is to ensure that any change observed is not artifactually due to high t1 (t0) values (regression to the mean). If the effect of the condition is significant, the ANCOVA was repeated for the pairs of conditions (each intervention is compared separately with the control) to test whether a difference between the interventions is greater than chance as indicated by the change in the control condition, which in turn would be tested by the condition term in the ANCOVA.

An overall ANOVA was not calculated, because we expected an independent induction and an independent intervention effect. The effect sizes were calculated using the R package "rstatix" (Version 0.4.0; Kassambara, 2019) whereby the adjusted partial eta-squares (η_p^2) are reported. A Shapiro Wilk Test for normality was conducted, and the normality assumption was violated for both dependent variables. However, due to the sample size, it is possible to assume an approximate asymptotic normal distribution for each of these variables (Field, 2013). A Levene test for the homogeneity of variance was conducted for the dependent variables across the time and the homogeneity of variance was not violated for shame and fear at any time point, $p > .05$. To exploratively test the effect of individual traits on shame and fear reduction, either an analysis of covariance (ANCOVA) or correlational analyses (Pearson's product-moment correlations) was calculated, depending on whether the experimental conditions differed or not.

Results

Demographic Characteristics

The three conditions were not statistically different concerning age, sex, level of social anxiety (SIAS), level of shame (TOSCA-3), level of self-compassion, or any of the other demographic or clinical data (see Table 1). The median-split resulted in a low socially anxious group ($n = 56$, sex: 51 females [91%], age = 29.86 [$SD = 11.54$]) and a high socially anxious group ($n = 59$, sex: 45 females [76%], age = 28.98 [$SD = 8.77$]). These two groups did not differ a priori concerning age, $t(113) = .456$, $p = .647$, $d = .086$, trait compassion, $t(113) = .784$, $p = .435$, $d = .146$, and trait shame, $t(113) = .292$, $p = .771$, $d = .055$, but did differ – as expected – concerning social anxiousness, $t(113) = .14.01$, $p < .001$, $d = 2.61$.

Table 1

Demographics and Clinical Characteristics by Condition

	COMP (*n* = 39)		**REAP (*n* = 37)**		**CONT (*n* = 39)**				
Characteristics	***M***	***SD* / %**	***M***	***SD* / %**	***M***	***SD* / %**	**stats**	***p***	η_p^2
Sex (% female)	33:6	86%	31:6	84%	32:7	82%	$X^2(2) = .064$	.969	
Age	28.49	8.019	31.76	12.23	28.10	9.935	*F*(2, 112) = 1.810	.168	.032
Highest Education[a]	2.821	.644	2.757	.641	2.872	.409	*F*(2, 112) = 0.382	.683	.007
Social Anxiety	33.05	21.15	28.30	16.69	26.23	14.24	*F*(2, 112) = 1.535	.220	.027
Shame	33.63	2.56	32.22	2.729	32.68	2.85	*F*(2, 112) = 2.674	.073	.046
Compassion[b]	3.29	.32	3.17	.38	3.23	.38	*F*(2, 112) = 1.036	.358	.018

Note. COMP = participants with the self-compassion intervention; REAP = participants with the cognitive reappraisal intervention; Social Anxiety (SIAS = Social-Interaction-Anxiety-Scale); Shame (TOSCA-3 = Tangney's Test of Self-Conscious Affect) Compassion (SCS = Self-Compassion Scale).

[a]Educational level was recorded in four levels matching the German school system from 1 [= highest secondary school level achieved (Abitur)] to 4 [= basic secondary school level achieved (Hauptschule)]. [b]Self-compassion level was the average of the SCS-Score without the self-criticism subscales (Neff, 2003a).

Hypothesis Testing

Manipulation Check: Induction Between t0 and t1

A repeated-measured ANOVA investigating the effects of *condition* and *anxiety group* between t0 (before induction) and t1 (after induction) for shame as dependent variable, with t0 as a covariate, shows a significant main effect of the *covariate t0*, $F(1,108) = 116.935$, $p < .001$, $\eta_p^2 = .73$, a main significant effect of *time*, $F(1,108) = 47.233$, $p < .001$, $\eta_p^2 = .304$, and a significant main effect of *anxiety group*, $F(1,108) = 4.998$, $p = .027$, $\eta_p^2 = .044$, but no significant main effect for *condition*, nor any other significant interaction, $p > .37$, $\eta_p^2 < .01$. The Bonferroni-corrected post-hoc results, controlled for the covariate, show a stronger shame experience in t1 compared to t0, $M = .767$, $t = 8.952$, $p < .001$, $d = .835$. However according to the Bonferroni-corrected post-hoc results, there was no significant stronger shame experience in the high socially anxious group, $M = .09$, $t = -.859$, $p = .392$, $d = .08$, compared to the low socially anxious group (see Figure 2a).

A repeated-measured ANOVA investigating the effects of *condition* and *anxiety group* between t0 (before induction) and t1 (after induction) for fear as dependent variable, with t0 as a covariate, shows a main significant effect of the *covariate t0*, $F(1,108) = 291.32$, $p < .001$, $\eta_p^2 = .73$, a main significant effect of *time*, $F(1,108) = 18.607$, $p < .001$, $\eta_p^2 = .147$, and a significant main effect of *anxiety group*, $F(1,108) = 4.44$, $p = .037$, $\eta_p^2 = .039$, but no significant main effect for *condition*, nor any other significant interaction, $p > .41$, $\eta_p^2 < .02$. The Bonferroni-corrected post-hoc results, controlled for the covariate, show a stronger fear experience in t1 compared to t0, $M = .51$, $t = 7.247$, $p < .001$, $d = .676$ as well as a stronger fear experience in the high socially anxious group, $M = .15$, $t = -2.39$, $p = .019$, $d = .22$, compared to the low socially anxious group (see Figure 2b).

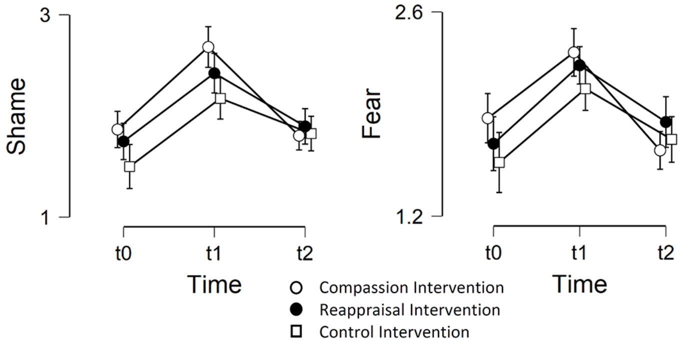
Hypothesis 1 and 2: Emotion Regulation Between t1 and t2 and the Impact of Group

An ANCOVA investigating the effects of *condition* and *anxiety group* for the difference t1 (after induction) – t2 (after intervention) as dependent variable for shame, with t1 as a covariate, shows a significant main effect of the *covariate t1*, $F(1,108) = 43.323$, $p < .001$, $\eta_p^2 = .267$, a marginally significant main effect of *condition*, $F(2,108) = 2.98$, $p = .055$, $\eta_p^2 = .037$, and a significant main effect of *anxiety group*, $F(1,108) = 4.378$, $p = .039$, $\eta_p^2 = .027$, but no significant interaction effect of *condition* and *anxiety group,* $p > .85$, $\eta_p^2 < .01$. The Tukey-corrected post-hoc results, controlled for the covariate, show a stronger shame reduction in the COMP compared to the CONT condition, $M = .395$, $t = -2.441$, $p = .043$, $d = .523$, but no significant differences between COMP and REAP, $M = .204$, $t = -1.253$, $p = .425$, $d = .24$, and REAP and COMP, $M = .191$, $t = -1.175$, $p = .471$, $d = .224$, as well as a stronger shame reduction in the high socially anxious group, $M = .311$, $t = -2.092$, $p = .039$, $d = .37$, compared to the low socially anxious group. The other post-hoc comparisons became not significant, $p > .42$, $d < .22$ (see Figure 2a).

An ANCOVA investigating the effects of *condition* and *anxiety group* for the difference t1 (after induction) – t2 (after intervention) as dependent variable for fear, with t1 as a covariate, shows a significant main effect of the *covariate t1*, $F(1,108) = 28.24$, $p < .001$, $\eta_p^2 = .19$, but no significant main effects of *condition* and *anxiety group,* nor a significant interaction effect, $p > .42$, $d < .22$ (see Figure 2b).

Figure 2

Means and Standard Error Bars of the Shame and Fear Experience Between t0 (Baseline), t1 (Post-Induction), and t2 (Post-Intervention) Across the Three Experimental Conditions



Note. Shame ratings (a) and fear ratings (b) were given on a scale between 1 and 4.

Explorative Analysis: Effects of Individual Traits on Changing Shame

Effect of Individual Traits on Shame Induction

While the factor *anxiety group* was unrelated to shame induction, we also calculated Pearson's product-moment correlations between the dependent variable *shame induction between t0 and t1* and the individual traits. The correlations between *trait social anxiety*, *trait self-compassion*, *trait shame* and the shame induction were all insignificant (all r between .07 and -.07, all $p > .45$).

Effect of Individual Traits on Shame Reduction

To examine the association between the trait variables and shame *reduction* in more detail, an ANCOVA investigating the effect of the three *conditions* for the difference t1 – t2 for shame, with t1, *trait social anxiety, trait self-compassion* and *trait shame* as a covariates, was calculated. The covariates *trait social anxiety*, $F(1, 108) = 6.56$, $p = .012$, $\eta_p^2 = .038$, and *trait self-compassion* are significantly related to experimental *condition*, $F(2, 108) = 1.874$, $p = .047$, $\eta_p^2 = .023$, while *trait shame* is not a significant covariate ($p > .29$). Controlling for the effect of *trait social anxiety* and *trait self-compassion*, the significant main effect of *condition* on *shame reduction between t1 and t2* becomes significant, $F(2, 108) = 3.854$, $p = .024$, $\eta_p^2 = .045$.

Discussion

The aim of the present study was a) to test an online-adapted method for inducing shame and b) to pilot-test two self-help interventions against heightened shame experiences. The results of this study show that the shame-based *Autobiographical Emotional Memory Task* could successfully induce both shame and fear. Furthermore, the results show that a micro-intervention based on self-compassion can reduce shame significantly better than the control condition although this effect could only be shown for shame and not for fear. An exploratory analysis also showed that trait social anxiety and trait self-compassion moderated this effect. Almost across all measurement time points, more shame and more fear were reported in the high socially anxious group compared to the low socially anxious group.

Confirming the first part of the manipulation check, the *Autobiographical Emotional Memory Task* used in this study was successful in inducing shame and fear and, accordingly, should be further applied in future experimental studies. Contrary to the second part of the manipulation check, shame and fear were induced with a similar intensity, when comparing the effect sizes. Accordingly, this induction procedure cannot be labeled as being shame-specific, which at least in part might be due to fear and shame being overlapping and highly correlated emotional states (Gilbert et al., 1994). However, to further validate the induction procedure, the introduction of a divergent variable, e.g., an emotion such as joy, is clearly recommendable. Furthermore, an induction task which is known to elicit feelings of shame even more precisely would certainly be desirable. In future research, of course, experimenters should further take care for applying the induction with ethical sensibility, as AEMT could lead to increased stress especially in samples with vulnerable individuals.

Even though there is a main effect of social anxiety group, which underlines the link between shame and social phobic symptoms (Fergus et al., 2010; Gilbert & Miles, 2000; Lutwak & Ferrari, 1997), the post-hoc effect did not become significant. Further, the results did not confirm the interaction effect hypothesized in H 2.1, which implies that the induction of shame elicit higher level of shame in subjects with higher compared to lower levels of social anxiety. This might in part be explainable by the non-clinical nature of the sample (with limited variance in social anxiety severity), but it seems more likely that there was a ceiling effect in the socially anxious group, with their initially higher shame experience scoring leaving virtually no room for further increase in shame experiences on the Likert scale. In the future, it might be useful to develop an empirical valence scale for shame (Lishner et al., 2008) that takes such ceiling effects into account.

In addition to exploring induction methods, the focus of the present study was on interventions to change exaggerated shame. Here, in line with H 1, the results show that the micro-intervention based on self-compassion (COMP) reduced shame with a medium effect size and significantly more strongly compared to a control condition (CONT). This finding supports previous ones showing that interventions based on self-compassion can be helpful in regulating shame (Cândea & Szentagotai-Tătar, 2018). This is particularly noteworthy because the control condition was an active distraction task that was also capable of producing an emotion-regulating effect. Moreover, the preliminary results show that trait social anxiety and trait self-compassion moderates this effect, and that these traits could thus influence the effectiveness of self-compassion strategies. However, this would need to be investigated in more detail in future studies. Interestingly, the stronger effect of the COMP condition specifically compared to CONT applies only to shame (and not fear) and thus does not seem to simply reflect a non-specific arousal effect. As the main effect condition became only marginally significant across all three conditions, the results need to be further verified, with the direct comparison between COMP and CONT reaching significance. Future studies should investigate whether any type of active emotion regulation conditions could be similarly effective. The results of the present study however suggest that COMP has benefits compared to active avoidance.

Not confirming to H 1, the comparison with the other active regulation intervention, cognitive reappraisal (REAP), did not show a significantly superior shame reduction of the COMP condition. REAP did not show superiority in shame reduction over the CONT condition either, thereby indirectly indicating that COMP might be preferable in reducing shame. Also contrary to H 2.2, REAP and COMP resulted not in stronger reduction of shame or fear in comparison to the control condition in subjects with higher levels of social anxiety. At the same time, the more socially anxious group reported more shame reduction during the intervention. Thus, this finding supports the proposition that using self-compassion can be a successful approach to regulate shame for individuals with higher social anxiety symptoms (cf. Blackie & Kocovski, 2018; Goldin & Gross, 2010; Koszycki et al., 2016). These results are also promising, in view of the transdiagnostic significance of shame in a number of other psychopathologies e.g. depression (for review: Kim et al., 2011), eating disorders (Nechita et al., 2021), post-traumatic stress disorder (Saraiya & Lopez-Castro, 2016).

Limitations

The present study has some limitations. Firstly, as this was an online study it could not directly be observed what the subjects did during the experiment and with what level of personal involvement they participated. We tried to experimentally control this limitation from the beginning with a series of open-ended questions. By asking the questions at the end, we hoped that it could reduce the effects of social desirability. In addition, we included subjects in the study conservatively, excluding *n* = 32 subjects from the analyses. Nevertheless, the duration and the demands of the study might have introduced some unknown bias. A second limitation of this study is the non-clinical population, although previous research (Abramowitz et al., 2003) postulated that thoughts and behaviors in psychological disorders differ more in their quantitative rather than qualitative aspects to those observed in non-clinical individuals and that basic aspects of psychological disorders (e.g., emotion regulation) can be investigated on a continuum between non-clinical individuals and patients. Yet, the sample shows on average relatively high social anxiety scores and a relatively wide variation of scores. Both aspects are favourable for investigating our research questions. Thirdly, this was a feasibility study with a piloting character without a formal a priori power analysis, and the study was not preregistered. These aspects are of course inevitable preconditions for possible replication studies in the future. Fourth, the results might have been influenced by responder bias because self-report questionnaires were used for measuring the dependent variable. However, the response tendencies affect all conditions equally and thus should have no influence on differences between the conditions, but rather increase the “noise” of the main effect time. Even though questions concerning shame and fear were not directly posed, further less biased measures, e.g., physiological measures should be included in future studies. Fifth, it cannot be definitively ruled out that the effect is artifactual due to scale effects, even though we controlled for the initial measurement (t0/t1). By scale effects we mean that details of the information are lost due to the small range and the upper and lower limits of the scale. In future studies it would be helpful to use longer-lasting, potentially even more intense interventions and follow-up measurements. Lastly, future experiments might consider recording the broader variable “gender” instead of “sex”.

Conclusions and Implications for Further Research

The study had two aims: first, to examine the extent to which the *Autobiographical Emotional Memory Task* is a helpful approach to induce shame experimentally in an online setting, and second, to find experimental evidence for superior effects of self-compassion in reducing subclinical shame (when compared with active control conditions). The results show that shame could be effectively induced in an experimental online study with the *Autobiographical Emotional Memory Task,* but that the induction also elicited fear which is why the procedure should be further developed and validated in future studies. In addition, the results show that even a short micro-intervention of self-compassion, unlike cognitive reappraisal, was significantly more efficacious in in down-regulating shame – also in healthy individuals with higher social anxiety symptoms – than a control intervention. Additionally, the traits of social anxiety and self-compassion seem to moderate this effect. These results support the importance of self-compassion in the treatment of shame-related disorders.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgments: We would like to thank the individuals who participated in this study. The authors also thank Lena Lämmerhirdt and Martin Schmieder for their assistance in conducting the study.

Competing Interests: The authors declare that they have no conflict of interest.

Ethics Statement: This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the local ethics committee (no. 2020.02.09).

Twitter Accounts: @jaemaf

Data Availability: Data and non-copyrighted materials can be publicly accessed (Fink-Lamotte et al., 2021a, 2021b, 2023) or will be made available by the authors on reasonable request.

Supplementary Materials

The Supplementary Materials contain the following items (for access see Index of Supplementary Materials below):

- The research data for this study
- The codebook for the dataframe
- The online appendices for the article

Index of Supplementary Materials

Fink-Lamotte, J., Hoyer, J., Platter, P., Stierle, C., & Exner, C. (2021a). *Supplementary materials to "Shame on me? Love me tender! Inducing and reducing shame and fear in social anxiety in an analogous sample"* [Research data]. OSF. https://osf.io/ne48g

Fink-Lamotte, J., Hoyer, J., Platter, P., Stierle, C., & Exner, C. (2021b). *Supplementary materials to "Shame on me? Love me tender! Inducing and reducing shame and fear in social anxiety in an analogous sample"* [Codebook]. OSF. https://osf.io/gkam8

Fink-Lamotte, J., Hoyer, J., Platter, P., Stierle, C., & Exner, C. (2023). *Supplementary materials to "Shame on me? Love me tender! Inducing and reducing shame and fear in social anxiety in an analogous sample"* [Online appendices]. PsychOpen GOLD. https://doi.org/10.23668/psycharchives.13165

References

Abramowitz, J. S., Franklin, M. E., Schwartz, S. A., & Furr, J. M. (2003). Symptom presentation and outcome of cognitive-behavioral therapy for obsessive-compulsive disorder. *Journal of Consulting and Clinical Psychology, 71*(6), 1049–1057. https://doi.org/10.1037/0022-006X.71.6.1049

Blackie, R. A., & Kocovski, N. L. (2018). Examining the relationships among self-compassion, social anxiety, and post-event processing. *Psychological Reports, 121*(4), 669–689. https://doi.org/10.1177/0033294117740138

Cândea, D.-M., & Szentagotai-Tătar, A. (2018). Shame-proneness, guilt-proneness and anxiety symptoms: A meta-analysis. *Journal of Anxiety Disorders, 58*, 78–106. https://doi.org/10.1016/j.janxdis.2018.07.005

Clark, D. M., & Wells, A. (1995). A cognitive model of social phobia. In R. G. Heimberg, M. Leibowitz, D. A. Hope, & F. R. Schneider (Eds.), *Social phobia: Diagnosis, assessment and treatment* (pp. 69-93). Guilford Press.

de Hooge, I. E., Zeelenberg, M., & Breugelmans, S. M. (2010). Restore and protect motivations following shame. *Cognition and Emotion, 24*(1), 111–127. https://doi.org/10.1080/02699930802584466

Desmond, T. (2017). *The Self-Compassion Skills Workbook: A 14-day plan to transform your relationship with yourself.* W. W. Norton & Company.

Dryman, M. T., & Heimberg, R. G. (2018). Emotion regulation in social anxiety and depression: A systematic review of expressive suppression and cognitive reappraisal. *Clinical Psychology Review, 65*, 17–42. https://doi.org/10.1016/j.cpr.2018.07.004

Englert, C., Bertrams, A., & Dickhäuser, O. (2011). Entwicklung der Fünf-Item-Kurzskala STAI-SKD zur Messung von Zustandsangst. *Zeitschrift für Gesundheitspsychologie, 19*(4), 173–180. https://doi.org/10.1026/0943-8149/a000049

Fehm, L., Pelissolo, A., Furmark, T., & Wittchen, H.-U. (2005). Size and burden of social phobia in Europe. *European Neuropsychopharmacology, 15*(4), 453–462. https://doi.org/10.1016/j.euroneuro.2005.04.002

Fergus, T. A., Valentiner, D. P., McGrath, P. B., & Jencius, S. (2010). Shame- and guilt-proneness: Relationships with anxiety disorder symptoms in a clinical sample. *Journal of Anxiety Disorders, 24*(8), 811–815. https://doi.org/10.1016/j.janxdis.2010.06.002

Fessler, D. M. T. (2004). Shame in two cultures: Implications for evolutionary approaches. *Journal of Cognition and Culture, 4*(2), 207–262. https://doi.org/10.1163/1568537041725097

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4th edition). SAGE.

Fink, J., & Exner, C. (2019). Does transcranial direct current stimulation (tDCS) improve disgust regulation through imagery rescripting? *Frontiers in Human Neuroscience, 13*, Article 192. https://doi.org/10.3389/fnhum.2019.00192

Fink, J., Pflugradt, E., Stierle, C., & Exner, C. (2018). Changing disgust through imagery rescripting and cognitive reappraisal in contamination-based obsessive-compulsive disorder. *Journal of Anxiety Disorders, 54*, 36–48. https://doi.org/10.1016/j.janxdis.2018.01.002

Fink-Lamotte, J., Platter, P., Stierle, C., & Exner, C. (2022). Mechanisms and effectiveness of imagery strategies in reducing disgust in contamination-related obsessive–compulsive disorder: Comparing imagery rescripting, imagery self-compassion and mood-focused imagery. *Cognitive Therapy and Research, 46*, 747–763. https://doi.org/10.1007/s10608-021-10275-9

Friis, A. M., Johnson, M. H., & Consedine, N. S. (2017). Paradoxical effects of self-compassion on mood and teeth flossing behavior in an experimental setting. *Mindfulness, 8*(1), 150–158. https://doi.org/10.1007/s12671-016-0585-x

Gilbert, P. (2010). An introduction to compassion focused therapy in cognitive behavior therapy. *International Journal of Cognitive Therapy, 3*(2), 97–112. https://doi.org/10.1521/ijct.2010.3.2.97

Gilbert, P. (2013). *Compassion focused therapy*. Junfermann.

Gilbert, P., & Miles, J. N. V. (2000). Sensitivity to social put-down: It’s relationship to perceptions of social rank, shame, social anxiety, depression, anger and self-other blame. *Personality and Individual Differences, 29*(4), 757–774. https://doi.org/10.1016/S0191-8869(99)00230-5

Gilbert, P., Pehl, J., & Allan, S. (1994). The phenomenology of shame and guilt: An empirical investigation. *The British Journal of Medical Psychology, 67*(1), 23–36. https://doi.org/10.1111/j.2044-8341.1994.tb01768.x

Goffnett, J., Liechty, J. M., & Kidder, E. (2020). Interventions to reduce shame: A systematic review. *Journal of Behavioral and Cognitive Therapy, 30*(2), 141–160. https://doi.org/10.1016/j.jbct.2020.03.001

Goldin, P. R., & Gross, J. J. (2010). Effects of mindfulness-based stress reduction (MBSR) on emotion regulation in social anxiety disorder. *Emotion, 10*(1), 83–91. https://doi.org/10.1037/a0018441

Gross, J. J., & Thompson, R. A. (2007). Emotion regulation: Conceptual foundations. In J. J. Gross (Ed.), *Handbook of emotion regulation* (pp. 3–24). Guilford Press.

Hedman, E., Ström, P., Stünkel, A., & Mörtberg, E. (2013). Shame and guilt in social anxiety disorder: Effects of cognitive behavior therapy and association with social anxiety and depressive symptoms. *PLoS ONE, 8*(4), Article e61713. https://doi.org/10.1371/journal.pone.0061713

Hofmann, S. G., & Hayes, S. C. (2019). The future of intervention science: Process-based therapy. *Clinical Psychological Science, 7*(1), 37–50. https://doi.org/10.1177/2167702618772296

Houazene, S., Leclerc, J. B., O'Connor, K., & Aardema, F. (2021). "Shame on you": The impact of shame in body-focused repetitive behaviors and binge eating. *Behaviour Research and Therapy, 138*, Article 103804. https://doi.org/10.1016/j.brat.2021.103804

Hupfeld, J., & Ruffieux, N. (2011). Validierung einer deutschen Version der Self-Compassion Scale (SCS-D). *Zeitschrift für Klinische Psychologie und Psychotherapie, 40*(2), 115–123. https://doi.org/10.1026/1616-3443/a000088

JASP Team. (2020). *JASP* (Version 0.14.1) [Computer software].

Kassambara, A. (2019). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests* (R package Version 0.3.1) [Computer software]. https://CRAN.R-project.org/package=rstatix

Keng, S.-L., & Tan, J. X. (2017). Effects of brief mindful breathing and loving-kindness meditation on shame and social problem solving abilities among individuals with high borderline personality traits. *Behaviour Research and Therapy, 97*, 43–51. https://doi.org/10.1016/j.brat.2017.07.004

Kessler, R. C. (2003). The impairments caused by social phobia in the general population: Implications for intervention. *Acta Psychiatrica Scandinavica, 108*(s417), 19–27. https://doi.org/10.1034/j.1600-0447.108.s417.2.x

Kim, S., Thibodeau, R., & Jorgensen, R. S. (2011). Shame, guilt, and depressive symptoms: A meta-analytic review. *Psychological Bulletin, 137*(1), 68–96. https://doi.org/10.1037/a0021466

Koszycki, D., Thake, J., Mavounza, C., Daoust, J.-P., Taljaard, M., & Bradwejn, J. (2016). Preliminary investigation of a mindfulness-based intervention for social anxiety disorder that integrates compassion meditation and mindful exposure. *Journal of Alternative and Complementary Medicine, 22*(5), 363–374. https://doi.org/10.1089/acm.2015.0108

Lewis, H. B. (1971). Shame and guilt in neurosis. *Psychoanalytic Review, 58*(3), 419–438.

Lishner, D. A., Cooter, A. B., & Zald, D. H. (2008). Addressing measurement limitations in affective rating scales: Development of an empirical valence scale. *Cognition and Emotion, 22*(1), 180–192. https://doi.org/10.1080/02699930701319139

Lutwak, N., & Ferrari, J. R. (1997). Understanding shame in adults: Retrospective perceptions of parental-bonding during childhood. *The Journal of Nervous and Mental Disease, 185*(10), 595–598. https://doi.org/10.1097/00005053-199710000-00001

Marschall, D., Saftner, J., & Tangney, J. P. (1994). *State Shame and Guilt Scale (SSGS)* [Unpublished Master's thesis]. George Mason University, Fairfax, VA, USA.

Mattick, R. P., & Clarke, J. C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour Research and Therapy, 36*(4), 455–470. https://doi.org/10.1016/S0005-7967(97)10031-6

Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin, 113*(1), 181–190. https://doi.org/10.1037/0033-2909.113.1.181

Mills, C., & D'Mello, S. (2014). On the validity of the autobiographical emotional memory task for emotion induction. *PLoS ONE, 9*(4), Article e95837. https://doi.org/10.1371/journal.pone.0095837

Mulder, R., Murray, G., & Rucklidge, J. (2017). Common versus specific factors in psychotherapy: Opening the black box. *The Lancet Psychiatry, 4*(12), 953–962. https://doi.org/10.1016/S2215-0366(17)30100-1

Nechita, D.-M., Bud, S., & David, D. (2021). Shame and eating disorders symptoms: A meta-analysis. *International Journal of Eating Disorders, 54*(11), 1899–1945. https://doi.org/10.1002/eat.23583

Neff, K. D. (2003a). The development and validation of a scale to measure self-compassion. *Self and Identity, 2*(3), 223–250. https://doi.org/10.1080/15298860309027

Neff, K. D. (2003b). Self-compassion: An alternative conceptualization of a healthy attitude toward oneself. *Self and Identity, 2*(2), 85–101. https://doi.org/10.1080/15298860309032

Ochsner, K. N., & Gross, J. J. (2007). The neural architecture of emotion regulation. In J. J. Gross (Ed.), *Handbook of emotion regulation* (pp. 87–109). Guilford Press.

Prkachin, K. M., Williams-Avery, R. M., Zwaal, C., & Mills, D. E. (1999). Cardiovascular changes during induced emotion: An application of Lang's theory of emotional imagery. *Journal of Psychosomatic Research, 47*(3), 255–267. https://doi.org/10.1016/S0022-3999(99)00036-7

R Development Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Rüsch, N., Corrigan, P. W., Bohus, M., Jacob, G. A., Brueck, R., & Lieb, K. (2007). Measuring shame and guilt by self-report questionnaires: A validation study. *Psychiatry Research, 150*(3), 313–325. https://doi.org/10.1016/j.psychres.2006.04.018

Saraiya, T., & Lopez-Castro, T. (2016). Ashamed and afraid: A scoping review of the role of shame in post-traumatic stress disorder (PTSD). *Journal of Clinical Medicine, 5*(11), Article 94. https://doi.org/10.3390/jcm5110094

Schuster, P., Beutel, M. E., Hoyer, J., Leibing, E., Nolting, B., Salzer, S., Strauss, B., Wiltink, J., Steinert, C., & Leichsenring, F. (2021). The role of shame and guilt in social anxiety disorder. *Journal of Affective Disorders Reports, 6*, Article 100208. https://doi.org/10.1016/j.jadr.2021.100208

Stangier, U., Heidenreich, T., Berardi, A., Golbs, U., & Hoyer, J. (1999). Die Erfassung sozialer Phobie durch Social Interaction Anxiety Scale (SIAS) und die Social Phobia Scale (SPS) [Assessment of social phobia by the Social Interaction Anxiety Scale (SIAS) and the Social Phobia Scale (SPS)]. *Zeitschrift für Klinische Psychologie und Psychotherapie, 28*(1), 28–36. https://doi.org/10.1026//0084-5345.28.1.28

Swee, M. B., Hudson, C. C., & Heimberg, R. G. (2021). Examining the relationship between shame and social anxiety disorder: A systematic review. *Clinical Psychology Review, 90*, Article 102088. https://doi.org/10.1016/j.cpr.2021.102088

Tangney, J. P., Dearing, R. L., Wagner, P. E., & Gramzow, R. (2000). *Test of Self-Conscious Affect–3.* APA PsycTests. https://doi.org/10.1037/t06464-000

Tangney, J. P., Mashek, D., & Stuewig, J. (2005). Shame, guilt, and embarrassment: Will the real emotion please stand up? *Psychological Inquiry, 16*(1), 44–48.

Werner, K. H., Jazaieri, H., Goldin, P. R., Ziv, M., Heimberg, R. G., & Gross, J. J. (2012). Self-compassion and social anxiety disorder. *Anxiety, Stress, and Coping, 25*(5), 543–558. https://doi.org/10.1080/10615806.2011.608842

EACLIPT

Clinical Psychology in Europe (CPE) is the official journal of the European Association of Clinical Psychology and Psychological Treatment (EACLIPT).

leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.

Clinical Psychology in Europe

PsychOpen GOLD

Research Articles

Scheduled Support Versus Support on Demand in Internet-Delivered Cognitive Behavioral Therapy for Social Anxiety Disorder: Randomized Controlled Trial

Anton Käll[1,2], Cecilia Olsson Lynch[3], Kajsa Sundling[3], Tomas Furmark[3], Per Carlbring[4], Gerhard Andersson[1,2,5]

[1] *Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden.* **[2]** *Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden.* **[3]** *Department of Psychology, Uppsala University, Uppsala, Sweden.* **[4]** *Department of Psychology, Stockholm University, Stockholm, Sweden.* **[5]** *Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden.*

Clinical Psychology in Europe, 2023, Vol. 5(3), Article e11379, https://doi.org/10.32872/cpe.11379

Received: 2023-02-14 • **Accepted:** 2023-08-01 • **Published (VoR):** 2023-09-29

Handling Editor: Cornelia Weise, Philipps-University of Marburg, Marburg, Germany

Corresponding Author: Gerhard Andersson, Department of Behavioural Sciences and Learning, Linköping University, SE-581 83 Linköping, Sweden. E-mail: gerhard.andersson@liu.se

Abstract

Objectives: Clinician-supported internet-delivered cognitive behavioral therapy (ICBT) can be an effective treatment option when treating social anxiety disorder (SAD). Unguided ICBT is often found to be less effective. One possible solution to reduce the costs of clinician support is to provide support on demand. In this format of guidance, participants have the option to contact their clinician if needed. In a few studies, this mode of support has been compared favorably to scheduled support.

Method: Participants in a previously reported controlled trial on SAD who had been in a waitlist control group were randomly allocated to ICBT with either on-demand guidance or scheduled weekly therapist guidance. A total of 99 participants were included. Data were collected weekly on the primary outcome measure, the Liebowitz Social Anxiety Scale self-report (LSAS-SR), and at pre- and post-treatment for secondary measures. Data were analyzed in accordance with the intention-to-treat principle using mixed-effects models.

Results: Both groups improved significantly during the treatment according to the LSAS-SR ratings. The groups did not differ in their estimated change during the treatment period, with a between-group effect of $d = 0.02$, 95% CI [-0.37, 0.43]. Both groups experienced similar improvement also on the secondary outcome measures, with small between-group effect sizes on all outcomes.

Conclusions: The findings indicate that support on demand can be an effective way of providing guidance in ICBT for SAD, although more research on this topic is needed. A limitation of the study is that it was conducted in 2009, and the findings were in the file drawer. Subsequent published studies support our initial findings, but more research is needed.

Keywords

social anxiety disorder, ICBT, Internet-delivered treatments, guided ICBT

Highlights

- One way to handle the need for therapist support in internet-delivered cognitive behaviour therapy (ICBT) is to offer support on demand.
- A randomized controlled trial was conducted comparing scheduled versus support on demand when completing ICBT for social anxiety disorder (SAD).
- Both groups improved and there were no major differences in outcome.
- Support on demand can be an effective way of providing guidance in ICBT for SAD.

Social anxiety disorder (SAD) is a common and debilitating mental health problem characterized by a persistent and intense fear of being evaluated in social situations (American Psychiatric Association, 2013). Global estimates suggest that SAD has an average lifetime prevalence of around 4%, often coupled with an early onset (Stein et al., 2017) and, when left untreated, a chronic course (Steinert et al., 2013).

Psychological treatments have been shown to assist people with this problem (Acarturk et al., 2009). Cognitive behavioral therapy (CBT) is often seen as the gold standard among these treatments, producing large effect sizes (Mayo-Wilson et al., 2014) and lasting effects that are maintained years after therapy termination (van Dis et al., 2020). Additionally, CBT targeting SAD has been disseminated successfully using modes other than traditional individual therapy, for example, in group settings (Barkowski et al., 2016) and via the Internet (Guo et al., 2021), most commonly in the form of internet-delivered CBT (ICBT; Andersson, 2018). ICBT provides a resource-effective way of delivering psychological treatment, as it requires less time from the therapist and can increase access to CBT in underserved areas and populations (Andersson, 2016). It has also been shown to be a cost-effective option (Donker et al., 2015). Specifically for SAD, ICBT has been shown to be an effective option in a regular care setting (El Alaoui et al., 2015) and to have lasting effects five years after termination (Hedman et al., 2011).

ICBT is often administered with scheduled support from a therapist (Andersson, 2016), and studies suggest that this is more effective than pure self-help versions of ICBT (Baumeister et al., 2014). However, there are exceptions. For example, one study conducted in China reported that a pure self-help condition produced comparable results to a condition which received regular therapist guidance (Kishimoto et al., 2016). Furmark et al. (2009) also found that a bibliotherapy condition with minimal therapist contact led to similar improvement compared to a therapist-supported ICBT condition, and that both active conditions outperformed a waitlist control group (Furmark et al., 2009).

One alternative to providing scheduled clinical support in ICBT is to provide support on demand (also referred to as optional support). This requires clients to contact their clinician when they want feedback, support, or have questions regarding the treatment material. This resembles helplines and usually requires less clinician time. Support on demand has been found to generate similar results to guided ICBT interventions in the treatment of anxiety and depression in a routine care setting (Hadjistavropoulos et al., 2017; Hadjistavropoulos et al., 2019). Additionally, the results indicated no significant differences in satisfaction with the treatment. Dear et al. (2015) did not find any significant differences between the optional support condition and scheduled therapist support in a trial on chronic pain, with high satisfaction and completion ratings across conditions (Dear et al., 2015). Support on demand has also been shown to have similar long-term outcomes compared to scheduled guidance in a study examining the outcomes of ICBT for loneliness two years after treatment (Käll et al., 2020). In a small factorial design trial on generalized anxiety disorder, the authors reported that support on demand was as effective as scheduled support, but that scheduled support was rated as more positive (Dahlin et al., 2022). Also, it has been suggested that scheduled guidance compared to optional guidance, is slightly more favorable at least in terms of adherence (Koelen et al., 2022). In conclusion, controlled trials on clinician support on demand provides initial support for this guidance format. This way of disseminating ICBT could increase access to ICBT and reduce costs for the support function while still not sacrificing effects and safety.

Given the increasing interest in ICBT and the need to make ICBT scalable, the aim of the current study was to compare the support on demand mode with scheduled support in ICBT treatment for SAD. Here we report findings from an unpublished part, i.e. a waiting list control group, of a previous randomized controlled trial (Andersson et al., 2012). After initial waiting-time individuals were randomized to the two forms of guidance. We had originally hypothesized that the support on demand group would experience smaller reductions in symptoms of social anxiety and related psychopathology and smaller increase in quality of life. In addition, the support on demand group was expected to lead to less demand for therapist input.

Method

The current study was part of the SOFIE-6 project, a study investigating the efficacy of ICBT for SAD (Andersson et al., 2012). Here, we report the results from the waitlist control group, which received treatment directly following the first group in the controlled trial.

Participants and Recruitment

A flowchart of the recruitment and treatment processes is presented in Figure 1. More information about the initial phase of the study can be found in Andersson et al. (2012). Participants were recruited via an email sent out to a waitlist who had registered interest on a public site hosted by the research group (www.studie.nu). An email invitation was sent to the first 600 names on the list. A total of 359 participants completed the screening questionnaires, and 272 completed the subsequent Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) (First et al., 1997) via telephone. The SCID-I interviews were conducted by 10 final-year students from the clinical psychologist program at Uppsala University, Sweden. They received training in administering the interviews before the study. Inclusion criteria were: a) at least 18 years old, b) living in Sweden, c) having access to a computer and an internet connection, d) meeting the criteria for SAD on the Social Phobia Screening Questionnaire (SPSQ) (Furmark et al., 1999), e) meeting the SCID-I criteria for SAD without meeting the criteria for a comorbid eating disorder or psychotic disorder, f) if applicable, having a stable dose of medication for the past two months, g) not undergoing current psychological treatment or having received psychological treatment during the past six months, h) providing informed consent via mail. In the original study, 204 participants met these criteria and were randomized to receive treatment either immediately during the autumn of 2008 or later. Once the post treatment data were collected, the control group participants (n = 99 after accounting for dropout) were randomized once again using a true random number generator (www.random.org) to receive either scheduled support or support on demand. Randomization during both phases was conducted by researchers not involved in other aspects of the study.

Treatment

The treatment was divided into nine modules that were unlocked one at a time, given that the participants had completed the assignments in the previous module. Modules were unlocked on a weekly basis, and participants were informed that this was the expected pace to keep during the treatment period. All unlocked modules were available for the duration of the treatment phase. Each module consisted of a PDF containing texts and practical exercises to complete during the week. A quiz was placed at the end of the modules boost adherence to the important principles of the treatment. Participants also provided a short written summary of the module in their weekly correspondence with the therapist (for the regular support group) or in a separate email to a non-specific therapist (in the support on-demand group). The content of the modules was identical to those used in previous studies within the SOFIE project (Furmark et al., 2009), which contained psychoeducation, cognitive restructuring, behavioral experiments, exposure exercises, and social skills training. An outline of the treatment is presented in Table 1. The modules spanned 188 pages, ranging from 17 to 30 pages per module.

Figure 1

Flowchart of the Recruitment and Assessments Throughout the SOFIE-6 Study

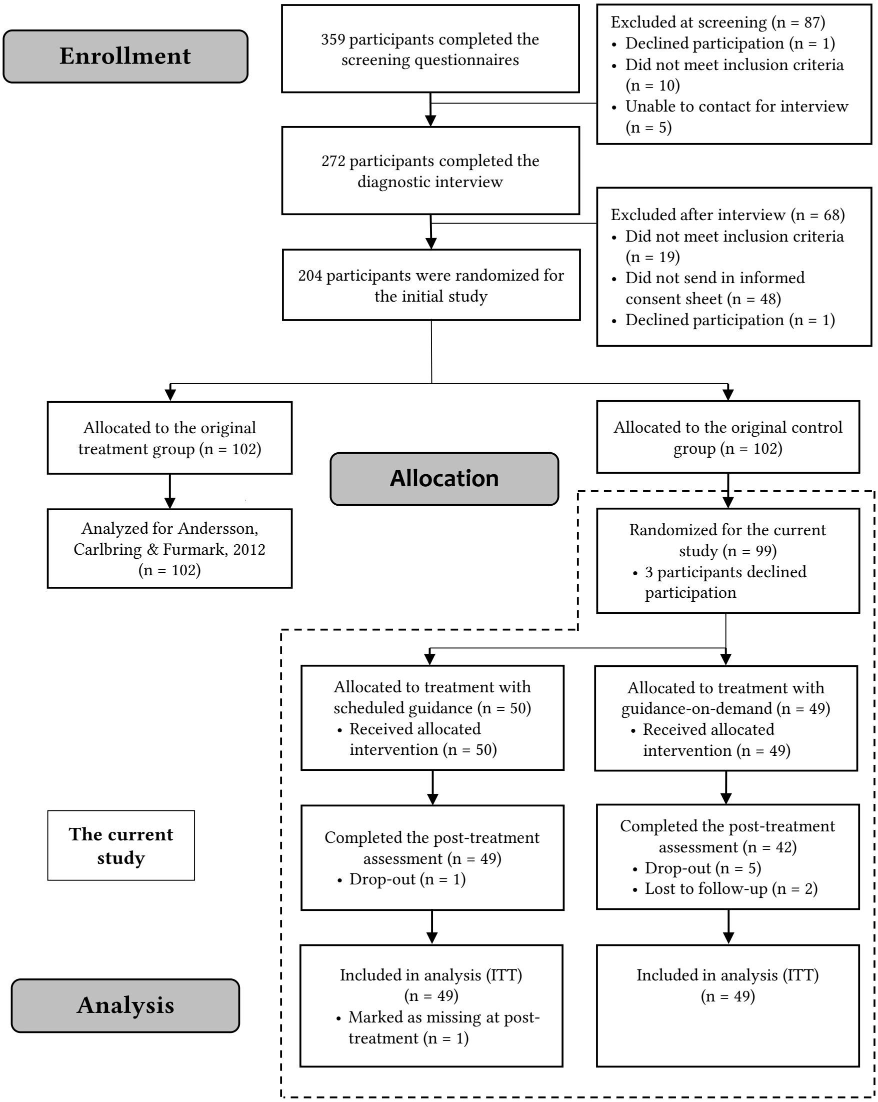


Table 1

Content of the Modules

Module	Content	Exercises	Number of pages (A4)
1	Introduction and psychoeducation	Learning about symptoms, anxiety hierarchy	18
2	Clarks and Wells' cognitive model of social anxiety	Personal model of social anxiety, thought record	20
3	Cognitive restructuring I	Reality testing, cognitive distortions, goals for the treatment	30
4	Cognitive restructuring II	Negative automatic thoughts, behavioral experiments	23
5	Exposure I	Exposure based on anxiety hierarchy	21
6	Shifting focus	Safety behaviors, exposure	19
7	Exposure II	Safety behaviors, exposure	17
8	Social skills	Social skills, exposure	19
9	Relapse prevention	Summary, plan for relapse prevention	21

Ten clinical psychologists served as clinicians during treatment. Communication between the participants and the clinician was conducted via a messaging system on the encrypted study website (Vlaescu et al., 2016). In addition to the messaging system, all participants had access to one of two anonymous discussion forums where they could write about the progress and experiences of the exercises conducted during the week. The scheduled support and support on demand groups had separate forums, and both forums were monitored by the study staff for safety. All participants received an introductory message, but for the group with scheduled support, this message was sent from their personal clinician, while the support on demand group participants received a generic message. The group with scheduled support received feedback on their exercises on a fixed day each week, which was the same day as when they received access to the next module. The support on demand group participants were told that they could contact the study staff via the messaging system. Participants in this condition were not assigned a specific clinician; rather, the clinicians had a schedule with days during which they would monitor the activity of the participants and respond to requests for help and feedback.

Measures

Primary Outcome Measure

Liebowitz Social Anxiety Scale – Self Report (LSAS-SR) — The LSAS-SR was the primary outcome measure. The LSAS-SR measures fear and avoidance related to social situations using 24 items (Fresco et al., 2001). Respondents are asked to rate their fear and anxiety regarding a social situation on a scale between 0 (no fear or anxiety) and 3 (severe fear or anxiety). They also rate how often they avoid the situation or scenario, ranging from 0 (never) to 3 (usually). The ratings are summed up to provide a general rating of social anxiety, ranging between 0 and 144. The self-report version of the scale has been noted to have excellent internal consistency (Cronbach's $\alpha = 0.95$) and a 12-week test–retest reliability of $r = .83$ (Baker et al., 2002). It has been validated for internet administration (Hedman et al., 2010). The questionnaire was administered online as a screening tool before the treatment began, weekly during the treatment (at a fixed day each week which was also the same time as participants were sent a new module if they had completed the previous module), and at post treatment.

Secondary Outcome Measures

All secondary outcome measures were administered at the screening (before this part of the study took place), at the pretreatment time point (the start of the current study), and at the post treatment time point.

Social Interaction Anxiety Scale (SIAS) — The SIAS consists of 20 items aimed at measuring the respondent's anxiety during social interactions (Heimberg et al., 1992). Ratings are made on a Likert scale from 0 (not at all characteristic or true of me) to 4 (extremely characteristic or true of me), with the total sum ranging from 0 to 80. Psychometric properties include excellent internal consistency (Cronbach's $\alpha = .93$) and a 12-week test–retest reliability of $r = .92$ (Mattick & Clarke, 1998).

Social Phobia Scale (SPS) — The SPS consists of 20 items administered with the intention of measuring respondents' fear of evaluation in social situations (Heimberg et al., 1992). Ratings are made on a Likert scale from 0 (not at all characteristic or true of me) to 4 (extremely characteristic or true of me), with a total sum range of 0 to 80. Psychometric properties include an internal consistency (Cronbach's $\alpha = .94$) and a 12-week test–retest reliability of $r = .93$ (Mattick & Clarke, 1998).

Beck Anxiety Inventory (BAI) — The BAI consists of 21 items that measure the physiological and cognitive symptoms of anxiety (Beck et al., 1988). Ratings are made on a four-point Likert scale, with possible sum scores ranging from 0 to 63. The instrument's psychometric properties include internal consistency (Cronbach's $\alpha = .92$) and a one-week test–retest reliability of $r = .75$ (Beck et al., 1988).

Montgomery Åsberg Depression Rating Scale – Self Report (MADRS-S) — The MADRS-S is a nine-item scale measuring symptoms of depression based on the 10-item clinician-administered version of the scale (Montgomery & Åsberg, 1979). Respondents rate the frequency of cognitive, emotional, and physiological symptoms during the past three days on a seven-point scale. Total sum scores can range from 0 to 54, with higher scores indicating an increased severity of symptoms. Psychometric properties for the self-report version have been reported to include an internal consistency (Cronbach's $\alpha = .84$) and a one-week intraclass correlation of .78 (Fantino & Moore, 2009). Both BAI and MADRS-S are validated for internet use (Thorndike et al., 2009).

Quality of Life Inventory (QoLI) — The QoLI is a 16-item instrument measuring respondents' subjectively rated quality of life (Frisch et al., 1992). The respondent is asked to indicate how important a specific domain is on a scale from 0 (not important) to 2 (very important), and then how satisfied they are with their current situation within that domain. The two ratings are multiplied and divided by the number of areas that the respondent considers to be somewhat or very important. The test–retest coefficient was measured between $r = .80$ and .91 (mean duration between measurements = 33 days) during the validation of the instrument. The range of values for internal consistency was reported as being between Cronbach's $\alpha = .77$ and .89. The QoLI, has been validated for internet use (Lindner et al., 2013).

Power

A formal power analysis was not conducted, as this was a spin-off study following the first phase of the trial. However, given the sample size and a power of 80% and $p < .05$, we had statistical power to detect an effect size of $d = 0.40$ on the LSAS-SR. This would correspond to a clinically relevant effect, with the expected direction being the superiority of scheduled support over support on demand.

Statistical Analyses

Statistical analyses were conducted using R version 4.0.3 (R Core Team, 2020) and SPSS version 25. Across the analyses, the alpha level was set to .05. Confidence intervals were reported at 95%. The assumption of normality was controlled using Shapiro-Wilks tests. Tests of pretreatment differences and differences between responders and non-responders on the post treatment assessment were evaluated using independent sample *t*-tests, Mann-Whitney *U* tests (when the assumption of parametric data was not met) and Fisher's exact tests. Independent *t*-tests were also used to investigate potential differences in the number of modules accessed (i.e., read) and completed (defined as completing the exercises in a module). A multiple regression model using residualized change scores as the dependent variable was used to investigate the relationship between completion of modules and change in the primary outcome measure. The data were analyzed according to the intention-to-treat principle (ITT), meaning that all available data were included in the analysis and all randomized participants were included in the analysis. The post treatment data from one of the participants in the scheduled support condition was flagged, as the scores on all the outcome measures were 0 (including both symptom measures, such as the LSAS and the quality-of-life ratings). Due to this likely mistake/error the post treatment data for this participant were marked as missing.

The model used to investigate the outcome of the primary outcome measure (LSAS-SR) was a mixed-effects model fitted using the *lme4* package (Bates et al., 2015). Model fit, including the form of change and covariance structure for the primary outcome where we had weekly measurements, was investigated iteratively using a likelihood ratio test (by using the *ANOVA* function in R). The final model for the primary outcome measure incorporated a linear rate of change, random intercept and slope, and an unstructured residual variance structure. For the secondary outcomes with only two data points, we estimated a random intercept but not a random slope. Q-Q plots were used to assess the normal distribution of the residuals for all the mixed models. Significance for the fixed effects in the models was evaluated using the Wald test, in which the estimate was divided by the standard error and compared against a *z*-distribution. Inferences about the random effects of the model for the primary outcome measure are not evaluated by the Wald test but rather from the estimated confidence intervals, where an interval not containing zero is interpreted in the same way as a significant *p*-value. Confidence intervals were calculated using the *ConfintMermod* function with the profile method. The models were estimated using restricted maximum likelihood estimation, thus making use of all available data. The use of maximum likelihood estimation is one of two recommended approaches for dealing with missing data (Schafer & Graham, 2002). Maximum likelihood estimators provide unbiased estimates in situations where data can be assumed to be missing at random (MAR), meaning that the data are not missing systematically as a function of the would-be value. This is a less restrictive assumption than missing completely at random (MCAR), where missingness is assumed to be independent of both the would-be value and the values of the other variables.

Due to differences in means between the conditions at pretreatment for the outcome measures, the parameter deemed to be of interest was the time x group interaction rather than the endpoint difference between the conditions. The conditions were coded as scheduled support = -0.5 and support on demand = 0.5.

The Cohen’s *d* between-group effect size for the estimated parameters of the models was calculated with the *lme.dscore* function using the Satterwaites degrees of freedom according to the formula $d = 2t/\text{Sqrt}(\text{df})$ (Rosenthal, 1994). Observed within-group effect sizes were calculated with the pooled standard deviations from the pre- and post treatment measurements. Between-group effect sizes were interpreted according to the recommendation provided by Cohen, with 0.20, 0.50, and 0.80 corresponding to small, moderate, and large effect sizes, respectively (Cohen, 1988).

Reliable change/deterioration was calculated according to the formula provided by Jacobson and Truax (1991), where the pretreatment mean was subtracted from the post treatment mean and divided by the pooled standard deviation adjusted for the instrument's test–retest reliability (Jacobson & Truax, 1991). The critical value for the LSAS-SR was set at ± 28 points.

Results

Baseline Characteristics

The demographic characteristics of the sample are presented in Table 2. The conditions did not differ significantly with regard to age, gender, civil status, or education level, all of which were $p > .05$.

Attrition, Missing Data, Activity Statistics, and Adherence

Five participants in the support-on-demand group (10%) dropped out of the study during the treatment period. One of the participants from the scheduled support group dropped out during the treatment period (2%). There was no significant difference in the proportion of dropouts between the two conditions, $\chi^2(1) = 2.93$, $p = .087$. For the sample as a whole, data were provided for 87% of the primary outcome measurements during the study. A total of 49 participants (98%) in the group with scheduled support completed all post treatment measurements. In the support on demand group, 42 participants (86%) completed all post treatment measures. The groups differed significantly in this regard, $\chi^2(1) = 5.03$, $p = .025$, suggesting that the support on demand group was less likely to complete the post treatment measurement. For the clinical and demographic variables, there were no significant differences between those who completed the post treatment assessment and those who did not (all $p > .05$).

Activity statistics are presented in Table 3. For the support on demand group, the average total number of messages sent to the clinician during the treatment was 0.6 ($SD = 1.10$, range = 0–4), which was much lower than in the scheduled support group ($M = 15.04$, $SD = 8.03$, range = 0–45) ($p < .001$). On average, the support on demand group accessed 77.4% of the modules, while the scheduled support group accessed 83.3%. This difference was not significant, $p = .359$. However, participants with scheduled support completed significantly more modules (79.2% on average) than participants in the support on demand group (64.2% on average), $p < .001$. The group receiving scheduled support also posted more on the discussion forum ($p < .001$). As expected, clinicians supporting the scheduled support group also spent more time on average attending to their participants than the clinician responsible for the support on demand group ($p <$ .001).

Table 2

Demographic and Clinical Characteristics of the Sample (n = 99)

Characteristic	Scheduled support		Support on demand			
	M	***SD***	***M***	***SD***	***t*(97)**	***p***
Age	39.44	10.60	37.59	11.42	0.84	.33
	n	**%**	***n***	**%**	χ^2	***p***
Gender						
Female	17	34.0	22	44.9	1.23	.32
Male	33	66.2	27	55.1		
Civil status						
Single	19	38.0	17	32.7	0.23	.63
In a relationship/Married	31	62.0	32	65.3		
Highest educational degree						
Primary school	1	2	3	6.1	5.98	.11
High school	12	24	13	26.5		
University	31	62	20	40.8		
Other post-secondary education	6	12	13	26.5		
	M	***SD***	***M***	***SD***	***t*(97)**	***p***
Outcome measure						
LSAS	58.76	24.14	69.71	21.99	-2.36	.020
SIAS	43.24	15.09	48.82	14.14	-2.74	.007
QoLI	1.29	1.66	0.22	1.59	3.30	.001
	M	***SD***	***M***	***SD***	***U***	***p***
SPS	28.80	13.73	36.69	14.94	2.822	.005
BAI	11.56	7.30	16.82	8.45	3.211	.001
MADRS-S	12.50	6.12	17.02	7.63	3.007	.003

Note. LSAS = Liebowitz Social Anxiety Scale – Self-rated; SIAS = Social Interaction Anxiety Scale; SPS = Social Phobia Scale; BAI = Beck Anxiety Inventory; MADRS-S = Montgomery-Åsberg Depression Rating Scale – Self-rated; QoLI = Quality of Life Inventory; *U* = Mann-Whitney *U*-test statistic.

Table 3

Statistics on Activity and Comparisons Between the Conditions

	Scheduled support	Support on demand		
Variable	*M* (*SD*)	*M* (*SD*)	*t*(97)	*p*
Number of emails sent by participants to the clinician	15.04 (8.03)	0.44 (1.09)	12.48	< .001
Number of posts made on the discussion forum	11.76 (7.99)	3.57 (5.24)	6.01	< .001
Modules accessed during treatment (out of nine)	7.50 (2.49)	6.97 (2.82)	0.92	.359
Modules completed during treatment (out of nine)	7.13 (2.48)	5.78 (2.89)	2.36	.020
Clinician time per week and participant (minutes)	14.00 (6.08)	0.6 (1.10)	15.30	< .001

The multiple regression model showed no significant predictive value in residualized gain score for neither condition, $\beta = -.33$, $p = .226$, or the number of completed modules, $\beta = .08$, $p = .585$. There was, however, an interaction between condition and module completion for the gain scores, $\beta = .55$, $p = .045$. This suggests that the number of completed modules was significantly related to a greater reduction in symptoms but only in the support on demand group. The explained variance in the LSAS-SR outcome was $R^2 = 0.141$.

Primary Outcome

Liebowitz Social Anxiety Scale – Self Report (LSAS-SR)

Observed means including effect sizes are reported in Table 4. For the LSAS-SR ratings, the mixed-effects model revealed significant heterogeneity in both the intercept, *SD* = 23.26, 95% CI [19.99, 26.81], and the slope, *SD* = 2.43, 95% CI [2.05, 2.84], across the sample. Additionally, the results showed a strong correlation between intercept and slope, $r = -.53$, 95% CI [-.67, -.35], suggesting that higher initial ratings were related to a steeper decline in symptoms during the treatment period. The fixed effects showed a significant difference between the groups at pretreatment, $b = 11.40$, 95% CI [2.02, 20.79], *SE* = 4.79, $p = .019$, indicating that the support on demand group had significantly higher ratings on the LSAS-SR at the start of the treatment. There was a significant linear decrease in symptoms over each unit of time (one week) for the entire sample, $b = -2.61$, 95% CI [-3.13, -2.09], *SE* = 0.26, $p < .001$. The interaction between time and group was not significant, *b* = -0.07, 95% CI [-1.10, 0.97], *SE* = 0.53, *p* = .898, suggesting that there was no significant difference in slope between the two conditions. The effect size for this comparison was *d* = 0.02, 95% CI [-0.37, 0.43], with the slight difference favoring the condition with support on demand.

Table 4

Observed Means for the Outcome Measures at Pre- and Post-Treatment With Within-Group Effect Sizes

	Pre-treatment		**Post-treatment**		**Observed within-group effect size**
Outcome measure	***M* (*SD*)**	***n***	***M* (*SD*)**	***n***	***d* [95% CI]**
LSAS					
Scheduled	58.76 (24.14)	50	37.80 (22.79)	49	-0.89 [-1.31, -0.48]
On demand	69.71 (21.99)	49	47.62 (22.43)	42	-1.00 [-1.43, -0.56]
SIAS					
Scheduled	43.24 (15.09)	50	32.90 (16.53)	49	-0.65 [-1.06, -0.25]
On demand	48.82 (14.14)	49	38.43 (15.96)	42	-0.69 [-1.12, -0.27]
SPS					
Scheduled	28.80 (13.73)	50	18.71 (13.27)	49	-0.75 [-1.16, -0.34]
On demand	36.69 (14.94)	49	23.76 (15.11)	42	-0.86 [-1.29, -0.43]
BAI					
Scheduled	11.56 (7.30)	50	8.18 (7.62)	49	-0.45 [-0.85, -0.05]
On demand	16.82 (8.45)	49	11.05 (7.68)	42	-0.71 [-1.14, -0.29]
MADRS-S					
Scheduled	12.50 (6.12)	50	7.86 (6.20)	49	-0.75 [-1.16, -0.35]
On demand	17.02 (7.63)	49	10.76 (7.36)	42	-0.83 [-1.26, -0.40]
QoLI					
Scheduled	1.29 (1.66)	50	2.04 (1.67)	49	0.45 [0.05, 0.85]
On demand	0.22 (1.59)	49	0.88 (1.61)	42	0.41 [0.00, 0.82]

Note. LSAS = Liebowitz Social Anxiety Scale – Self-rated; SIAS = Social Interaction Anxiety Scale; SPS = Social Phobia Scale; BAI = Beck Anxiety Inventory; MADRS-S = Montgomery-Åsberg Depression Rating Scale – Self-rated; QoLI = Quality of Life Inventory.

Secondary Outcomes

Social Interaction Anxiety Scale

The model did not indicate a significant initial difference between the conditions in the SIAS ratings *b* = 5.08, 95% CI [-4.46, 14.62], *SE* = 4.89, *p* = .301. Overall, the SIAS scores decreased during the treatment, *b* = -10.12, 95% CI [-16.42, -4.26], *SE* = 1.36, $p < .001$. The interaction between time and group was not significant, *b* = 0.50, 95% CI [-4.46, 5.81], *SE* = 2.71, *p* = .855. This difference in change equaled an effect size of *d* = 0.17, 95% CI [-0.35, 0.45] in favor of the group with scheduled support.

Social Phobia Scale

The analysis showed a significant initial difference between the conditions on the SPS, *b* = 10.04, 95% CI [1.32, 18.75], *SE* = 4.47, *p* = .026. There was a significant overall average decrease from pre to post treatment, *b* = -11.03, 95% CI [-15.88, -4.29], *SE* = 1.21, *p* < .001. The time x group interaction was not statistically significant, *b* = -2.15, 95% CI [-6.87, 2.59], *SE* = 2.42, *p* = .377. This difference in change corresponded to an effect size of *d* = 0.18, 95% CI [-0.22, 0.59] in favor of the support on demand group.

Beck Anxiety Inventory

The groups differed significantly in their initial BAI ratings, *b* = 7.49, 95% CI [2.61, 12.38], *SE* = 2.50, *p* = .003. There was an overall decrease in the BAI scores, *b* = -4.36, 95% CI [-5.74, -2.99], *SE* = 0.70, *p* < .001. The interaction between time and group was not statistically significant, *b* = -2.24, 95% CI [-6.83, 2.04], *SE* = 1.40, *p* = .115. The corresponded to an effect size of *d* = 0.34, 95% CI [-0.31, 1.03] favoring the support on demand group.

Montgomery Åsberg Depression Rating Scale – Self Report

There was a significant pretreatment difference in the MADRS-S scores, *b* = 5.73, 95% CI [1.40, 10.07], *SE* = 2.22, *p* = .011. After the treatment period, the analysis showed a significant decrease for the sample, *b* = -5.30, 95% CI [-6.54, -4.08], *SE* = 0.63, *p* < .001. The interaction between time and group was again not significant, *b* = -1.21, 95% CI [-3.68, 1.24], *SE* = 1.26, *p* = .338. The effect size for the difference in change between the groups was *d* = 0.20, 95% CI [-0.21, 0.61] favoring the support on demand group.

Quality of Life Inventory

The groups did not differ significantly in their pretreatment QoLI scores, *b* = -0.88, 95% CI [-2.46, 0.46], *SE* = 0.47, *p* = .064. The sample showed a significant increase in the QoLI during the treatment period, *b* = 0.66, 95% CI [0.23, 1.17], *SE* = 0.12, *p* < .001. The groups did not differ significantly in their changes during this period, as indicated by the interaction between group and time, *b* = -0.20, 95% CI [-1.01, 0.86], *SE* = 0.24, *p* = .423. The effect size for the difference in change between the groups during the treatment was *d* = 0.17, 95% CI [-0.76, 0.89] favoring the group with scheduled support.

Reliable Change/Deterioration

In total, 27 of the respondents (27%) at post treatment met the criteria for reliably improving during the treatment period. None of the participants were classified as reliably deteriorated. The proportion of clinically significantly improved participants did not differ between the scheduled support group (n = 12) and the support on demand group (n = 15), Fisher's exact p = .504.

Discussion

The aim of this study was to investigate the effects of a support on demand model for delivering ICBT targeting SAD relative to a standardized form of clinical support. Results suggested overall significant reductions in symptoms of SAD and related psychopathology, along with an increase in quality of life with no significant between-group differences. The effect sizes for the estimated within-group pre-to-post comparisons on the measures of social anxiety were all large. The reduction in the symptoms of social anxiety is consistent with earlier findings indicating that ICBT can be an effective alternative for treating SAD (Guo et al., 2021). Results further suggest that active therapist guidance may be reduced with support-on-demand without significant loss of treatment gains.

The lack of significant differences in change between the groups and the nonexistent-to-small between-group effect sizes are in line with the notion that support on demand can be a way of delivering ICBT for SAD. The findings also add to the literature on comparisons between the support on demand format and traditional ways of administering ICBT with weekly clinician support. Like earlier studies (Hadjistavropoulos et al., 2017; Hadjistavropoulos et al., 2019), the analyses indicated that the two conditions did not differ significantly in change during the treatment. Although the randomization "failed" as the groups differed at baseline, the lack of significant interactions between time and group suggests that support on demand can be a sufficiently effective way of delivering ICBT compared to the more established clinician-guided format. That has positive implications for scalability (Andersson et al., 2019). As expected, participants in the support on demand condition required significantly less clinician time per module than scheduled support participants. The average number of requests for help and/or feedback was low (M = 0.44), and none of the participants sent more than four messages to the clinician. Extrapolating from this, it is likely that a support on demand model could be a resource-effective way of disseminating ICBT for SAD, given that there are clinicians who are prepared to provide support when needed. This differentiates on demand ICBT from fully self-guided versions in which contact with clinicians is not offered or only possible in urgent cases. However, the reduced need for clinician support gives credibility to the idea that ICBT could be administered to a larger number of patients with relatively few clinicians, thus making it easier to disseminate in contexts where a lack of trained clinicians is a problem. As unguided interventions have sometimes been deemed less effective than interventions with scheduled support (Ciuca et al., 2018), a support on demand model could serve as a compromise, making it possible to disseminate more broadly with the decreased need for clinician support.

While the two guidance conditions produced comparable reductions in social anxiety, there were some differences in the activity levels between them. Participants in the group with scheduled support completed more modules, sent more emails to their clinician, and made more posts on the discussion forum than participants with support on demand. It is unclear whether activity levels such as these are important in relation to the outcome of ICBT in general, but the fact that module completion predicted a stronger reduction in symptoms in the support on demand group could be important. Future studies could investigate this relationship and whether module engagement in the support on demand condition can be increased with the addition of optional components such as personalized reminders (Hilvert-Bruce et al., 2012). For unguided ICBT, treatment credibility has also been noted as relevant to adherence (Nordgreen et al., 2012), and this would be interesting to investigate in relation to the on-demand format. Of note is that a significantly larger proportion of the participants in support on demand failed to complete the post treatment measures. This is likely due to a larger dropout rate during the treatment period. As the participants who provided post treatment ratings did have lower pretreatment scores on the BAI, the results for this outcome measure should be interpreted with caution. Inquiring about the reasons for dropout and non-adherence could be important going forward. Such information may inform decisions about who the support on demand format is a good match compared to a scheduled and structured mode of clinician support.

The results of the study should be viewed with some limitations in mind in addition to the fact that it is a file drawer study and hence could be less relevant even if technology in many ways has remained the same. First, the sample size was suboptimal for testing the differences between the two active treatment conditions. As Cuijpers et al. (2019) noted, studies investigating the components of psychological treatments often have far too small a sample to serve as outright non-inferiority trials (Cuijpers et al., 2019). It is important to note that the present study was not intended as such but rather a proof-of-concept trial regarding the ability to provide a new way of guiding participants through an ICBT treatment. When the SOFIE-6 study was conducted, no such trials had been published apart from studies testing the added value of scheduled telephone calls (Andersson et al., 2003; Kenwright et al., 2005). The results should not be interpreted as conclusive but rather as an indication that support on demand can be feasible in the treatment of SAD and possibly other conditions. Additional, better-powered trials are needed, along with studies on change mechanisms, as we do not know what works for whom in terms of support.

Second, the randomization procedure did yield unbalanced group in terms of their pre-treatment differences. Though the statistical analyses focused on the differences in change over time, rather than just the endpoint differences between the conditions, this should be kept in mind when interpreting the results and the outcome ratings.

Third, the study lacked data on some variables that might be of interest in addition to ratings of symptoms. For example, we did not measure treatment satisfaction or working alliance during and at post treatment. Although the groups did not differ significantly with regard to changes in the outcome measures, such information could be valuable when seeking to understand other factors that might be important, such as adherence and module completion.

Fourth, data were not collected beyond the post treatment assessment. Although the comparison of changes between the two conditions did not differ during the treatment period, the findings by Ivanova et al. (2016) indicated that differences in effect may occur later (Ivanova et al., 2016). While the long-term effects of ICBT in general is favorable (Andersson, 2018), future studies should strive to investigate the long-term effects of different support forms.

Lastly, though both the conditions had access to a forum, the condition with scheduled support made use of this function significantly more often. Given that a similar forum may produce symptom reductions (Griffiths et al., 2009), the fact that the design of the present study did not control for the specific effect of forum usage is a limitation.

In conclusion, the present study provides support for the role of support on demand as a way of delivering ICBT, and that the format is suitable in the treatment of SAD. It can also serve as an example of the importance of still reporting studies in which the data (in this case, randomization group differences) do not fulfill expectations. The findings are important, as groups exhibited very similar symptom trajectories during the treatment period, regardless of whether they received scheduled weekly support, or had the option to contact a clinician when needed. Additionally, no significant differences were found for any of the secondary measures. Given the small number of studies testing the support on demand format, we look forward to replications and systematic reviews when a sufficient number of trials have been conducted.

Funding: This study was sponsored in part by a grant from the Swedish Research Council to Professor Furmark, and a grant from Linköping University to Professor Andersson. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgments: We thank members of SOFIE-6 research group for their help with the original study.

Competing Interests: Three of the authors (TF, PC, GA) have published a self-help book based on the material tested in the study.

Ethics Statement: The study was approved by the ethics committee of Uppsala University and registered (Identifier: UMIN000001383).

References

Acarturk, C., Cuijpers, P., van Straten, A., & de Graaf, R. (2009). Psychological treatment of social anxiety disorder: A meta-analysis. *Psychological Medicine, 39*(2), 241–254. https://doi.org/10.1017/S0033291708003590

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Press.

Andersson, G. (2016). Internet-delivered psychological treatments. *Annual Review of Clinical Psychology, 12*, 157–179. https://doi.org/10.1146/annurev-clinpsy-021815-093006

Andersson, G. (2018). Internet interventions: Past, present and future. *Internet Interventions, 12*, 181–188. https://doi.org/10.1016/j.invent.2018.03.008

Andersson, G., Carlbring, P., Furmark, T., & the SOFIE Research Group. (2012). Therapist experience and knowledge acquisition in Internet-delivered CBT for social anxiety disorder: A randomized controlled trial. *PLoS One, 7*(5), Article e37411. https://doi.org/10.1371/journal.pone.0037411

Andersson, G., Lundström, P., & Ström, L. (2003). Internet-based treatment of headache: Does telephone contact add anything? *Headache, 43*(4), 353–361. https://doi.org/10.1046/j.1526-4610.2003.03070.x

Andersson, G., Titov, N., Dear, B. F., Rozental, A., & Carlbring, P. (2019). Internet-delivered psychological treatments: From innovation to implementation. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA), 18*(1), 20–28. https://doi.org/10.1002/wps.20610

Baker, S. L., Heinrichs, N., Kim, H. J., & Hofmann, S. G. (2002). The Liebowitz Social Anxiety Scale as a self-report instrument: A preliminary psychometric analysis. *Behaviour Research and Therapy, 40*(6), 701–715. https://doi.org/10.1016/S0005-7967(01)00060-2

Barkowski, S., Schwartze, D., Strauss, B., Burlingame, G. M., Barth, J., & Rosendahl, J. (2016). Efficacy of group psychotherapy for social anxiety disorder: A meta-analysis of randomized-controlled trials. *Journal of Anxiety Disorders, 39*, 44–64. https://doi.org/10.1016/j.janxdis.2016.02.005

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-efects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baumeister, H., Reichler, L., Munzinger, M., & Lin, J. (2014). The impact of guidance on Internet-based mental health interventions – A systematic review. *Internet Interventions, 1*(4), 205–215. https://doi.org/10.1016/j.invent.2014.08.003

Beck, A. T., Epstein, N., Brown, G., & Steer, R. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology, 56*(6), 893–897. https://doi.org/10.1037/0022-006X.56.6.893

Ciuca, A. M., Berger, T., Crişan, L. G., & Miclea, M. (2018). Internet-based treatment for panic disorder: A three-arm randomized controlled trial comparing guided (via real-time video sessions) with unguided self-help treatment and a waitlist control. PAXPD study results. *Journal of Anxiety Disorders, 56*, 43–55. https://doi.org/10.1016/j.janxdis.2018.03.009

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Cuijpers, P., Cristea, I. A., Karyotaki, E., Reijnders, M., & Hollon, S. D. (2019). Component studies of psychological treatments of adult depression: A systematic review and meta-analysis. *Psychotherapy Research, 29*(1), 15–29. https://doi.org/10.1080/10503307.2017.1395922

Dahlin, M., Johansson, A., Romare, H., Carlbring, P., & Andersson, G. (2022). Worry-specific versus self-tailored internet-based treatments for generalized anxiety disorder, with scheduled support or support on demand: A pilot factorial design trial. *Internet Interventions, 28*, Article 100531. https://doi.org/10.1016/j.invent.2022.100531

Dear, B. F., Gandy, M., Karin, E., Staples, L. G., Johnston, L., Fogliati, V. J., Wootton, B. M., Terides, M. D., Kayrouz, R., Perry, K. N., Sharpe, L., Nicholas, M. K., & Titov, N. (2015). The Pain Course: A randomised controlled trial examining an internet-delivered pain management program when provided with different levels of clinician support. *Pain, 156*(10), 1920–1935. https://doi.org/10.1097/j.pain.0000000000000251

Donker, T., Blankers, M., Hedman, E., Ljótsson, B., Petrie, K., & Christensen, H. (2015). Economic evaluations of Internet interventions for mental health: A systematic review. *Psychological Medicine, 45*(16), 3357–3376. https://doi.org/10.1017/S0033291715001427

El Alaoui, S., Hedman, E., Kaldo, V., Hesser, H., Kraepelien, M., Andersson, E., Rück, C., Andersson, G., Ljótsson, B., & Lindefors, N. (2015). Effectiveness of internet-based cognitive behavior therapy for social anxiety disorder in clinical psychiatry. *Journal of Consulting and Clinical Psychology, 83*(5), 902–914. https://doi.org/10.1037/a0039198

Fantino, B., & Moore, N. (2009). The self-reported Montgomery-Asberg Depression Rating Scale is a useful evaluative tool in Major Depressive Disorder. *BMC Psychiatry, 9*, Article 26. https://doi.org/10.1186/1471-244X-9-26

First, M. B., Gibbon, M., Spitzer, R. L., & Williams, J. B. W. (1997). *Structured clinical interview for DSM-IV Axis I Disorders (SCID-I).* American Psychiatric Press.

Fresco, D. M., Coles, M. E., Heimberg, R. G., Liebowitz, M. R., Hami, S., Stein, M. B., & Goetz, D. (2001). The Liebowitz Social Anxiety Scale: A comparison of the psychometric properties of self-report and clinician-administered formats. *Psychological Medicine, 31*(6), 1025–1035. https://doi.org/10.1017/S0033291701004056

Frisch, M. B., Cornell, J., Villanueva, M., & Retzlaff, P. J. (1992). Clinical validation of the Quality Of Life Inventory: A measure of life satisfaction for use in treatment planning and outcome assessment. *Psychological Assessment, 4*(1), 92–101. https://doi.org/10.1037/1040-3590.4.1.92

Furmark, T., Carlbring, P., Hedman, E., Sonnenstein, A., Clevberger, P., Bohman, B., Eriksson, A., Hållén, A., Frykman, M., Holmström, A., Sparthan, E., Tillfors, M., Nilsson Ihrfelt, E., Spak, M., Eriksson, A., Ekselius, L., & Andersson, G. (2009). Guided and unguided self-help for social anxiety disorder: Randomised controlled trial. *The British Journal of Psychiatry, 195*(5), 440–447. https://doi.org/10.1192/bjp.bp.108.060996

Furmark, T., Tillfors, M., Everz, P., Marteinsdottir, I., Gefvert, O., & Fredrikson, M. (1999). Social phobia in the general population: Prevalence and sociodemographic profile. *Social Psychiatry and Psychiatric Epidemiology, 34*(8), 416–424. https://doi.org/10.1007/s001270050163

Griffiths, K. M., Calear, A. L., & Banfield, M. (2009). Systematic review on Internet Support Groups (ISGs) and depression (1): Do ISGs reduce depressive symptoms? *Journal of Medical Internet Research, 11*, Article e40. https://doi.org/10.2196/jmir.1270

Guo, S., Deng, W., Wang, H., Liu, J., Liu, X., Yang, X., He, C., Zhang, Q., Liu, B., Dong, X., Yang, Z., Li, Z., & Li, X. (2021). The efficacy of internet-based cognitive behavioural therapy for social anxiety disorder: A systematic review and meta-analysis. *Clinical Psychology & Psychotherapy, 28*(3), 656–668. https://doi.org/10.1002/cpp.2528

Hadjistavropoulos, H. D., Schneider, L. H., Edmonds, M., Karin, E., Nugent, M. N., Dirkse, D., Dear, B. F., & Titov, N. (2017). Randomized controlled trial of internet-delivered cognitive behaviour therapy comparing standard weekly versus optional weekly therapist support. *Journal of Anxiety Disorders, 52*, 15–24. https://doi.org/10.1016/j.janxdis.2017.09.006

Hadjistavropoulos, H. D., Schneider, L. H., Mehta, S., Karin, E., Dear, B. F., & Titov, N. (2019). Preference trial of internet-delivered cognitive behaviour therapy comparing standard weekly versus optional weekly therapist support. *Journal of Anxiety Disorders, 63*, 51–60. https://doi.org/10.1016/j.janxdis.2019.02.002

Hedman, E., Furmark, T., Carlbring, P., Ljótsson, B., Rück, C., Lindefors, N., & Andersson, G. (2011). A five-year follow-up of internet-based cognitive behaviour therapy for social anxiety disorder. *Journal of Medical Internet Research, 13*, Article e39. https://doi.org/10.2196/jmir.1776

Hedman, E., Ljótsson, B., Rück, C., Furmark, T., Carlbring, P., Lindefors, N., & Andersson, G. (2010). Internet administration of self-report measures commonly used in research on social anxiety disorder: A psychometric evaluation. *Computers in Human Behavior, 26*(4), 736–740. https://doi.org/10.1016/j.chb.2010.01.010

Heimberg, R. G., Mueller, G. P., Holt, C. S., Hope, D. A., & Liebowitz, M. R. (1992). Assessment of anxiety in social interaction and being observed by others: The social interaction anxiety scale and the Social Phobia Scale. *Behavior Therapy, 23*(1), 53–73. https://doi.org/10.1016/S0005-7894(05)80308-9

Hilvert-Bruce, Z., Rossouw, P. J., Wong, N., Sunderland, M., & Andrews, G. (2012). Adherence as a determinant of effectiveness of internet cognitive behavioural therapy for anxiety and depressive disorders. *Behaviour Research and Therapy, 50*(7–8), 463–468. https://doi.org/10.1016/j.brat.2012.04.001

Ivanova, E., Lindner, P., Ly, K. H., Dahlin, M., Vernmark, K., Andersson, G., & Carlbring, P. (2016). Guided and unguided Acceptance and Commitment Therapy for social anxiety disorder and/or panic disorder provided via the Internet and a smartphone application: A randomized controlled trial. *Journal of Anxiety Disorders, 44*, 27–35. https://doi.org/10.1016/j.janxdis.2016.09.012

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*(1), 12–19. https://doi.org/10.1037/0022-006X.59.1.12

Käll, A., Backlund, U., Shafran, R., & Andersson, G. (2020). Lonesome no more? A two-year follow-up of internet-administered cognitive behavioral therapy for loneliness. *Internet Interventions, 19*, Article 100301. https://doi.org/10.1016/j.invent.2019.100301

Kenwright, M., Marks, I., Graham, C., Franses, A., & Mataix-Cols, D. (2005). Brief scheduled phone support from a clinician to enhance computer-aided self-help for obsessive-compulsive disorder: Randomized controlled trial. *Journal of Clinical Psychology, 61*(12), 1499–1508. https://doi.org/10.1002/jclp.20204

Kishimoto, T., Krieger, T., Berger, T., Qian, M., Chen, H., & Yang, Y. (2016). Internet-based cognitive behavioral therapy for social anxiety with and without guidance compared to a wait list in China: A propensity score study. *Psychotherapy and Psychosomatics, 85*(5), 317–319. https://doi.org/10.1159/000446584

Koelen, J. A., Vonk, A., Klein, A., de Koning, L., Vonk, P., de Vet, S., & Wiers, R. (2022). Man vs. machine: A meta-analysis on the added value of human support in text-based internet treatments ("e-therapy") for mental disorders. *Clinical Psychology Review, 96*, Article 102179. https://doi.org/10.1016/j.cpr.2022.102179

Lindner, P., Andersson, G., Öst, L.-G., & Carlbring, P. (2013). Validation of the Internet-administered Quality of Life Inventory (QOLI) in different psychiatric conditions. *Cognitive Behaviour Therapy, 42*(4), 315–327. https://doi.org/10.1080/16506073.2013.806584

Mattick, R. P., & Clarke, J. C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour Research and Therapy, 36*(4), 455–470. https://doi.org/10.1016/S0005-7967(97)10031-6

Mayo-Wilson, E., Dias, S., Mavranezouli, I., Kew, K., Clark, D. M., Ades, A. E., & Pilling, S. (2014). Psychological and pharmacological interventions for social anxiety disorder in adults: A systematic review and network meta-analysis. *The Lancet Psychiatry, 1*(5), 368–376. https://doi.org/10.1016/S2215-0366(14)70329-3

Montgomery, S. A., & Åsberg, M. (1979). A new depression scale designed to be sensitive to change. *The British Journal of Psychiatry, 134*(4), 382–389. https://doi.org/10.1192/bjp.134.4.382

Nordgreen, T., Havik, O. E., Öst, L.-G., Furmark, T., Carlbring, P., & Andersson, G. (2012). Outcome predictors in guided and unguided self-help for social anxiety disorder. *Behaviour Research and Therapy, 50*(1), 13–21. https://doi.org/10.1016/j.brat.2011.10.009

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The Handbook of research synthesis* (pp. 231–244). Russell Sage Foundation.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147–177. https://doi.org/10.1037/1082-989X.7.2.147

Stein, D. J., Lim, C. C. W., Roest, A. M., de Jonge, P., Aguilar-Gaxiola, S., Al-Hamzawi, A., Alonso, J., Benjet, C., Bromet, E. J., Bruffaerts, R., de Girolamo, G., Florescu, S., Gureje, O., Haro, J. M., Harris, M. G., He, Y., Hinkov, H., Horiguchi, I., Hu, C., . . . Scott, K. M. (2017). The cross-national epidemiology of social anxiety disorder: Data from the World Mental Health Survey Initiative. *BMC Medicine, 15*(1), Article 143. https://doi.org/10.1186/s12916-017-0889-2

Steinert, C., Hofmann, M., Leichsenring, F., & Kruse, J. (2013). What do we know today about the prospective long-term course of social anxiety disorder? A systematic literature review. *Journal of Anxiety Disorders, 27*(7), 692–702. https://doi.org/10.1016/j.janxdis.2013.08.002

Thorndike, F. P., Carlbring, P., Smyth, F. L., Magee, J., Gonder-Frederick, L., Öst, L.-G., & Ritterband, L. M. (2009). Web-based measurement: Effect of completing single or multiple items per webpage. *Computers in Human Behavior, 25*(5), 393–401. https://doi.org/10.1016/j.chb.2008.05.006

van Dis, E. A. M., van Veen, S. C., Hagenaars, M. A., Batelaan, N. M., Bockting, C. L. H., van den Heuvel, R. M., Cuijpers, P., & Engelhard, I. M. (2020). Long-term outcomes of cognitive behavioral therapy for anxiety-related disorders: A systematic review and meta-analysis. *JAMA Psychiatry, 77*(3), 265–273. https://doi.org/10.1001/jamapsychiatry.2019.3986

Vlaescu, G., Alasjö, A., Miloff, A., Carlbring, P., & Andersson, G. (2016). Features and functionality of the Iterapi platform for internet-based psychological treatment. *Internet Interventions, 6*, 107–114. https://doi.org/10.1016/j.invent.2016.09.006

EACLIPT

Clinical Psychology in Europe (CPE) is the official journal of the European Association of Clinical Psychology and Psychological Treatment (EACLIPT).

leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.

PsychOpen GOLD

Clinical Psychology
in Europe

PsychOpen GOLD

Research Articles

Check for updates

Building an Early Warning System for Depression: Rationale, Objectives, and Methods of the WARN-D Study

Eiko I. Fried[1], Ricarda K. K. Proppert[1], Carlotta L. Rieble[1]

[1] *Department of Clinical Psychology, Leiden University, Leiden, The Netherlands.*

Clinical Psychology in Europe, 2023, Vol. 5(3), Article e10075, https://doi.org/10.32872/cpe.10075

Received: 2022-08-15 • **Accepted:** 2023-07-02 • **Published (VoR):** 2023-09-29

Handling Editor: Winfried Rief, Philipps-University of Marburg, Marburg, Germany

Corresponding Author: Eiko I. Fried, Leiden University, Wassenaarseweg 52, 2333 AK Leiden, The Netherlands. E-mail: eikofried@gmail.com

Supplementary Materials: Materials [see Index of Supplementary Materials]

Abstract

Background: Depression is common, debilitating, often chronic, and affects young people disproportionately. Given that only 50% of patients improve under initial treatment, experts agree that prevention is the most effective way to change depression's global disease burden. The biggest barrier to successful prevention is to identify individuals at risk for depression in the near future. To close this gap, this protocol paper introduces the WARN-D study, our effort to build a personalized early warning system for depression.

Method: To develop the system, we follow around 2,000 students over 2 years. Stage 1 comprises an extensive baseline assessment in which we collect a broad set of predictors for depression. Stage 2 lasts 3 months and zooms into participants' daily experiences that may predict depression; we use smartwatches to collect digital phenotype data such as sleep and activity, and we use a smartphone app to query participants about their experiences 4 times a day and once every Sunday. In Stage 3, we follow participants for 21 months, assessing transdiagnostic outcomes (including stress, functional impairment, anxiety, and depression) as well as additional predictors for future depression every 3 months. Collected data will be utilized to build a personalized prediction model for depression onset.

Discussion: Overall, WARN-D will function similarly to a weather forecast, with the core difference that one can only seek shelter from a thunderstorm and clean up afterwards, while depression may be successfully prevented before it occurs.

Keywords

ecological momentary assessment, digital health, student mental health, depression, prediction, early warning system, prevention

Highlights

- Prevention of depression in students may help prevent a lifetime of chronic illness.
- But we do not currently know whom to target in prevention programs, and when exactly.
- We introduce the protocol of the WARN-D study, aimed at building an early warning system for depression in students.
- To do so, we leverage advances in theory (complexity science), measurement (smartphone, smartwatch, and registry data), and statistical modelling (machine learning, network models).

Depressive disorders are prevalent, debilitating, and costly, and therefore among the most pressing health problems of modern living. They affect around 300 million people worldwide (Arias-de la Torre et al., 2021; Ferrari et al., 2013; James et al., 2018), are the leading cause of disability in the world, and are among the leading causes of global disease burden (Lopez et al., 2006; Mathers & Loncar, 2006). Major Depressive Disorder (MDD) is the strongest predictor for suicide (Berman, 2009), with 1 million lives lost annually (World Health Organization, 2019). Being depressed worsens the impact of common diseases like cancer and cardiovascular disease (Cuijpers et al., 2012), and about 60% of people living with depression report severe, long-lasting impairment of functioning, compromising the capacity for self-care and independent living (Kessler, Chiu, et al., 2005; Mathers & Loncar, 2006; Murray & Lopez, 1996). MDD is often chronic: Over half of depressed patients will develop multiple episodes, and many will spend a considerable part of their lifetime in a state of emotional agony and despair (Cuijpers et al., 2012).

Compared to progress in treating diseases like cancer (Biemar & Foti, 2013), breakthroughs for treating depression have lagged far behind. Treatment effectiveness has remained stable over the last decades (Khan & Brown, 2015). Around half of patients remain depressed following initial treatment with psychological therapies or pharmacotherapy (Cuijpers et al., 2018; Khan & Brown, 2015), and treatments reduce only one-third of the disease burden (van Zoonen et al., 2014). Mechanisms underlying MDD remain largely opaque, despite considerable efforts and investments into trying to understand biological underpinnings (Kapur et al., 2012; Rogers, 2017).

It is for these reasons that experts agree that *prevention*—stopping depression before it occurs—is the most important way forward to make a real difference in people’s lives (Cuijpers et al., 2012; Muñoz et al., 2010). Since 60-75% of all mental health problems develop before the age of 24, young people are an especially important group for prevention (Kessler, Berglund, et al., 2005; Solmi et al., 2022). While some progress has been made in developing and testing prevention programs that can effectively lower incidence rates by levels considered clinically relevant, improving prevention crucially relies on the reliable detection of specific individuals at risk for depression in the near future, which is currently not possible (Cuijpers et al., 2012; Muñoz et al., 2010; van Zoonen et al., 2014).

The study we describe here aims to tackle one of the largest barriers to implementing successful, tailored prevention programs: knowing when to intervene, and in which people. We address this problem by developing the personalized early warning system WARN-D. In the following, we will introduce the guiding principles of WARN-D; discuss the challenges of conceptualizing and measuring depression; describe the design, procedure, and measures of WARN-D; and conclude with strengths and challenges of the study.

Principles Guiding the Development of WARN-D

Our study's design, methods, and measures are guided by 6 primary goals and principles.

First, our most ambitious goal is to identify at-risk individuals before they transition into depression. We hope that our efforts will result in the first personalized early warning system for depression.

Second, we will develop this system in and for students, because timely detection of depression onset in young people promises to enable prevention programs to alleviate a potential lifetime of suffering for many, given the often-chronic nature of MDD. Students are at considerable risk for developing depression and comorbid mental health problems (Auerbach et al., 2016; Ebert et al., 2019), and the recent WHO World Mental Health Surveys International College Student Project reported that of ~14,000 full time students across 9 countries, including the US, Mexico, Germany, Belgium, and South Africa, the 12-month prevalence for any mental health disorder was ~31% (Auerbach et al., 2018). Another reason we focus on students is because MDD is highly heterogeneous in terms of both etiology and the problems people experience (Fried, Flake, & Robinaugh, 2022; Kendler, 2012a; Zimmerman et al., 2015), and efforts to understand and predict depression onset are more likely to succeed in more homogeneous populations (Cai et al., 2015, 2020). Moreover, prediction projects such as WARN-D require large samples, which are feasible to recruit in student populations, and students have the skills to operate the smartphone and smartwatch applications required for remote participation.

Third, WARN-D should be feasible for implementation in real-world settings. This precludes repeated lab visits and costly, time-intensive measurement such as brain scans and other biomarkers, which also do not appear to robustly predict depression onset (Border et al., 2019; Kennis et al., 2020; Winter et al., 2022). We instead focus on types of data that can readily be collected in the daily lives of students, including self-report surveys collected via smartphones, smartwatch data, and registry data. We will investigate the feasibility of our data collection protocol by querying participants about perceived burden of and barriers to participation.

Fourth, we aim to build a generic infrastructure that can be transferred and applied to many other disorders and target populations. If successful, WARN-D may spawn a host of follow-up projects that use the same infrastructure to provide personalized prediction of e.g., PTSD in military personnel, burnout in at-risk teachers, or manic episodes in recovered patients with bipolar disorders at risk for relapse. This promises to answer important scientific questions about personalized prediction across a range of mental disorders, such as which risk factors are transdiagnostic, which risk factors are disorder-specific, and which risk factors are specific to people with certain (e.g., demographic) features. This, in turn, relates to the identification of potentially novel mechanisms of change to inform future prevention programs (Nock, 2007).

Fifth, our study is guided by open scholarship principles. We are excited to make our design, measures, code, and data available to the research community. Information on design and measures are available in the accompanying Supplementary Materials. All empirical papers will be accompanied by open code; and we are currently developing a data sharing protocol with all relevant stakeholders which will be ready in 2025/2026 by the time data collection is finished. See our WARN-D project hub for all future updates and publications.

The final principle driving our design, methods, and measures is to conceptualize depression consistent with what we have learned about the complexities of the construct in the last decades (Fried, Flake, & Robinaugh, 2022). The next section is dedicated to this challenge.

Conceptualization of Depression

Depression is a complex construct, and any study aiming to understand and predict MDD onset must grapple with these complexities. Challenges include (1) the heterogeneity of MDD in terms of etiology and symptoms; (2) depression severity as a continuum; (3) inter-individual differences of people diagnosed with MDD; (4) and the dynamic nature of MDD. We discuss these one by one below, and address how we aim to tackle them in WARN-D.

Heterogeneity of Risk Factors and Symptoms

MDD is highly multifactorial, with many identified risk factors, all of which explain comparably little variance in isolation (Kendler, 2012a). Depression is also highly multifaceted: common rating scales for depression encompass over 50 separate symptoms (Fried, Flake, & Robinaugh, 2022), and there is increasing evidence that symptoms are not interchangeable (Fried & Nesse, 2015b). For example, specific individual symptoms feature differential relations to constructs including impairment (Fried & Nesse, 2014; Tweed, 1993), biological markers (Frank et al., 2021; Fried et al., 2020; Hilland et al., 2020; Nagel et al., 2018; Van Eeden et al., 2020), life events (Fried et al., 2015; Keller et al., 2007; Keller & Nesse, 2005), and treatments (Boschloo, Bekhuis, et al., 2019; Boschloo, Cuijpers, et al., 2019; Snippe et al., 2021). Further, there is evidence that depression is not unidimensional, i.e., cannot be adequately described as *one* process (Fried et al., 2016). Together, this calls into question the practice of modeling depression as a single variable or process, and its etiology as driven by a small number of factors. Instead, it suggests the study of a broad set of biological, psychological, and social risk factors, protective factors, as well as problems or symptoms participants experience nested under the umbrella of the depressive phenotype (Engel, 1977).

Depression Severity as a Continuum

Case-control studies are commonplace in depression research, where 2 groups (healthy vs depressed) are compared. This is widely recognized as a fundamental barrier to insights (Fried, Flake, & Robinaugh, 2022; Hitchcock et al., 2022), and categorical conceptualizations ignore subclinical cases who have increased levels of functional impairment, socioeconomic burden, service use, suicide attempts, and worse prognosis (Cuijpers & Smit, 2004; Gotlib et al., 1995; Hetrick et al., 2008; Judd et al., 1997). Dimensional perspectives in which subclinical cases are not subsumed into the category of healthy individuals offer ways forward that conceptualize depression as a continuum between healthy and sick, and align with evidence that depression behaves as a continuum at the between-subjects level (Conway et al., 2019; Haslam, 2003; Haslam et al., 2012), rather than a category or taxon.

Inter-Individual Differences Within MDD

People diagnosed with MDD often differ from each other in fundamental ways regarding symptoms and etiology, and subsuming them into one group can obfuscate pronounced inter-individual differences (Fried, Flake, & Robinaugh, 2022; Fried & Nesse, 2015a; Kendler, 2012b; Zimmerman et al., 2015). Two patients can have the same DSM-5 diagnosis of MDD without sharing a single symptom, and knowing that a person is diagnosed with MDD tells us little about the actual problems they face in daily life (McWilliams, 2021; Parker, 2005). Longitudinal data combined with statistical approaches that can leverage such data efficiently (e.g., network models, machine learning models) allow researchers to disentangle group-level processes (i.e., the nomothetic) from personalized processes (i.e., the idiographic) in order to find out to which degree processes are shared across people (Fisher et al., 2018).

Depression as a Dynamic Phenotype

This leads to the next challenge: the dynamic nature of depression (Hetrick et al., 2008; Judd et al., 1998; Wichers, 2014). There is sparse data on the nature of transitions into depression in the first place: are they largely categorical (i.e., a catastrophic transition), continuous (i.e., a process that unfolds slowly over weeks), or are there considerable inter-individual differences in how people transition into depression? Further, comparably little empirical work has been conducted on the depressive prodrome: what are the prominent features that could serve as early warning signals (EWS) for upcoming transitions into depression? Studies have identified a host of prodromal signs such as anxiety, sleep disturbances, worthlessness, sad mood, and concentration problems (Fava & Tossani, 2007; Iacoviello et al., 2010; Murphy et al., 2002), but results are inconsistent across studies, and prospective studies in large samples, including a long period of daily assessments, do not exist. One of the most comprehensive studies on the topic collected data every 6 weeks (Iacoviello et al., 2010), but cannot provide insights into daily fluctuations of problems. Such dynamic challenges require dynamic data, including daily reports of experiences, affect states, problems, and contextual variables whose fluctuations may shed light on upcoming transitions (Kuppens, 2015; van de Leemput et al., 2014; Wichers et al., 2016).

WARN-D Embraces the Complexity of Depression

In sum, depression is a complex, dynamic, heterogeneous phenotype. To embrace this complexity, WARN-D is guided by the rationale of depression as emerging from a system of biopsychosocial elements (Fried, 2022), which we term the *human mood system.* We conceptualize this system broadly, including time-invariant (or very slow-moving) features such as personality; time-varying features such as a person's thoughts, feelings, and behaviors; as well as the context in which experiences are made; a detailed list of all assessed features is provided later. Understanding this human mood system and its development requires the study of a broad set of system elements as well as their interrelations (Borsboom, 2017; Olthof et al., 2023), which is why we use smartphones and smartwatches to gather dynamic data. Conceptualizing complex processes as multivariate, multicausal systems has resulted in many breakthroughs in disciplines such as ecology, meteorology, medicine, public health, and social dynamics (Barabási, 2012; Castellano et al., 2009; Luke & Stamatakis, 2012; Olde Rikkert et al., 2016; Quax et al., 2018). In clinical psychology and psychiatry, recent studies have demonstrated the potential utility of a systems approach for understanding mental health problems like depression (Hayes & Andrews, 2020; Lutz et al., 2018; Olthof et al., 2023; Robinaugh et al., 2020; Wichers, 2014). Of particular interest are EWS that have been uncovered in many different areas of research, showing that systems close to transitions into alternative states (e.g., from healthy states to disordered states) show particular behavior that can be leveraged to forecast upcoming transitions (Olthof et al., 2020; van de Leemput et al., 2014; Wichers et al., 2016).

WARN-D hopes to identify such markers in the human mood system to predict individuals at risk for imminent system shifts into depression. In our communication with participants, we use weather forecasting and thunderstorms as a metaphor for this: thunderstorms are not best predicted by increases in thunderstorms, and in the same way, monitoring symptoms over time may not be the best way to predict depression onset. Instead, thunderstorms are best predicted by monitoring features of the weather system, along with the dynamic relations among these features. Together, these can provide evidence of upcoming changes in the system. The main difference between forecasting thunderstorms and depression is that in the former case, if we successfully anticipate an upcoming storm, all we can do is to accept the incoming storm, seek shelter, and try to clean up afterwards. For depression, successful prediction may allow us to prevent depression before it occurs in the first place.

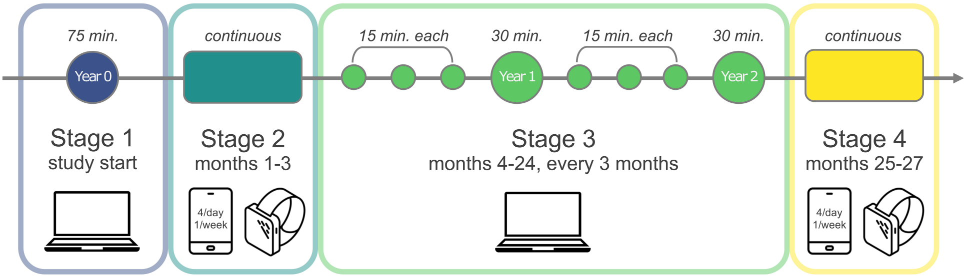
WARN-D Design, Procedure, and Measurement

Design

We plan to follow 2,000 students from vocational schools, technical universities, and universities in the Netherlands for ~2 years, using a multicohort design with 4 cohorts of 500 students each. The 4 cohorts start in November 2021, May 2022, November 2022, and May 2023, respectively. The timeline of the project is visualized in Figure 1 and includes 4 stages.

Figure 1

Overview of Design and Procedure of the WARN-D Study



Note. The study takes place in 4 cohorts with a target n = 500 per cohort, and each cohort runs for 2 years through Stages 1, 2, and 3. Starting times for cohorts are November 2021, May 2022, November 2022, and May 2023. For Stage 4, the repetition of Stage 2, we will re-invite all participants from cohorts 1 and 2. Attribution of images: laptop, phone, and smartwatch by Mello, Rabi'ah Al Adawiyyah, and Smashicons, respectively (Noun Project, CC BY 3.0).

After participants meet inclusion criteria based on a brief online screener, Stage 1 consists of a 75-minute online survey with the goal to assess risk factors for depression broadly. Stage 2 collects daily smartwatch and smartphone data, obtaining detailed insights into students' lives. Stage 3 consists of 8 online surveys, every 3 months, to determine if changes in mental health have occurred, and to assess risk and resilience factors.

The study officially terminates after Stage 3, at which point we plan to re-invite all participants from cohorts 1 and 2 for Stage 4, which is a repetition of Stage 2, i.e., another 3 months of daily monitoring. Study design, procedure, inclusion and exclusion criteria, and measurement are described in more detail in the Supplementary Materials.

Procedure

The WARN-D study is funded by the European Research Council under the European Union's Horizon 2020 research and innovation program (No. 949059). The data collection was approved by the Leiden University Research Ethics Committee Leiden (2021-09-06-E.I.Fried-V2-3406). The study was exempted from having to obtain ethics approval under the Medical Research Involving Human Subjects Act.

Although data collection is finished for some cohorts, it is still running for others; therefore, we will use present tense in the remainder of the procedure section. We advertise the study both online and offline, and partnered with several initiatives (e.g., Caring Universities) and educational institutions (e.g., MBO Rijnland) to reach students. Participants interested in participating receive a link to an online survey. Upon signing up, they can choose their preferred language (Dutch or English), and then read and sign the informed consent materials. After a screener on inclusion and exclusion criteria described below, participants are invited to Stage 1 of the study; completing Stage 1 is mandatory to be invited to Stage 2.

We pay participants up to 90€ for completing all surveys in Stage 1 (7.50€), Stage 2 (45€), and Stage 3 (37.50€; 7.50€ for the 30-minute surveys at 12 and 24 months, and 3.75€ for the 15-minute surveys at 3, 6, 9, 15, 18, and 21 months). Participation in Stage 4 yields up to 45€. Participants who complete the 1-year and 2-year follow-up surveys in Stage 3 can participate in 500€ lotteries for each survey. Further, participants completing Stage 2 receive a personalized report of the self-report data collected via smartphones, based on our experiences in a recent study that this is of great interest to many participants (Fried, Papanikolaou, & Epskamp, 2022).

Inclusion and Exclusion Criteria

Participants qualify for the study if they meet the following criteria: ≥18 years old; fluent in reading Dutch or English; studying at a Dutch educational institution pursuing an MBO (vocational school), HBO (higher vocational school), or WO (university) degree (no PhD students); currently living in the Netherlands, Germany, or Belgium (this is to ensure that smartwatches can be shipped in time); having a European bank account (for reimbursement purposes); and having a smartphone that runs on Android or iOS so that the apps required for Stage 2 work without problems.

Participants are excluded if they meet any of the following 6 criteria. First, at least moderate levels of current depression, operationalized via a score of ≥2 on the 2-item Patient Health Questionnaire (PHQ-2; Kroenke et al., 2003) and then a score of ≥14 on the 9-item Patient Health Questionnaire (PHQ-9; Kroenke & Spitzer, 2013). Second, current mania, operationalized via the corresponding items on the American Psychiatric Association's (APA) "DSM-5 Self-Rated Level 1 Cross-Cutting Symptom Measure—Adult" (Narrow et al., 2013), from here on referred to as the Level 1 screener, followed by APA's recommended Level 2 screener, the Altman Self-Rating Mania Scale (Altman et al., 1997); participants are excluded if they meet thresholds on both Level 1 (≥2 on either of the 2 items) and Level 2 (sum score ≥6) screeners. Third, current thought disorders, operationalized via the Level 1 screener (sum score ≥1). Fourth, substance use disorder via the Alcohol, Smoking and Substance Involvement Screening Test (ASSIST v3.0), using the cutoff of ≥27 for each substance (WHO ASSIST Working Group, 2010). Fourth, we exclude participants reporting that they are currently in treatment or waiting for treatment for the mental health problems described above. Fifth, we exclude students with at least moderate current suicidal ideation, operationalized via a score of 2 on item 4 of the Beck Scale for Suicide Ideation (BSS; Beck et al., 1979), which has shown excellent psychometric properties to screen for suicidal ideation, including in a Dutch sample (De Beurs et al., 2014). Finally, we exclude participants who indicate that they would find seeing an estimate of daily calories burned very stressful, given that the smartwatches worn in Stage 2 provide such an estimate.

Measurement

There are many tools to measure constructs in clinical psychology and psychiatry. For our baseline and follow-up assessments, we based our selection of measures on 5 guiding principles:

1. Scales should assess constructs relevant to understanding the human mood system and predict changes of the system (e.g., protective and risk factors).
2. Scales should be free to use and in the public domain.
3. Scales should have adequate psychometric properties.
4. Scales should be validated in both English and Dutch.
5. Scales should be short without sacrificing content validity.

Some of the measures had to be created, translated, or adapted. Guiding principles for measure adaptation were:

1. Adapt as little as possible.

2. Minimize burden for participants. We did so by streamlining time periods (e.g., we adapted the Perceived Stress Scale (Cohen & Williamson, 1988) from the last 4 weeks to the last 2 weeks, so it is aligned with all our other measures that capture 2 weeks); by shortening scales to remove items not of interest to our research; and by shortening repetitive instructions (e.g., many scales instruct participants to “read these items carefully”).
3. Adapt measures to ensure they are adequate for most participants in a student sample in the Netherlands. Three examples are: we changed the unit “stone” to “kilogram” in the SCOFF scale (Morgan et al., 1999); we removed the item “combat or exposure to a war-zone” from the Life Events Checklist for DSM-5 (Gray et al., 2004); and we replaced the examples “gardening”, “collecting”, and “sewing” with “playing computer games” in the leisure domain of the Work and Social Adjustment Scale (Mundt et al., 2002).

Because our selection of constructs may miss important aspects of participants’ lives, every stage affords participants the opportunity to indicate further relevant information in open text fields. Overall, we have made all questionnaires and codebooks for all measures available in the Supplementary Materials.

Stage 1: Baseline

Stage 1 consists of a 75-minute Qualtrics survey to collect research data. Table 1 contains an overview of our measurement battery, resulting from a detailed literature review and several expert meetings, followed by a short Delphi study with 12 clinicians and researchers from clinical psychology and psychiatry.

In addition to this survey, we ask participants for permission to link their postal code to Dutch registry data containing neighborhood information such as air pollution, green spaces, and traffic noise (see Table 1); such data may be helpful as indicators for socioeconomic status, which in turn has been shown to be related to depression (Platania, 2023). More information about registry data is available at gecco.nl; permission to link postal code to registry data is not necessary for participation in WARN-D.

Stage 2: Daily Monitoring

Stage 2 aims to provide a detailed mapping of the biopsychosocial components of the human mood system. This includes the temporal dynamics of important variables like depression and anxiety symptoms, affect states, stress, functional impairment, activity, sleep, as well as contextual variables.

To assess these data, we use ecological momentary assessment (EMA) to follow participants in their daily lives via smartphones for 85 days (Bos et al., 2019; Larson & Csikszentmihalyi, 1983; Myin-Germeys & Kuppens, 2021). Specifically, we use the Ethica app to query people 4 times a day, between around 10 am and 9:30pm at intervals of around 225 minutes with a normally distributed 30-minute jitter for each survey; each

Table 1

Stage 1 (Baseline) Measurements in the WARN-D Study

Category	**Examples**
Demographics	Age, nationality, population group
Physical appearance	Height, weight and satisfaction therewith, satisfaction physical appearance
Sex and gender	Biological sex, gender identity and struggles, sexual orientation and struggles
Internationality	Time spent in the Netherlands, integration into Dutch Society, international student status
SES and finances	Subjective socioeconomic status, current work, income sources, income vs spending, satisfaction work and finances, parents' and own highest education
Education	Current studies and satisfaction, academic standing and satisfaction
Living situation	Children, household composition, satisfaction living situation
Religion	Religious affiliation, connection to church, place of worship
Physical health	Global physical health and impairment rating last 2 weeks and last year, chronic health issues, pain, medication
Menstruation-related questions	Detailed menstruation information, pregnancy / breastfeeding, contraception
COVID-19	Impact of pandemic on mental health, prior COVID-19 diagnoses, COVID-19 symptom severity, long COVID-19 symptoms
Sleep habits	Chronotype, sleep schedule, sleep problems like nightmares, worry about sleep, impairment, satisfaction
Mental health	Family history, global mental health and impairment rating last 2 weeks and last year, lifetime emotional problems, current / prior problems and diagnoses, recent changes in mental health, current need for treatment, current / prior treatment, current and lifetime depression, current seasonal affective disorder / (hypo)mania / generalized anxiety disorder / social anxiety disorder / obsessive-compulsive disorder / eating disorder / borderline personality disorder, current and past suicidal ideation, prior suicide attempts, non-suicidal self-injury
Substance use	Current, past, and lifetime substance use problems
Wellbeing and stressors	Hedonic and eudaemonic wellbeing, general life satisfaction, current stress and stress domains, childhood and lifetime adversity, discrimination, bullying, feelings of safety, negative and positive life events
Social	Social network online / offline, social media use, positive / negative interpersonal experiences, satisfaction relationship with friends / family, relationship status and satisfaction, satisfaction sex life, satisfaction independence from parents, loneliness
Leisure and activity	Physical activity, sedentary behavior, time spent outside, leisure activities and satisfaction
Traits and tendencies	Attachment style, negative affect, big five personality traits, repetitive negative thinking, intolerance to uncertainty, pessimism, behavioral and cognitive emotional regulation strategies, affective lability, anger/irritability, perfectionism, workaholism, dependency/separation anxiety/insecurity, procrastination
Resilience	Perceived stress recovery, self-efficacy, self-esteem, locus of control
Meta	Motivation to participate, survey difficulty, attention paid while answering, feedback on survey
Registry data	Air pollution, educational facilities, green spaces, income, urbanization, traffic noise, poverty, value of houses

Note. For a full list of variables, phrasing, response options, translations, and bibliography of measurement instruments, see codebook in the Supplementary Materials.

survey expires after 20 minutes. All 4 surveys contain the same block of 18 questions and take about 1-2 minutes to complete. The morning survey contains 3 additional questions about the last night and outlook for the day, and the evening survey contains 18 additional questions about the day as a whole. In addition, we query people every Sunday at noon for a 46-item survey that takes around 5-7 minutes to complete, expiring after 10 hours. Table 2 summarizes EMA measurement design and content. An example item is “How sad are you right now”, which we query using a 7-point Likert scale from 1 (not at all) to 7 (very much). Our measures are based on the literature, our prior work, currently ongoing projects, and discussions with EMA experts. The number of prompts and items per prompt were chosen based on discussions with numerous experts as well as our own experience regarding the compliance rates of EMA data in student populations, with the overarching goal to obtain insightful momentary data whilst ensuring that the EMA protocol is feasible for students; for that reason, we also assess if participants experience the monitoring as burdensome.

Table 2

Stage 2 Measurements in the WARN-D Study

Category	Examples
Mental health	Stress and stress domains, mental health and interference with daily activities, depression and anxiety symptoms, bad dreams, non-suicidal self-injury
Positive and negative affect	Happy/cheerful, motivated, relaxed, stressed, sad, nervous/anxious, overwhelmed, annoyed/irritated
Satisfaction and wellbeing	Ability to concentrate, feeling productive, general satisfaction
Physical experiences	Physical health and interference with daily activities, pain/discomfort, sleep, substance use, menstruation, sleep and tiredness
Experiences	Best and worst experiences of the day and the week, category of experiences such as finances, education, and love life
Social experiences	Feeling connected to others, being able to rely on others for support, current social offline/online contact, social media use
Context	Current activity and enjoyment of activity, current location
Coping and Appraisal	Being able to handle daily and weekly challenges, emotion regulation
Meta	Enjoying study participation, reasons for missing surveys
Garmin smartwatch	Heart rate (constant, daily resting), blood oxygen saturation monitor, energy monitor, stress, body battery, sleep, step counter

Note. For a full list of variables, phrasing, response options, and translations, see codebook in the Supplementary Materials.

We also collect digital phenotype data via the Garmin smartwatch Vivosmart 4, including sleep phases and duration, activity, heart rate, and stress.

Stage 3: Follow-Up Surveys

Stage 3 consists of 8 follow-up surveys. Two of these (the yearly ones) last ~30 minutes, the others ~15 minutes; see Table 3 for an overview of the assessed constructs.

Table 3

Stage 3 (Follow-Up) Measurements in the WARN-D Study

Category	Variables
Physical appearance	Height, weight and satisfaction therewith, satisfaction physical appearance
Sex and gender	Struggles with gender identify / sexual orientation
Internationality	Integration into Dutch Society
SES and finances	Satisfaction work and finances, highest education
Education	Current studies and satisfaction
Living situation	Children, satisfaction living situation
Physical health	Global physical health and impairment rating last 2 weeks and last 3 months, medication
Menstruation-related questions	Pregnancy/breastfeeding, contraception
COVID-19	Impact of pandemic on mental health, prior COVID-19 diagnoses, COVID-19 symptom severity, long COVID-19 symptoms
Sleep habits	Sleep, nightmares, satisfaction
Mental health	Global mental health and impairment rating last 2 weeks and last 3 months, current emotional problems and diagnoses, recent changes in mental health, current need for treatment, current treatment, current depression, current generalized anxiety disorder, current/prior suicidal ideation, suicide attempts, non-suicidal self-injury
Substance use	Current substance use habits
Wellbeing and stressors	Hedonic and eudaemonic wellbeing, current stress and stress domains, negative and positive life events
Social	Social network online / offline, social media use, positive / negative interpersonal experiences, satisfaction relationship with friends / family, relationship status & satisfaction, satisfaction sex life, satisfaction independence from parents, loneliness
Leisure and activity	Physical activity, satisfaction leisure activities
Traits and tendencies	Neuroticism, behavioral and cognitive emotional regulation strategies, affective lability, perfectionism, dependency/separation anxiety/insecurity, procrastination
Resilience	Perceived stress recovery, perceived recent resilience, forecast resilience
Meta	Motivation to continue participation, feedback on survey

Note. For a full list of variables, phrasing, response options, translations, and bibliography of measurement instruments, see codebook in the Supplementary Materials.

Stage 4: Repetition of Stage 2

Stage 4 is a repetition of Stage 2. We aim to recruit ~500 participants from Cohorts 1 and 2 who previously completed Stage 2 to obtain insight into the temporal stability of the human mood system.

Strengths and Challenges

We hope to achieve our ambitious goal of building a personalized early warning system for depression by embracing the complexity of the human mood system (Fried, 2022; Olthof et al., 2023). A multi-disciplinary approach integrating advances from systems theory, multi-modal measurement, and statistical models will be crucial to achieve this aim. The project also faces several challenges, and many open questions remain.

First, there is a large literature on EWS in other disciplines such as ecology (Dakos et al., 2012; Scheffer et al., 2012), and the psychological literature is growing rapidly (e.g., Adler et al., 2020; Cabrieto et al., 2019; Olthof et al., 2020). Which EWS may be predictive of depression remains to be seen, and we will focus on both data-driven and theory-driven EWS by leveraging all collected data and using machine-learning models to analyze what particular features are predictive of an upcoming transition, but also by testing various EWS proposed in the literature. One is critical slowing down, a feature that has been shown to predict transitions in systems such as lakes before they turn from clean to turbid states, non-linear physical systems such as Earth's climate, as well as the stock market (Olde Rikkert et al., 2016; Quax et al., 2018; Scheffer et al., 2012, 2018). Slowing down is a marker that a system becomes more vulnerable for an upcoming transition, because vulnerable systems take longer to recover from perturbations, which goes together with changes in parameters of systems that can be observed. Another EWS is higher connectivity, defined as more and stronger relations among components in a system, which could confer vulnerability for future depression. This is because in a more strongly connected causal system of problems (e.g., sleep problems, sad mood, concentration problems, fatigue), activating one problem may lead to a cascade that activates others (Cramer et al., 2016; Schweren et al., 2018; van Borkulo et al., 2015; van de Leemput et al., 2014).

A second challenge is that WARN-D is focused on forecasting depression, a complex and fuzzy phenotype for which many defensible operationalizations exist. For this reason, we will predict several outcome variables, rather than restricting ourselves to one arbitrary operationalization. Outcomes include: stress, anxiety and depression severity, as well as probable MDD diagnosis; wellbeing and impairment of functioning; changes in any of these constructs over time (as observed by longitudinal data), as well as perceived changes in these outcomes over time as retrospectively reported by participants. Further outcomes include self-report information participants provide on diagnoses by health care professionals, as well as starting psychological or pharmacological treatments for MDD or related conditions. A robust predictor is one that predicts a larger number of these operationalizations of significant mental health changes.

Third, attrition rates are a concern in EMA studies and prospective studies. To mitigate attrition, we incentivize participants in various ways: we pay them per completed survey (up to 90€ in total); organize 500€ lotteries per cohort for completing the 1-year and 2-year surveys, respectively; provide participants with Garmin VivoSmart 4 smartwatches they can use freely in Stage 2; and offer participants a personalized data report of their EMA data after completing Stage 2. We also continuously ask participants about their experience with WARN-D to learn about participation barriers with the goal of minimizing attrition rates in future cohorts.

Finally, WARN-D is an observational study, since our primary goal is prediction of onset—interventions carried out by WARN-D itself would stand in the way of accurate forecasting. However, one could argue that especially Stage 2 (tracking people via smartphones and smartwatches) may itself be an intervention. Fortunately, we hope that such effects are held constant across time. That is, after we have developed the WARN-D app in a few years to predict onset, people using it will go through a very similar program as described here, tracking themselves via smartphones and smartwatches to collect data to enable prediction of future onset. The app will likely also support a functionality where users can view the data they provide, similar to the personalized data reports. In that sense, our observational validation cohort for WARN-D, and the people using the app in the future, will receive similar self-tracking 'interventions', holding potential intervention effects constant in our prediction and validation samples.

Funding: Eiko Fried, Ricarda Proppert, and Carlotta Rieble are supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program, Grant No. 949059.

Acknowledgments: We wholeheartedly thank all participants of WARN-D; all prior and current WARN-D team members; all members of the scientific advisory board of WARN-D; and all experts who participated in the Delphi study to inform our baseline measurement battery.

Competing Interests: The authors have declared that no competing interests exist.

Author Contributions: Conceptualization, funding acquisition, project administration, visualization, writing – original draft: E.I.F. Data curation, investigation, software, supervision, validation, writing – review & editing: E.I.F., R.P., and C.L.R.

Twitter Accounts: @eikofried, @RProppert, @carlottarieble

Data Availability: We are excited to make our design, measures, code, and data available to the research community. Information on design and measures are available in the accompanying Supplementary Materials (Fried, Proppert, & Rieble, 2023a). All empirical papers will be accompanied by open code; we are currently developing a data sharing protocol with all relevant stakeholders which will be ready in 2025/2026 by the time data collection is finished. See our WARN-D project hub (Fried, Proppert, & Rieble, 2023b) for all future updates and publications, and, eventually, data.

Supplementary Materials

Supplementary Materials are available online (Fried, Proppert, & Rieble, 2023a), which contain further information regarding:

1. Consent sheets and general information sheets for participants, inclusion and exclusion screener, codebook for the screener
2. Design and procedures for all stages

3. All questionnaires and codebooks in Dutch and English for all stages
4. Stage 1: data journey for participant data, mental health information package for participants, information on the delphi study, information on the registry data
5. Stage 2: participant instruction materials and video for setting up smartphones and watches
6. Procedures for the personalized data reports and an example report

Index of Supplementary Materials

Fried, E. I., Proppert, R. K. K., & Rieble, C. L. (2023a). *Supplementary materials to "Building an early warning system for depression: Rationale, objectives, and methods of the WARN-D study"* [Additional information]. OSF. https://doi.org/10.17605/OSF.IO/2JD9H

Fried, E. I., Proppert, R. K. K., & Rieble, C. L. (2023b). *Supplementary materials to "Building an early warning system for depression: Rationale, objectives, and methods of the WARN-D study"* [Project hub]. OSF. https://doi.org/10.17605/OSF.IO/FRQDV

References

Adler, D. A., Ben-Zeev, D., Tseng, V. W.-S., Kane, J. M., Brian, R., Campbell, A. T., Hauser, M., Scherer, E. A., & Choudhury, T. (2020). Predicting early warning signs of psychotic relapse from passive sensing data: An approach using encoder-decoder neural networks. *JMIR mHealth and uHealth, 8*(8), Article e19962. https://doi.org/10.2196/19962

Altman, E. G., Hedeker, D., Peterson, J. L., & Davis, J. M. (1997). The Altman Self-Rating Mania Scale. *Biological Psychiatry, 42*(10), 948–955. https://doi.org/10.1016/S0006-3223(96)00548-3

Arias-de la Torre, J., Vilagut, G., Ronaldson, A., Serrano-Blanco, A., Martín, V., Peters, M., Valderas, J. M., Dregan, A., & Alonso, J. (2021). Prevalence and variability of current depressive disorder in 27 European countries: A population-based study. *The Lancet: Public Health, 6*(10), e729–e738. https://doi.org/10.1016/S2468-2667(21)00047-5

Auerbach, R. P., Alonso, J., Axinn, W. G., Cuijpers, P., Ebert, D. D., Green, J. G., Hwang, I., Kessler, R. C., Liu, H., Mortier, P., Nock, M. K., Pinder-Amaker, S., Sampson, N. A., Aguilar-Gaxiola, S., Al-Hamzawi, A., Andrade, L. H., Benjet, C., Caldas-De-Almeida, J. M., Demyttenaere, K., . . . Bruffaerts, R. (2016). Mental disorders among college students in the World Health Organization World Mental Health Surveys. *Psychological Medicine, 46*(14), 2955–2970. https://doi.org/10.1017/S0033291716001665

Auerbach, R. P., Mortier, P., Bruffaerts, R., Alonso, J., Benjet, C., Cuijpers, P., Demyttenaere, K., Ebert, D., Green, J., Murray, E., Nock, M., Pinder-Amaker, S., Sampson, N., Stein, D., Vilagut, G., Zaslavsky, A., & Kessler, R. (2018). WHO World Mental Health Surveys International College Student Project: Prevalence and Distribution of Mental Disorders. *Journal of Abnormal Psychology, 127*(7), 623–638. https://doi.org/10.1037/abn0000362

Barabási, A.-L. (2012). The network takeover. *Nature Physics, 8*(1), 14–16. https://doi.org/10.1038/nphys2188

Beck, A. T., Kovacs, M., & Weissman, A. (1979). Assessment of suicidal intention: The Scale for Suicide Ideation. *Journal of Consulting and Clinical Psychology, 47*(2), 343–352. https://doi.org/10.1037/0022-006X.47.2.343

Berman, A. L. (2009). Depression and suicide. In I. H. Gotlib & C. L. Hammen (Eds.), *Handbook of depression* (2nd ed., pp. 510–530). Guilford Press.

Biemar, F., & Foti, M. (2013). Global progress against cancer-challenges and opportunities. *Cancer Biology & Medicine, 10*(4), 183–186. https://doi.org/10.7497/j.issn.2095-3941.2013.04.001

Border, R., Johnson, E. C., Evans, L. M., Smolen, A., Berley, N., Sullivan, P. F., & Keller, M. C. (2019). No support for historical candidate gene or candidate gene-by-interaction hypotheses for major depression across multiple large samples. *The American Journal of Psychiatry, 176*(5), 376–387. https://doi.org/10.1176/appi.ajp.2018.18070881

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA), 16*(1), 5–13. https://doi.org/10.1002/wps.20375

Bos, F. M., Snippe, E., Bruggeman, R., Wichers, M., & van der Krieke, L. (2019). Insights of patients and clinicians on the promise of the experience sampling method for psychiatric care. *Psychiatric Services, 70*(11), 983–991. https://doi.org/10.1176/appi.ps.201900050

Boschloo, L., Bekhuis, E., Weitz, E. S., Reijnders, M., DeRubeis, R. J., Dimidjian, S., Dunner, D. L., Dunlop, B. W., Hegerl, U., Hollon, S. D., Jarrett, R. B., Kennedy, S. H., Miranda, J., Mohr, D. C., Simons, A. D., Parker, G., Petrak, F., Herpertz, S., Quilty, L. C., . . . Cuijpers, P. (2019). The symptom-specific efficacy of antidepressant medication vs. cognitive behavioral therapy in the treatment of depression: Results from an individual patient data meta-analysis. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA), 18*(2), 183–191. https://doi.org/10.1002/wps.20630

Boschloo, L., Cuijpers, P., Karyotaki, E., Berger, T., Moritz, S., Meyer, B., & Klein, J. P. (2019). Symptom-specific effectiveness of an internet-based intervention in the treatment of mild to moderate depressive symptomatology: The potential of network estimation techniques. *Behaviour Research and Therapy, 122*(August), Article 103440. https://doi.org/10.1016/j.brat.2019.103440

Cabrieto, J., Adolf, J., Tuerlinckx, F., Kuppens, P., & Ceulemans, E. (2019). An objective, comprehensive and flexible statistical framework for detecting early warning signs of mental health problems. *Psychotherapy and Psychosomatics, 88*(3), 184–186. https://doi.org/10.1159/000494356

Cai, N., Bigdeli, T. B., Kretzschmar, W., Li, Y., Liang, J., Song, L., Hu, J., Li, Q., Jin, W., Hu, Z., Wang, G., Wang, L., Qian, P., Liu, Y., Jiang, T., Lu, Y., Zhang, X., Yin, Y., Li, Y., . . . Flint, J. (2015). Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature, 523*(7562), 588–591. https://doi.org/10.1038/nature14659

Cai, N., Choi, K. W., & Fried, E. I. (2020). Reviewing the genetics of heterogeneity in depression: Operationalizations, manifestations, and etiologies. *Human Molecular Genetics, 29*(R1), R10–R18. https://doi.org/10.1093/hmg/ddaa115

Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of Modern Physics, 81*(2), 591–646. https://doi.org/10.1103/RevModPhys.81.591

Cohen, S., & Williamson, G. M. (1988). Perceived stress in a probability sample of the United States. In S. Spacapan & S. Oskamp (Eds.), *The social psychology of health: Claremont symposium on applied social psychology* (pp. 31–67). Sage.

Conway, C. C., Forbes, M., Forbush, K. T., Fried, E. I., Hallquist, M. N., Kotov, R., Mullins-Sweatt, S. N., Shackman, A. J., Skodol, A. E., South, S. C., Sunderland, M., Waszczuk, M. A., Zald, D. H., Afzali, M. H., Pincus, A. L., Tackett, J. L., & Reininghaus, U. (2019). A hierarchical taxonomy of psychopathology can transform mental health research. *Perspectives on Psychological Science, 14*(3), 419–436. https://doi.org/10.1177/1745691618810696

Cramer, A. O. J., van Borkulo, C. D., Giltay, E. J., van der Maas, H. L. J., Kendler, K. S., Scheffer, M., & Borsboom, D. (2016). Major depression as a complex dynamic system. *PLoS One, 11*(12), Article e0167490. https://doi.org/10.1371/journal.pone.0167490

Cuijpers, P., Beekman, A. T. F., & Reynolds, C. F., III. (2012). Preventing depression: A global priority. *JAMA, 307*(10), 1033–1034.

Cuijpers, P., Karyotaki, E., Reijnders, M., & Huibers, M. J. H. (2018). Who benefits from psychotherapies for adult depression? A meta-analytic update of the evidence meta-analytic update of the evidence. *Cognitive Behaviour Therapy, 47*(2), 91–106. https://doi.org/10.1080/16506073.2017.1420098

Cuijpers, P., & Smit, F. (2004). Subthreshold depression as a risk indicator for major depressive disorder: A systematic review of prospective studies. *Acta Psychiatrica Scandinavica, 109*(5), 325–331. https://doi.org/10.1111/j.1600-0447.2004.00301.x

Dakos, V., Carpenter, S. R., Brock, W. A., Ellison, A. M., Guttal, V., Ives, A. R., Kéfi, S., Livina, V., Seekell, D. A., van Nes, E. H., & Scheffer, M. (2012). Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. *PLoS One, 7*(7), Article 41010. https://doi.org/10.1371/journal.pone.0041010

De Beurs, D. P., de Vries, A. L., de Groot, M. H., de Keijser, J., & Kerkhof, A. J. (2014). Applying computer adaptive testing to optimize online assessment of suicidal behavior: A simulation study. *Journal of Medical Internet Research, 16*(9), Article e3511. https://doi.org/10.2196/jmir.3511

Ebert, D. D., Buntrock, C., Mortier, P., Auerbach, R. P., Weisel, K. K., Kessler, R. C., Cuijpers, P., Green, J. G., Kiekens, G., Nock, M. K., Demyttenaere, K., & Bruffaerts, R. (2019). Prediction of major depressive disorder onset in college students. *Depression and Anxiety, 36*(4), 294–304. https://doi.org/10.1002/da.22867

Engel, G. L. (1977). The need for a new medical model: A challenge for biomedicine. *Science, 196*(4286), 129–136. https://doi.org/10.1126/science.847460

Ethica. (2023). *Patient-centric Clinical Trial and Health Research Software Platform.* https://ethicadata.com

Fava, G. A., & Tossani, E. (2007). Prodromal stage of major depression. *Early Intervention in Psychiatry, 1*(1), 9–18. https://doi.org/10.1111/j.1751-7893.2007.00005.x

Ferrari, A. J., Somerville, A. J., Baxter, A. J., Norman, R., Patten, S. B., Vos, T., & Whiteford, H. A. (2013). Global variation in the prevalence and incidence of major depressive disorder: A systematic review of the epidemiological literature. *Psychological Medicine, 43*(3), 471–481. https://doi.org/10.1017/S0033291712001511

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences of the United States of America, 115*(27), E6106–E6115. https://doi.org/10.1073/pnas.1711978115

Frank, P., Jokela, M., Batty, G. D., Cadar, D., Steptoe, A., & Kivimäki, M. (2021). Association between systemic inflammation and individual symptoms of depression: A pooled analysis of 15 population-based cohort studies. *The American Journal of Psychiatry, 178*(12), 1107–1118. https://doi.org/10.1176/appi.ajp.2021.20121776

Fried, E. I. (2022). Studying mental health problems as systems, not syndromes. *Current Directions in Psychological Science, 31*(6), 500–508. https://doi.org/10.1177/09637214221114089

Fried, E. I., Bockting, C. L. H., Arjadi, R., Borsboom, D., Amshoff, M., Cramer, A. O. J., Epskamp, S., Tuerlinckx, F., Carr, D., & Stroebe, M. (2015). From loss to loneliness: The relationship between bereavement and depressive symptoms. *Journal of Abnormal Psychology, 124*(2), 256–265. https://doi.org/10.1037/abn0000028

Fried, E. I., Flake, J. K., & Robinaugh, D. J. (2022). Revisiting the theoretical and methodological foundations of depression measurement. *Nature Reviews Psychology, 1*, 358–368. https://doi.org/10.1038/s44159-022-00050-2

Fried, E. I., & Nesse, R. M. (2014). The impact of individual depressive symptoms on impairment of psychosocial functioning. *PLoS One, 9*(2), Article e90311. https://doi.org/10.1371/journal.pone.0090311

Fried, E. I., & Nesse, R. M. (2015a). Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study. *Journal of Affective Disorders, 172*, 96–102. https://doi.org/10.1016/j.jad.2014.10.010

Fried, E. I., & Nesse, R. M. (2015b). Depression sum-scores don’t add up: Why analyzing specific depression symptoms is essential. *BMC Medicine, 13*, Article 72. https://doi.org/10.1186/s12916-015-0325-4

Fried, E. I., Papanikolaou, F., & Epskamp, S. (2022). Mental health and social contact during the COVID-19 pandemic: An ecological momentary assessment study. *Clinical Psychological Science, 10*(2), 340–354. https://doi.org/10.1177/21677026211017839

Fried, E. I., van Borkulo, C. D., Epskamp, S., Schoevers, R. A., Tuerlinckx, F., & Borsboom, D. (2016). Measuring depression over time. . . Or not? Lack of unidimensionality and longitudinal measurement invariance in four common rating scales of depression. *Psychological Assessment, 28*(11), 1354–1367. https://doi.org/10.1037/pas0000275

Fried, E. I., von Stockert, S., Haslbeck, J. M. B., Lamers, F., Schoevers, A., & Penninx, B. W. J. H. (2020). Using network analysis to examine links between individual depressive symptoms, inflammatory markers, and covariates. *Psychological Medicine, 50*(16), 2682–2690. https://doi.org/10.1017/S0033291719002770

Gotlib, I. H., Lewinsohn, P. M., & Seeley, J. R. (1995). Symptoms versus a diagnosis of depression: Differences in psychosocial functioning. *Journal of Consulting and Clinical Psychology, 63*(1), 90–100. https://doi.org/10.1037/0022-006X.63.1.90

Gray, M. J., Litz, B. T., Hsu, J. L., & Lombardo, T. W. (2004). Psychometric properties of the Life Events Checklist. *Assessment, 11*(4), 330–341. https://doi.org/10.1177/1073191104269954

Haslam, N. (2003). Categorical versus dimensional models of mental disorder: The taxometric evidence. *The Australian and New Zealand Journal of Psychiatry, 37*(6), 696–704. https://doi.org/10.1080/j.1440-1614.2003.01258.x

Haslam, N., Holland, E., & Kuppens, P. (2012). Categories versus dimensions in personality and psychopathology: A quantitative review of taxometric research. *Psychological Medicine, 42*(5), 903–920. https://doi.org/10.1017/S0033291711001966

Hayes, A. M., & Andrews, L. A. (2020). A complex systems approach to the study of change in psychotherapy. *BMC Medicine, 18*, Article 197. https://doi.org/10.1186/s12916-020-01662-2

Hetrick, S. E., Parker, A. G., Hickie, I. B., Purcell, R., Yung, A. R., & McGorry, P. D. (2008). Early identification and intervention in depressive disorders: Towards a clinical staging model. *Psychotherapy and Psychosomatics, 77*(5), 263–270. https://doi.org/10.1159/000140085

Hilland, E., Landrø, N. I., Kraft, B., Tamnes, C. K., Fried, E. I., Maglanoc, L. A., & Jonassen, R. (2020). Exploring the links between specific depression symptoms and brain structure: A network study. *Psychiatry and Clinical Neurosciences, 74*(3), 220–221. https://doi.org/10.1111/pcn.12969

Hitchcock, P. F., Fried, E. I., & Frank, M. J. (2022). Computational psychiatry needs time and context. *Annual Review of Psychology, 73*(1), 243–270. https://doi.org/10.1146/annurev-psych-021621-124910

Iacoviello, B., Alloy, L., Abramson, L., & Choi, J. (2010). The early course of depression: A longitudinal investigation of prodromal symptoms and their relation to the symptomatic course of depressive episodes. *Journal of Abnormal Psychology, 119*(3), 459–467. https://doi.org/10.1037/a0020114

James, S. L., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A., Abdollahpour, I., Abdulkader, R. S., Abebe, Z., Abera, S. F., Abil, O. Z., Abraha, H. N., Abu-Raddad, L. J., Abu-Rmeileh, N. M. E., Accrombessi, M. M. K., . . . Murray, C. J. L. (2018). Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *Lancet, 392*(10159), 1789–1858. https://doi.org/10.1016/S0140-6736(18)32279-7

Judd, L. L., Akiskal, H. S., Maser, J. D., Zeller, P. J., Endicott, J., Coryell, W., Paulus, M. P., Kunovac, J. L., Leon, A. C., Mueller, T. I., Rice, J. A., & Keller, M. B. (1998). A prospective 12-year study of subsyndromal and syndromal depressive symptoms in unipolar major depressive disorders. *Archives of General Psychiatry, 55*(8), 694–700. https://doi.org/10.1001/archpsyc.55.8.694

Judd, L. L., Akiskal, H. S., & Paulus, M. P. (1997). The role and clinical significance of subsyndromal depressive symptoms (SSD) in unipolar major depressive disorder. *Journal of Affective Disorders, 45*(1–2), 5–18. https://doi.org/10.1016/S0165-0327(97)00055-4

Kapur, S., Phillips, A. G., & Insel, T. (2012). Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Molecular Psychiatry, 17*(12), 1174–1179. https://doi.org/10.1038/mp.2012.105

Keller, M. C., Neale, M. C., & Kendler, K. (2007). Association of different adverse life events with distinct patterns of depressive symptoms. *The American Journal of Psychiatry, 164*(10), 1521–1529. https://doi.org/10.1176/appi.ajp.2007.06091564

Keller, M. C., & Nesse, R. M. (2005). Is low mood an adaptation? Evidence for subtypes with symptoms that match precipitants. *Journal of Affective Disorders, 86*(1), 27–35. https://doi.org/10.1016/j.jad.2004.12.005

Kendler, K. S. (2012a). The dappled nature of causes of psychiatric illness: Replacing the organic-functional/hardware-software dichotomy with empirically based pluralism. *Molecular Psychiatry, 17*(4), 377–388. https://doi.org/10.1038/mp.2011.182

Kendler, K. S. (2012b). The dappled nature of causes of psychiatric illness: Replacing the organic-functional/hardware-software dichotomy with empirically based pluralism. *Molecular Psychiatry, 17*(4), 377–388. https://doi.org/10.1038/mp.2011.182

Kennis, M., Gerritsen, L., van Dalen, M., Williams, A., Cuijpers, P., & Bockting, C. (2020). Prospective biomarkers of major depressive disorder: A systematic review and meta-analysis. *Molecular Psychiatry, 25*, 321–338. https://doi.org/10.1038/s41380-019-0585-z

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime Prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey replication. *Archives of General Psychiatry, 62*(6), 593–602. https://doi.org/10.1001/archpsyc.62.6.593

Kessler, R. C., Chiu, W. T., Demler, O., Merikangas, K. R., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry, 62*(6), 617–627. https://doi.org/10.1001/archpsyc.62.6.617

Khan, A., & Brown, W. A. (2015). Antidepressants versus placebo in major depression: An overview. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA), 14*(3), 294–300. https://doi.org/10.1002/wps.20241

Kroenke, K., & Spitzer, R. L. (2013). The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric Annals, 32*(9), 509–515. https://doi.org/10.3928/0048-5713-20020901-06

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2003). The Patient Health Questionnaire-2: Validity of a two-item depression screener. *Medical Care, 41*(11), 1284–1292. https://doi.org/10.1097/01.MLR.0000093487.78664.3C

Kuppens, P. (2015). It's about time: A Special Section on affect dynamics. *Emotion Review, 7*(4), 297–300. https://doi.org/10.1177/1754073915590947

Larson, R., & Csikszentmihalyi, M. (1983). The experience sampling method. *New Directions for Methodology of Social & Behavioral Science, 15*, 41–56.

Lopez, A. D., Mathers, C. D., Ezzati, M., Jamison, D. T., & Murray, C. J. L. (2006). *Global Burden of Disease and Risk Factors*. World Bank.

Luke, D. A., & Stamatakis, K. A. (2012). Systems science methods in public health: Dynamics, networks, and agents. *Annual Review of Public Health, 33*(1), 357–376. https://doi.org/10.1146/annurev-publhealth-031210-101222

Lutz, W., Schwartz, B., Hofmann, S. G., Fisher, A. J., Husen, K., & Rubel, J. A. (2018). Using network analysis for the prediction of treatment dropout in patients with mood and anxiety disorders: A methodological proof-of-concept study. *Scientific Reports, 8*(1), Article 7819. https://doi.org/10.1038/s41598-018-25953-0

Mathers, C. D., & Loncar, D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Medicine, 3*(11), Article e442. https://doi.org/10.1371/journal.pmed.0030442

McWilliams, N. (2021). Diagnosis and its discontents: Reflections on our current dilemma. *Psychoanalytic Inquiry, 41*(8), 565–579. https://doi.org/10.1080/07351690.2021.1983395

Morgan, J. F., Reid, F., & Lacey, J. H. (1999). The SCOFF questionnaire: Assessment of a new screening tool for eating disorders. *BMJ, 319*(7223), 1467–1468. https://doi.org/10.1136/bmj.319.7223.1467

Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. H. (2002). The Work and Social Adjustment Scale: A simple measure of impairment in functioning. *The British Journal of Psychiatry, 180*(5), 461–464. https://doi.org/10.1192/bjp.180.5.461

Muñoz, R. F., Cuijpers, P., Smit, F., Barrera, A. Z., & Leykin, Y. (2010). Prevention of major depression. *Annual Review of Clinical Psychology, 6*(1), 181–212. https://doi.org/10.1146/annurev-clinpsy-033109-132040

Murphy, J. M., Nierenberg, A. A., Monson, R. R., Laird, N. M., Sobol, A. M., & Leighton, A. H. (2002). Self-disparagement as feature and forerunner of depression: Findings from the Stirling County Study. *Comprehensive Psychiatry, 43*(1), 13–21. https://doi.org/10.1053/comp.2002.29842

Murray, C., & Lopez, A. (1996). *Global burden of disease: A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020.* Harvard School of Public Health.

Myin-Germeys, I., & Kuppens, P. (2021). *The open handbook of experience sampling methodology: A step-by-step guide to designing, conducting, and analyzing ESM studies.* Center for Research on Experience Sampling and Ambulatory Methods.

Nagel, M., Watanabe, K., Stringer, S., Posthuma, D., & Van Der Sluis, S. (2018). Item-level analyses reveal genetic heterogeneity in neuroticism. *Nature Communications, 9*(1), Article 905. https://doi.org/10.1038/s41467-018-03242-8

Narrow, W. E., Clarke, D. E., Kuramoto, S. J., Kraemer, H. C., Kupfer, D. J., Greiner, L., & Regier, D. A. (2013). DSM-5 field trials in the United States and Canada, Part III: Development and reliability testing of a cross-cutting symptom assessment for DSM-5. *The American Journal of Psychiatry, 170*(1), 71–82. https://doi.org/10.1176/appi.ajp.2012.12071000

Nock, M. K. (2007). Conceptual and design essentials for evaluating mechanisms of change. *Alcoholism, Clinical and Experimental Research, 31*(s3), 4s–12s. https://doi.org/10.1111/j.1530-0277.2007.00488.x

Olde Rikkert, M. G. M., Dakos, V., Buchman, T. G., de Boer, R., Glass, L., Cramer, A. O. J., Levin, S., van Nes, E., Sugihara, G., Ferrari, M. D., Tolner, E. A., van de Leemput, I., Lagro, J., Melis, R., & Scheffer, M. (2016). Slowing down of recovery as generic risk marker for acute severity transitions in chronic diseases. *Critical Care Medicine, 44*(3), 601–606. https://doi.org/10.1097/CCM.0000000000001564

Olthof, M., Hasselman, F., Oude Maatman, F., Bosman, A. M. T., & Lichtwarck-Aschoff, A. (2023). Complexity theory of psychopathology. *Journal of Psychopathology and Clinical Science, 132*(3), 314–323. https://doi.org/10.1037/abn0000740

Olthof, M., Hasselman, F., Strunk, G., van Rooij, M., Aas, B., Helmich, M. A., Schiepek, G., & Lichtwarck-Aschoff, A. (2020). Critical fluctuations as an early-warning signal for sudden gains and losses in patients receiving psychotherapy for mood disorders. *Clinical Psychological Science, 8*(1), 25–35. https://doi.org/10.1177/2167702619865969

Parker, G. (2005). Beyond major depression. *Psychological Medicine, 35*(4), 467–474. https://doi.org/10.1017/S0033291704004210

Platania, N. (2023). *Examining links between individual depressive symptoms, indicators of socioeconomic status, and stressors* [Master's thesis, Leiden University]. https://studenttheses.universiteitleiden.nl/handle/1887/3439862

Quax, R., Chliamovitch, G., Dupuis, A., Falcone, J.-L., Chopard, B., Hoekstra, A. G., & Sloot, P. M. A. (2018). Information processing features can detect behavioral regimes of dynamical systems. *Complexity, 2018*, Article 6047846. https://doi.org/10.1155/2018/6047846

Robinaugh, D. J., Hoekstra, R. H. A., Toner, E. R., & Borsboom, D. (2020). The network approach to psychopathology: A review of the literature 2008–2018 and an agenda for future research. *Psychological Medicine, 50*(3), 353–366. https://doi.org/10.1017/S0033291719003404

Rogers, A. (2017). Star neuroscientist Tom Insel leaves the google-spawned Verily for … A startup? *WIRED*. https://www.wired.com/2017/05/star-neuroscientist-tom-insel-leaves-google-spawned-verily-startup

Scheffer, M., Bolhuis, J. E., Borsboom, D., Buchman, T. G., Gijzel, S. M. W., Goulson, D., Kammenga, J. E., Kemp, B., van de Leemput, I. A., Levin, S., Martin, C. M., Melis, R. J. F., van Nes, E. H., Romero, L. M., & Olde Rikkert, M. G. M. (2018). Quantifying resilience of humans and other animals. *Proceedings of the National Academy of Sciences of the United States of America, 115*(47), 11883–11890. https://doi.org/10.1073/pnas.1810630115

Scheffer, M., Carpenter, S. R., Lenton, T. M., Bascompte, J., Brock, W., Dakos, V., Van De Koppel, J., Van De Leemput, I. A., Levin, S. A., Van Nes, E. H., Pascual, M., & Vandermeer, J. (2012). Anticipating critical transitions. *Science, 338*(6105), 344–348. https://doi.org/10.1126/science.1225244

Schweren, L., van Borkulo, C. D., Fried, E. I., & Goodyer, I. M. (2018). Assessment of symptom network density as a prognostic marker of treatment response in adolescent depression. *JAMA Psychiatry, 75*(1), 98–100. https://doi.org/10.1001/jamapsychiatry.2017.3561

Snippe, E., Doornbos, B., Schoevers, R. A., & Wardenaar, K. J. (2021). Individual and common patterns in the order of symptom improvement during outpatient treatment for major depression. *Journal of Affective Disorders, 290*(February), 81–88. https://doi.org/10.1016/j.jad.2021.04.097

Solmi, M., Radua, J., Olivola, M., Croce, E., Soardo, L., Salazar de Pablo, G., Il Shin, J., Kirkbride, J. B., Jones, P., Kim, J. H., Kim, J. Y., Carvalho, A. F., Seeman, M. V., Correll, C. U., & Fusar-Poli, P. (2022). Age at onset of mental disorders worldwide: Large-scale meta-analysis of 192 epidemiological studies. *Molecular Psychiatry, 27*, 281–295. https://doi.org/10.1038/s41380-021-01161-7

Tweed, D. L. (1993). Depression-related impairment: Estimating concurrent and lingering effects. *Psychological Medicine, 23*(2), 373–386. https://doi.org/10.1017/S0033291700028476

van Borkulo, C. D., Boschloo, L., Borsboom, D., Penninx, B. W. J. H., Waldorp, L. J., & Schoevers, R. A. (2015). Association of symptom network structure with the course of longitudinal depression. *JAMA Psychiatry, 72*(12), 1219–1226. https://doi.org/10.1001/jamapsychiatry.2015.2079

van de Leemput, I. A., Wichers, M., Cramer, A. O. J., Borsboom, D., Tuerlinckx, F., Kuppens, P., van Nes, E. H., Viechtbauer, W., Giltay, E. J., Aggen, S. H., Derom, C., Jacobs, N., Kendler, K., van der Maas, H. L. J., Neale, M. C., Peeters, F., Thiery, E., Zachar, P., & Scheffer, M. (2014). Critical slowing down as early warning for the onset and termination of depression. *Proceedings of the National Academy of Sciences of the United States of America, 111*(1), 87–92. https://doi.org/10.1073/pnas.1312114110

Van Eeden, W. A., van Hemert, A. M., Carlier, I. V. E., Penninx, B. W. J. H., Lamers, F., Fried, E. I., Schoevers, R., & Giltay, E. J. (2020). Basal and LPS-stimulated inflammatory markers and the course of individual symptoms of depression. *Translational Psychiatry, 10*, Article 235. https://doi.org/10.1038/s41398-020-00920-4

van Zoonen, K., Buntrock, C., Ebert, D. D., Smit, F., Reynolds, C. F., Beekman, A. T. F., & Cuijpers, P. (2014). Preventing the onset of major depressive disorder: A meta-analytic review of psychological interventions. *International Journal of Epidemiology, 43*(2), 318–329. https://doi.org/10.1093/ije/dyt175

WHO ASSIST Working Group. (2010). The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): Development, reliability and feasibility. *Addiction, 97*(9), 1183–1194.

Wichers, M. (2014). The dynamic nature of depression: A new micro-level perspective of mental disorder that meets current challenges. *Psychological Medicine, 44*(7), 1349–1360. https://doi.org/10.1017/S0033291713001979

Wichers, M., Groot, P. C., Psychosystems, ESM Group, & EWS Group. (2016). Critical slowing down as a personalized early warning signal for depression. *Psychotherapy and Psychosomatics, 85*(2), 114–116. https://doi.org/10.1159/000441458

Winter, N. R., Leenings, R., Ernsting, J., Sarink, K., Fisch, L., Emden, D., Blanke, J., Goltermann, J., Opel, N., Barkhau, C., Meinert, S., Dohm, K., Repple, J., Mauritz, M., Gruber, M., Leehr, E. J., Grotegerd, D., Redlich, R., Jansen, A., . . . Hahn, T. (2022). More alike than different:

Quantifying deviations of brain structure and function in major depressive disorder across neuroimaging modalities. *JAMA Psychiatry, 79*(9), 879–888. https://doi.org/10.1001/jamapsychiatry.2022.1780

World Health Organization. (2019). *Depression (WHO Factsheet)*. World Health Organization. http://www.who.int/news-room/factsheets/ detail/depression

Zimmerman, M., Ellison, W., Young, D., Chelminski, I., & Dalrymple, K. (2015). How many different ways do patients meet the diagnostic criteria for major depressive disorder? *Comprehensive Psychiatry, 56*, 29–34. https://doi.org/10.1016/j.comppsych.2014.09.007

EACLIPT

Clinical Psychology in Europe (CPE) is the official journal of the European Association of Clinical Psychology and Psychological Treatment (EACLIPT).

leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.

Clinical Psychology in Europe

PsychOpen GOLD

Systematic Reviews and Meta-analyses

Aetiological Understanding of Fibromyalgia, Irritable Bowel Syndrome, Chronic Fatigue Syndrome and Classificatory Analogues: A Systematic Umbrella Review

Maria Kleinstäuber[1 §], Andreas Schröder[2 §], Sarah Daehler[1],
Karen Johanne Pallesen[3], Charlotte U. Rask[2,4], Mathias Sanyer[1],
Omer Van den Bergh[5], Marie Weinreich Petersen[2], Judith G. M. Rosmalen[6]

[1] *Department of Psychology, Emma Eccles Jones College of Education and Human Services, Utah State University, Logan, UT, USA.* **[2]** *The Research Clinic for Functional Disorders and Psychosomatics, Aarhus University Hospital, Aarhus, Denmark.* **[3]** *Danish Centre for Mindfulness, Aarhus University, Aarhus, Denmark.* **[4]** *Department of Child and Adolescent Psychiatry, Psychiatry, Aarhus University Hospital, Aarhus, Denmark.* **[5]** *Department of Health Psychology, University of Leuven, Leuven, Belgium.* **[6]** *University Medical Centre Groningen, University of Groningen, Groningen, The Netherlands.*

[§] *These authors contributed equally to this work.*

Clinical Psychology in Europe, 2023, Vol. 5(3), Article e11179, https://doi.org/10.32872/cpe.11179

Received: 2023-01-21 • **Accepted:** 2023-08-27 • **Published (VoR):** 2023-09-29

Handling Editor: Winfried Rief, Philipps-University of Marburg, Marburg, Germany

Corresponding Author: Maria Kleinstäuber, Department of Psychology, Emma Eccles Jones College of Education and Human Services, Utah State University, 2810 Old Main Hill, Logan, UT 84322, USA. Phone: +1 435 797 1236. E-mail: maria.kleinstaeuber@usu.edu

Supplementary Materials: Materials [see Index of Supplementary Materials]

Abstract

Background: This umbrella review systematically assesses the variety and relative dominance of current aetiological views within the scientific literature for the three most investigated symptom-defined functional somatic syndromes (FSS) and their classificatory analogues within psychiatry and psychology.

Method: An umbrella review of narrative and systematic reviews with and without meta-analyses based on a search of electronic databases (PubMed, Web of Science, Embase, PsychINFO) was conducted. Eligible reviews were published in English, focused on research of any kind of aetiological factors in adults diagnosed with fibromyalgia syndrome (FMS), irritable bowel syndrome (IBS), chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME), and somatic symptom disorder (SSD)/somatoform disorder (SFD).

Results: We included 452 reviews (132 systematic reviews including meta-analyses, 133 systematic reviews, 197 narrative reviews), of which 132 (29%) focused on two or more of the investigated health conditions simultaneously. Across diagnoses, biological factors were addressed in 90% (k = 405), psychological in 33% (k = 150), social in 12% (k = 54), and healthcare factors in 5% (k = 23) of the reviews. The methodological quality of the included systematic reviews (k = 255) was low (low/critically low: 41% [k = 104]; moderate: 49% [k = 126]; high quality: 10% [k = 25]). The high-quality systematic reviews suggest that deficient conditioned pain modulation, genetic factors, changes in the immune, endocrinological, gastrointestinal, cardiovascular, and nervous system, and psychosocial factors such as sexual abuse and pain catastrophizing increase the risk for FSS.

Conclusion: Only very few systematic reviews have used comprehensive, biopsychosocial disease models to guide the selection of aetiological factors in FSS research. Future research should strive for higher scientific standards and broaden its perspective on these health conditions.

Keywords

chronic fatigue syndrome, myalgic encephalomyelitis, aetiology, fibromyalgia, irritable bowel syndrome, functional somatic syndromes, systematic review

Highlights

- This is an umbrella review of 452 reviews on the aetiology of FSS.
- Biological factors were most commonly examined as correlates of FSS.
- In 90% of the included reviews the methodological quality was low to moderate.
- High-quality reviews found genetic, immuno-endocrinologic, GI, and CV risk factors.
- High-quality reviews found psychosocial risk factors: sexual abuse, catastrophizing.

How physicians conceptualize disease determines their attitude towards their patients and the problems they present (Engel, 1977). This is particularly relevant in case of the so-called *functional somatic syndromes (FSS)* (Henningsen et al., 2007; Wessely et al., 1999). FSS are characterised by somatic symptoms that currently cannot be attributed to reproducibly observable pathophysiological processes, described by the rather outdated but in the past very popular term of *medically unexplained symptoms (MUS)*. Medicine has a long tradition of struggling with classifying and understanding FSS within the traditional disease model, resulting in a large variety of diagnostic labels that reflect the socio-cultural characteristics of a particular decade (neurasthenia, DaCosta syndrome, soldier’s heart syndrome, etc.) (Barsky & Borus, 1999). In recent decades, FSS have typically been investigated within a biopsychosocial model (Engel, 1977), but substantial differences exist between physicians in their belief about the relative importance of certain factors to understand these syndromes. Physicians who adhere to a purely biomedical model might consider these health problems as non-diseases, resulting in reduced scientific interest and neglect in patient care. In contrast, physicians’ overemphasis on psychosocial explanations might influence how such health problems are perceived publicly and might induce stigma.

Epidemiological research suggests that FSS are closely related and partly overlapping (Donnachie et al., 2020; Fink & Schröder, 2010; Janssens et al., 2015; Wessely et al., 1999), although some syndrome-specific aetiological factors have been found (Hamilton et al., 2009). Nevertheless, a variety of diagnostic labels are used, each based on the presence of a selected set of symptoms (Fink & Schröder, 2010), leading to unwanted diversity in diagnostic practice and clinical management (Budtz-Lilly et al., 2015; Creed, 2006; Wolfe, 2009). Diagnoses of FSS that are common in general medical settings as well as in medical specialties such as rheumatology, gastroenterology, or neurology are *fibromyalgia syndrome (FMS)* (Clauw, 2014), *irritable bowel syndrome (IBS)* (Ford et al., 2018), or *chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME)* (Haney et al., 2015). Whereas in psychiatry and in psychology the diagnostic label of somatoform disorders (SFD) was introduced in the 4th edition of the Diagnostic and Statistical Manual for Mental Disorders (DSM-IV; American Psychiatric Association, 2000). The diagnostic category of SFD which mainly focused on excluding a medical explanation of the somatic symptom(s) was replaced by the *somatic symptom disorders (SSD)* in the 5th edition of the DSM (DSM-5; American Psychiatric Association, 2013). SSD include syndromes of medically unexplained as well as syndromes of explained symptoms and – in comparison DSM-IV – rather emphasise the psychological distress associated with poor symptom management and psychological features such as extensive anxiety, dysfunctional thoughts and behaviours associated with the somatic symptom. Besides these diagnostic entities in DSM-IV and -5, several other concepts of MUS were established. For example in research literature concepts of multiple MUS such as the somatic symptom index-4/6 (Escobar et al., 1989) were introduced. In this paper we will refer to single FSS (including IBS, FMS, CFS/ME). We will summarise studies that include patients with multiple MUS in a sense of somatoform disorders, or other syndromes of multiple MUS under the umbrella term SSD.

This variety of diagnostic labels reflects also different aetiological views (Ford et al., 2018; Haney et al., 2015; Schröder & Fink, 2011) with important consequences. For example, compared to receiving a somatic diagnosis, a psychiatric diagnosis for FSS-related symptoms importantly impacts the patient's behaviour, the patient-physician interaction (Budtz-Lilly et al., 2015; Wolfe, 2009), and is associated with more stigma. Previous research has shown that if patients present their symptoms with a more somatic versus psychosocial focus, they are more likely to receive a somatic diagnosis (Salmon et al., 2007). Consequently, receiving a psychiatric vs. somatic diagnosis for the same problem might influence the availability of certain healthcare services for patients.

FSS are highly prevalent, up to 22% in primary care (De Waal et al., 2004) and up to 66% in some medical specialties (Nimnuan et al., 2001). Associated functional limitations are as severe as in well-defined chronic physical diseases (Joustra et al., 2015). Direct medical costs and indirect costs as a consequence of sick leave and disability are high (Ford et al., 2018; Rask et al., 2015). Given this high prevalence and associated burden, a shared conceptualisation of FSS is urgently needed in order to optimise clinical management (Murray et al., 2016; Yon et al., 2015).

Facing this challenge, we systematically assessed the variety and relative dominance of current aetiological views – i.e., from genes to biochemistry, pathophysiology, individual psychological features, and cultural and healthcare factors – as represented in the scientific reviews on these health conditions across different syndrome definitions. We selected the three most well-described FSS (IBS, FMS, CFS/ME) and the somatic symptom disorders (SSD, focusing on individuals with syndromes of multiple medically unexplained symptoms or the precedingly used diagnostic label of somatoform disorders). The objectives of this review were to identify the predominant aetiological factors and proposed illness mechanisms in existing research literature to explain FSS and SSD, and to explore the level of evidence for aetiological factors and proposed illness mechanisms according to systematic reviews and meta-analyses. The methodological quality of reviews for specific investigated aetiological factors across syndrome definitions was evaluated, and the few currently well-documented aetiological factors are discussed. Finally, we provide implications for research.

Method

Literature Search

A literature search was performed in Medline (PubMed) and Embase (Embase.com) in January 2016, in PsycInfo (OVID) and Web of Science (Clarivate Analytics) in February 2016. The searches were updated in August 2017, February 2020, and January 2022. We included meta-analyses, systematic reviews, and narrative reviews, published in English between 1990 and the search dates, which focus on research of aetiological factors and/or illness mechanisms in functional somatic syndromes (FSS) and somatic symptom disorders (SSD) in adults. Narrative reviews were included to give a comprehensive overview as they form a large part of the available reviews. More specifically, we included reviews on the three most investigated FSS – chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME), irritable bowel syndrome (IBS), and fibromyalgia syndrome (FMS) –, and on somatic symptom disorders (SSD) (focusing on individuals with syndromes of multiple medically unexplained symptoms [MUS] or the precedingly used diagnostic label of somatoform disorders and classificatory equivalents). The complete search strategy is available in the Supplementary Material 1 or from PROSPERO 2017 CRD42017053596[1].

---

[1] https://www.crd.york.ac.uk/PROSPEROFILES/53596_STRATEGY_20170105.pdf

Data Extraction

Titles, abstracts, and full texts of studies retrieved using the search strategy were entered at the Covidence platform[2]. Subsequently, the title and abstract of each retrieved reference were screened online by two review team members independently to identify reviews that met the eligibility criteria. Next, the full text of potentially eligible reviews was independently further assessed for inclusion by two review team members. Two raters tried to solve disagreements by finding consensus, if necessary, by involving a third review author.

A standardized, pilot tested form was entered at the REDCap-platform (Harris et al., 2009), hosted at Aarhus University and Utah State University, and was used to extract data from the finally included reviews (Box 1, for details, see Prospero protocol[3]). For each eligible review, two review authors extracted data independently, discrepancies were identified and resolved through consensus discussion, with a third author where necessary. We did not allow raters to extract data and score quality of reviews they had authored.

Box 1

Data That Were Extracted by Two Independent Researchers From Each Review

- Type of review: narrative vs. systematic with/without meta-analysis
- Diagnostic concept: broad diagnostic concept vs. specific diagnosis
- Diagnoses covered: IBS, FMS, CFS/ME, SSD, others
- Number of included studies
- Minimum and maximum number of participants in the included studies
- Type of sample: clinical vs. population-based
- Aetiological domains covered: biological, psychological, social, health care system
- Main finding, and magnitude of main finding(s) for meta-analyses
- Authors' interpretation of the main findings
- Methodological quality, based on AMSTAR-2

Methodological Quality

The methodological quality of each included review was assessed using the Assessment of Multiple SysTemAtic Reviews (AMSTAR) (Shea et al., 2007). AMSTAR was developed

---

[2] https://www.covidence.org

[3] See Footnote 1.

for critically appraising systematic reviews of randomised clinical trials. We adjusted the tool for our purposes as described below. According to the instructions of AMSTAR-2 (Shea et al., 2017), critical domains of the quality of reviews have to be identified. For this purpose three authors (MK, CR, and JR) independently indicated which of the 11 items of the original AMSTAR tool (Shea et al., 2007) indicated critical flaws, with discrepancies solved by consensus. The resulting unanimous critical items were Item 3 (Was a comprehensive literature search performed?), Item 6 (Were the characteristics of the included studies provided?), and Item 8 (Was the scientific quality of the included studies used appropriately in formulating conclusions?). Overall confidence in the results of the review was rated according to the AMSTAR-2 guidance as high (zero or one non-critical weakness), moderate (more than one non-critical weakness but no critical flaws), low (one critical flaw with or without non-critical weaknesses), or critically low (more than one critical flaw with or without non-critical weaknesses).

Data Synthesis and Analysis

First, we obtained descriptive statistics of the frequency of diagnoses studied in the included reviews (FMS, IBS, CFS/ME, SSD), and of the type of review (narrative, systematic, meta-analysis) according to the year of publication. We distinguished between reviews that were diagnosis-specific, i.e., explored only one FSS diagnosis, and reviews that were based on a broad diagnostic concept, i.e., investigated at least two FSS simultaneously, SFD, SSD, and classificatory equivalents. Reviews that, for instance, investigated both FMS and other pain syndromes simultaneously were regarded broad reviews, but only data on FMS were extracted. Second, we analysed the predominant aetiological approach per diagnostic category, i.e., the frequency with which each domain of aetiological factors (see Box 1) was addressed. We defined reviews that assessed multiple aetiological factors from at least two aetiological domains simultaneously as those investigating a broad biopsychosocial model. Third, we analysed the frequency of specific aetiological factors per diagnostic category. Fourth, we assessed the methodological quality (high – moderate – low – critically low) for systematic reviews with or without meta-analysis per year of publication. Fifth, we provide a detailed overview of the few high-quality systematic reviews and analysed the associations of the investigated aetiological factors with FSS.

Patient and Public Involvement

The central aim of our review was to systematically assess and analyse the variety and relative dominance of current aetiological views represented in systematic reviews of certain FSS. This research question did not provide opportunities to involve patients in the design, conduct, or reporting of our review. However, we plan to involve patients in disseminating our research findings (e.g., by presenting our results at meetings of patient interest and support groups).

Results

Search Results and Descriptive Variables of Included Reviews

We identified 5,605 reviews and assessed 980 full text articles for eligibility (see PRISMA checklist in the Supplementary Material 8 and PRISMA flow chart in Supplementary Material 10). We excluded 526 articles and included 454 articles (reporting on 452 reviews) in our descriptive analysis. Lists of all excluded and included reviews are in the online supplementary material (Supplementary Material 2 and 3).

Supplementary Material 4 provides characteristics of included systematic reviews with and without meta-analyses. Characteristics of included narrative reviews are summarised in Supplementary Material 5.

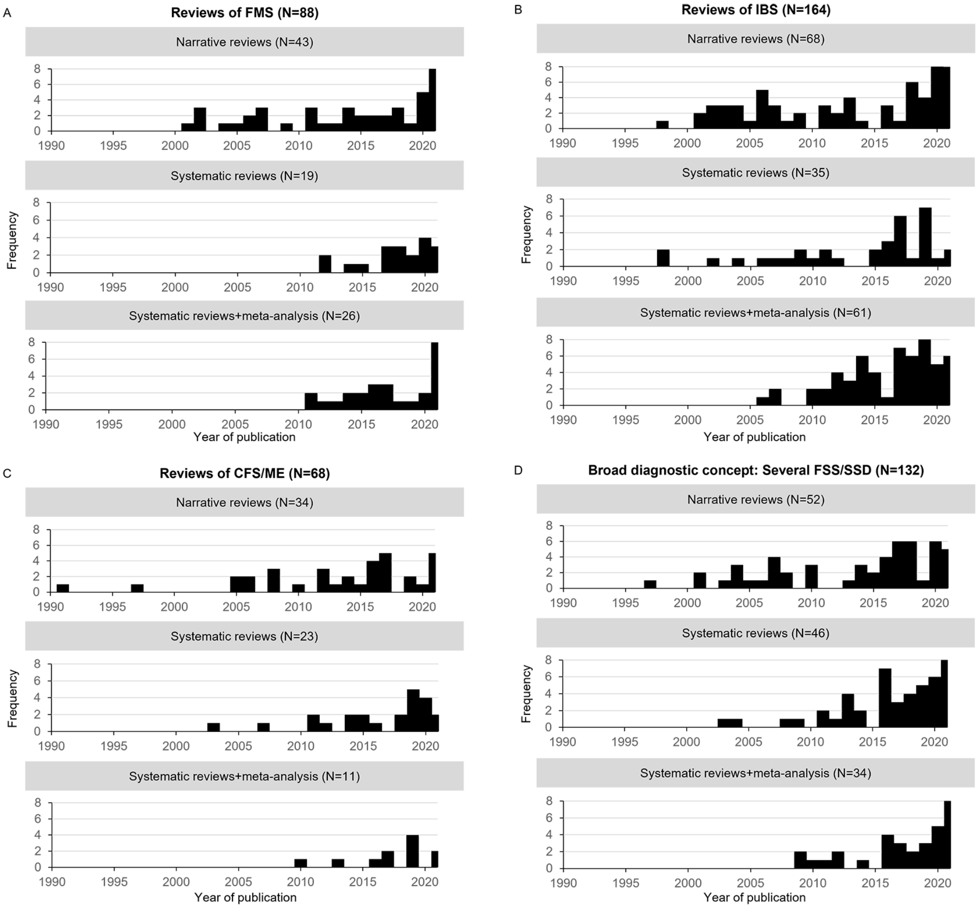
Figure 1 provides the frequency of diagnosis-specific reviews (A-C) and reviews with a broader diagnostic conceptualisation (D) per publication year since 1990, divided into narrative reviews (k = 197), systematic reviews without meta-analyses (k = 123), and with meta-analyses (k = 132). The majority of reviews (71%, 320/452) were diagnosis-specific and of these 51% (164/320) were done in IBS. Focusing on systematic reviews with meta-analyses only, we found the same tendency: 74% (98/132) were diagnosis-specific and of these 62% (61/98) were IBS-specific. While the numbers of reviews on FMS and CFS/ME (from the pool of all included reviews) were comparable (88 and 68, respectively), the number of systematic reviews with meta-analyses on FMS (k = 26) was almost twofold compared with CFS/ME (k = 11) (Figure 1, A and C). We identified 17 reviews on SSD only, 3 narrative, 8 systematic without meta-analyses, and 6 systematic with meta-analyses. These are reviews summarised under reviews with a broader diagnostic concept in Figure 1D.

Predominant Aetiological Approach in FMS, IBS, CFS/ME, and SSD

Figure 2 provides the frequency of reviews covering biological (A), psychological (B), social (C), and healthcare (D) aetiological factors per publication year since 1990, divided into narrative reviews, systematic reviews with and without meta-analyses. In total 90% (405/452) of all reviews proposed or investigated biological factors to explain FSS, while 33% (150/452) proposed psychological, and 12% (54/452) social factors (Figure 2A-C). Several reviews included more than one group of aetiological factors (28%, 127/452), i.e., investigated aetiology on different levels simultaneously. Healthcare factors were discussed in 5% of the reviews (23/452) only (Figure 2D). The primary scientific interest was also pronounced in systematic reviews with meta-analyses: 88% (116/132) investigated biological factors, while only 20% (27/132) explored psychological, social, or healthcare

Figure 1

A-D. *Frequency of Diagnosis-Specific (A-C) and Broad Reviews (i.e., Covering More Than One FSS Diagnosis and SSD) (D) Per Year of Publication Since 1990, Divided Into Narrative Reviews, Systematic Reviews Without and Systematic Reviews With Meta-Analyses*

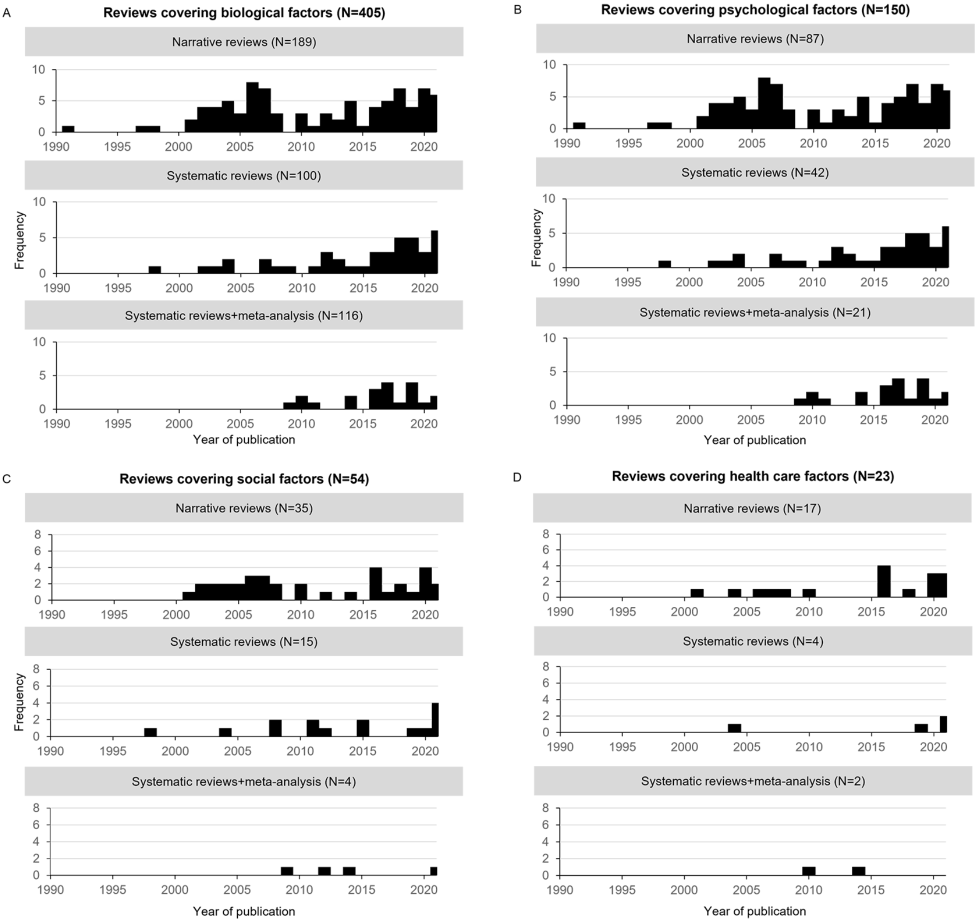


Note. Reviews focusing on somatic symptom disorder are summarised under reviews with a broad diagnostic concept. FMS = fibromyalgia syndrome; FSS = functional somatic syndrome; IBS = irritable bowel syndrome; CFS / ME = chronic fatigue syndrome / myalgic encephalomyelitis; SSD = somatic symptom disorder.

factors. There was no indication that this relative dominance of biologically oriented reviews and meta-analyses changed during the past 20 years. Supplementary Material 4 and 5 shows that only 19% (87/452) of the included reviews are published in journals that are categorised in the field of psychiatry or social sciences (e.g., psychology, behavioural sciences, multidisciplinary sciences, sport sciences, public/environmental/occupational

Figure 2

***A-D.** Frequency of Reviews (Regardless Investigated Diagnoses) Covering Biological (A), Psychological (B), Social (C) and Healthcare (D) Factors Per Year of Publication Since 1990, Divided Into Narrative Reviews, Systematic Reviews Without and Systematic Reviews With Meta-Analyses*



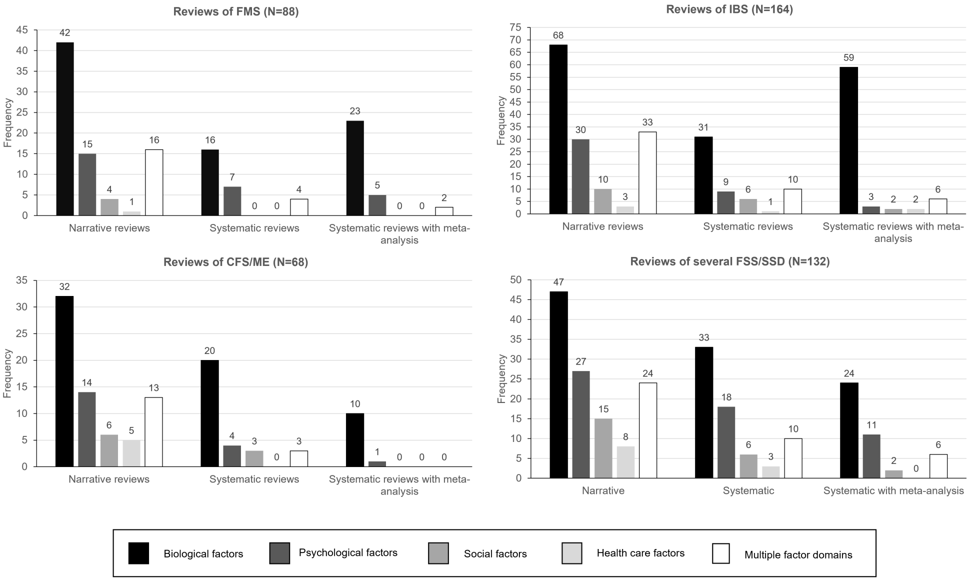
Note. One review may cover several factors, and hence appear in more than one of the Figures A-D.

health), whereas the remaining reviews were mostly published in medical journals (e.g., gastroenterology, neurology, rheumatology) or journals in biology and pharmacology.

Figure 3 provides the aetiological domains covered in diagnosis-specific reviews of FMS (A), IBS (B), CFS/ME (C) and reviews with broad diagnostic concepts (more than one FSS or SSD) (D). The dominance of a primarily biological approach (i.e., the attempt to describe the aetiology on a basic or “mechanistic” level only) was most evident in

Figure 3

A-D. *Aetiological Domains Covered in Diagnosis-Specific Reviews (A-C) and Reviews Applying a Broad Diagnostic Concept (D)*



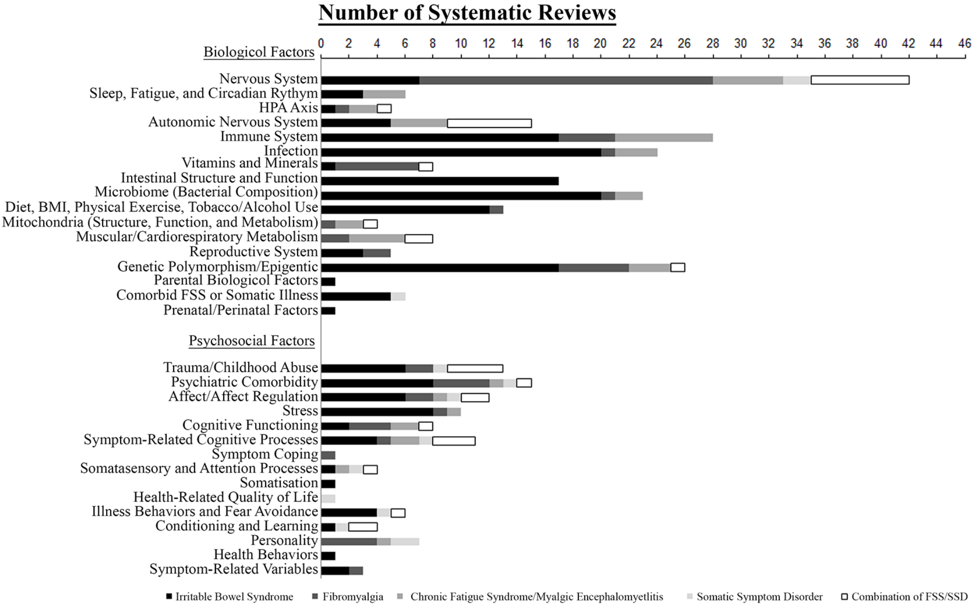
Note. One review may propose or investigate more than one aetiological factor. 'Multiple factor domains' indicates reviews that include two or more domains simultaneously. FMS = fibromyalgia syndrome; FSS = functional somatic syndrome; IBS = irritable bowel syndrome; CFS / ME = chronic fatigue syndrome / myalgic encephalomyelitis; SSD = somatic symptom disorder.

IBS-specific reviews: 96% (158/164) covered biological factors, while only 26% (42/164) covered psychological and 11% (18/164) social factors. A broad biopsychosocial model (i.e., acknowledging the interplay of aetiological factors) was proposed in 30% (49/164). The distribution of investigated domains of aetiological factors was similar in FMS, with 92% (81/88) addressing biological factors, 31% (27/88) psychological factors, 5% (4/88) social factors, and 25% (22/88) a broad biopsychosocial model. Regarding CFS/ME, in 28% (19/68) of the reviews addressed psychological factors, 13% (9/68) social factors, and broad models were included in 24% (16/68). However, the proportion of reviews investigating biological factors was with 91% (62/68) also high for CFS/ME.

Figure 4 displays the number of systematic reviews (with and without meta-analyses) that investigated specific biological or psychosocial aetiological factors for each FSS diagnosis. Investigated biological factors included nervous and autonomic nervous system, sleep, hypothalamus pituitary adrenal (HPA)-axis, immune system, infection, vitamins and minerals, intestinal structure and function, intestinal bacterial composition, diet

Figure 4

Frequency of Specific Investigated Biological and Psychosocial Factors in Systematic Reviews, Divided Into Reviews Addressing Irritable Bowel Syndrome, Fibromyalgia, Chronic Fatigue Syndrome, and Reviews Covering Somatic Symptom Disorder or More Than One of the Investigated Diagnoses Simultaneously



and body mass index (BMI), physical exercise, tobacco and alcohol use, mitochondrial structure or function and metabolism, muscular and cardiorespiratory metabolism, reproductive system, genetic polymorphisms and epigenetic changes, parental biological factors, comorbid FSS or somatic illness, and prenatal or perinatal factors (Supplementary Material 6A). Psychosocial factors also embraced a wide variety of aetiological theories, from developmental issues such as early trauma or impaired affect regulation, over learning processes such as attentional bias or conditioning, to specific illness behaviours or coping styles, and finally personality structure or interindividual (i.e., social or societal) factors (Supplementary Material 6B).

Methodological Quality

The figures in Supplementary Material 10 provide the number of systematic reviews per publication year within the four quality strata. In general, the quality of reviews was low. Of the 255 included systematic reviews (with and without meta-analyses), the quality of 35 (14%) was critically low, 69 (27%) was low, 126 (49%) was moderate, and only 25

(10%) were considered of high quality. Figures A-D in Supplementary Material 9 also display the median year of publication within each quality category, ranging from 2015 for reviews of critically low quality to 2019 for reviews of moderate quality. The median publication year of the 25 high quality reviews was 2019 (range 2007-2021). Of these, 12 were IBS-specific, 3 FMS-specific, 2 CFS/ME-specific and the remaining 8 reviews were on individuals with multiple FSS.

Currently Well-Documented Aetiological Factors

The table in the Supplementary Material 7 provides a detailed overview of the content and findings of the 25 systematic reviews of high quality. Moderate associations of *previous gastrointestinal infections* with IBS were reported in six reviews (Klem et al., 2017; Li et al., 2020; Saha et al., 2022; Schwille-Kiuntke et al., 2015; Svendsen et al., 2019; Wang et al., 2023). Limitations of these findings are high heterogeneity (Klem et al., 2017), and potential publication bias (Schwille-Kiuntke et al., 2015). Moreover, only a low number of reviews reported on specific pathogens and the results of the synthesis of data from these few reviews has to be interpreted with caution (Svendsen et al., 2019). IBS was the most commonly examined FSS regarding the previous infections as potential risk factor. There was only one study that examined relationship between a previous infection, a human herpes virus (HHV)-6 infection, and CFS (Mozhgani et al., 2022). The meta-analysis showed a 1.7 times increased risk in individuals with CFS to have a previous HHV-6 infection compared to healthy controls. *Lactose intolerance*, but not lactose maldigestion, was identified as risk factor of IBS in one review (Varjú et al., 2019). Differential analyses for IBS subtypes were done only in a small number of included studies (Varjú et al., 2019). Thus, conclusions of which subgroups are mostly affected by lactose intolerance cannot be drawn. Included studies vary substantially regarding the diagnostic criteria of IBS and diagnostic threshold of the lactose intolerance test (Varjú et al., 2019). *Gastrointestinal dysbiosis*, measured as count of lactobacillus, bifidobacterium, E coli, and enterobacter, significantly deviated between individuals with IBS and healthy control subjects (Wang et al., 2020). Here again, the increased heterogeneity of the studies included to the review limits the interpretability of the findings. Apart from lower *vitamin E levels*, no associations were found between vitamins or minerals and CFS or FMS (Joustra et al., 2017). The role of genetic factors, functional polymorphism in the gene encoding for activity of the *serotonin transporter protein* (SERT-P), was examined in individuals with IBS in two reviews (Van Kerkhoven et al., 2007; Zhu et al., 2018). Results are mixed, one analysis showed an increased risk of IBS associated with a functional polymorphism in the SERT-P gene (Zhu et al., 2018), whereas the other review showed no association (Van Kerkhoven et al., 2007). Two recent meta-analyses found low to moderate associations for *parasympathetic nervous system activity* as measured by means of high-frequency heart-rate variability (HRV) for IBS and FMS (Sadowski et al., 2021; Tracy et al., 2016). However, these estimates are based on only three to four studies per diagnosis, including a double publication on the same sample for FMS (Sadowski et al., 2021; Tracy et al., 2016). Additionally, one meta-analysis found changes in parasympathetic activity/HRV in a mixed group of CFS, FMS, and IBS patients, although these differences disappeared after correction for publication bias (Tak et al., 2009). A weak association between hypocortisolism (i.e. *HPA dysfunction*) and CFS/ME was found (Tak et al., 2011). One review examined the *immune status* in individuals with FMS (Andrés-Rodríguez et al., 2020). Compared to healthy controls, effect sizes indicated an increased level of different types of interleukins in subjects who are diagnosed with FMS. Another review showed a significantly decreased level of *conditioned pain modulation* in individuals with IBS compared to a healthy control sample (Albusoda et al., 2018). Amiri et al. (2021) demonstrated a significantly lower *nociceptive flexion reflex* threshold in patients with FMS. The nociceptive flexion reflex is a physiological, polysynaptic reflex triggered by a nociceptive stimulus activating a withdrawal response (Smith et al., 2017). A decreased nociceptive flexion reflex threshold has been discussed as a possible biomarker of central sensitization that may cause alteration of central nervous system processing in individuals with chronic musculoskeletal-related pain condition (Smith et al., 2017). Núñez-Fuentes et al. (2021) examined in their meta-analysis an association between alterations in *postural balance* and FMS. The authors demonstrated large effects indicating that patients with FMS show significantly worse scores on a variety of different measures of postural balance compared to healthy controls.

The following psychological variables were examined in high-quality reviews: history of sexual abuse and pain catastrophising. Moderate associations were reported for a *history of sexual abuse* and a lifetime diagnosis of IBS (Paras et al., 2009). The association of sexual abuse with FMS is less straightforward; it was only significant in a sensitivity analysis that was restricted to severe abuse, specifically rape (Paras et al., 2009). Limitations include unexplained heterogeneity, methodological limitations, recall bias, and the unknown generalisability to men, since studies were mainly performed in women. Finally, one review showed that *pain catastrophising* explains to a moderate extend variance of pain intensity and disability in individuals with a combination of FMS and CFS/ME (Martinez-Calderon et al., 2019).

Discussion

Statement of Principal Findings

This systematic umbrella review assessed the variety and relative dominance of aetiological factors in both narrative and systematic reviews of the three most acknowledged FSS and SSD (focusing on individuals with syndromes of multiple medically unexplained symptoms [MUS] or the precedingly used diagnostic label of somatoform disorders and classificatory equivalents). Although the number of systematic reviews has been increasing substantially in recent years, the review quality has only marginally improved. Almost three-quarter of the reviews was diagnosis-specific, with IBS being the most prominent syndrome. Very few reviews have taken a broad view across diagnoses. This is remarkable given the substantial diagnostic overlap among syndromes. This means that most individuals who are included in a study of IBS, for example, may suffer from other, co-morbid FSS – however, very few original studies on IBS distinguish between study participants who have IBS only, and those who have IBS with concomitant other FSS. Results of a recently published systematic review of cohort studies on predictors of the onset of persistent somatic symptoms confirm a similar focus on irritable bowel syndrome (Kitselaar et al., 2023).

It is important to note that the overlap between the medical and psychiatric diagnoses has been reduced in the most recent version of DSM (American Psychiatric Association, 2013). SFD were included in DSM-III (American Psychiatric Association, 1980) and DSM-IV (American Psychiatric Association, 2000), and these diagnoses were based on the presence of somatic symptoms for which there were no demonstrable organic findings or known physiologic mechanisms. DSM-5 (American Psychiatric Association, 2013) replaced this category, given that it was not considered appropriate to make a mental disorder diagnosis solely because a medical cause of the somatic symptoms cannot be demonstrated. The new diagnosis of SSD is made based on the presence of somatic symptoms, explained or unexplained, in combination with dysfunctional cognitions, emotions, or behaviours. This implies that the overlap in diagnostic criteria between FSS and SSD is largely reduced.

Our umbrella review clearly showed a predominance of the biological perspective: Biological factors were included in 90% of reviews, whereas 33% discussed psychological factors; only 28% discussed two or more domains of aetiological factors. Only 5% and 12% discussed healthcare or societal factors, respectively. This biological predominance, i.e. interest on the most basic aetiological level seems to be common and has also been demonstrated in other recently published systematic reviews, for example a review on cohort studies on predictors of the onset of persistent somatic symptoms (Kitselaar et al., 2023). It is in contrast with the views of many health care professionals and with current prevailing clinical management strategies that focus on doctor-patient communication, patients' illness perceptions and illness behaviours, and other healthcare and psychosocial factors (Henningsen et al., 2007; Henningsen, Zipfel, et al., 2018). In other words, the predominant management strategies are not backed up by firm aetiological research. This discrepancy between theoretical assumptions guiding clinical practice and those directing the dominant research focus seems to rely on fundamentally different views on causality and explanatory mechanisms, thereby contributing to enduring controversy and heated debates about legitimisation and "epistemic justice" to patients with FSS (Bernstein, 2016; Cohen, 2017; Mikocka-Walus et al., 2016; Spandler & Allen, 2018).

Specific aetiological factors included both previous or trait factors as well as current or state factors. While predisposing or triggering risk factors may help to identify people at risk or to prevent the development of FSS through control or even elimination of such factors, current or perpetuating factors may be of special interest, as these are potentially modifiable and therefore may be targets for intervention. The few high-quality reviews suggest that both biological (e.g., infection) and psychosocial (i.e., history of sexual abuse or pain catastrophising) factors can increase the risk of FSS. This is in line with current illness models for FSS (e.g., Deary et al., 2007). Four high quality reviews suggest involvement of the ANS in painful FSS (Sadowski et al., 2021; Tak et al., 2009; Tracy et al., 2016) in painful FSS and syndromes of fatigue and exhaustion (Tak et al., 2011). These findings suggest that different symptom clusters may be associated with specific pathophysiological pathways, while a more general dysfunction in interoception may be generic and of relevance for all FSS (Henningsen, Gündel, et al., 2018).

In summary, these high quality reviews show that there is a multiplicity of factors associated with FSS. This could be interpreted as an indicator of subgroups in a group of people diagnosed with a particular syndrome, who have different aetiological pathways. A recently published study (Kendler et al., 2022) examined genetic risk patterns in FSS such as IBS, CFS, and FMS as well as in a prototypic mental health condition, such as depression, and a prototypic somatic condition, rheumathoid arthritis. The authors could demonstrate unique profiles of family genetic risk scores in individuals with specific single FSS that were very different to major depression and rheumatoid arthritis. Another recently published study (Creed, 2022b) examined risk factors in individuals with self-reported IBS. This study identified partly overlapping and partly unique patterns of risk factors for a subgroup of IBS patients with previous mental health conditions compared to individuals with IBS but no previous mental health problem.

There is another important aspect that we have to consider when we interpret the high-quality reviews included in our umbrella review: We observe a relative infrequency of studies that measure several putative risk factors simultaneously. There are some single examples, such as the rather “biological“ study of IBS by Dunlop et al. (2003) which examined risk factors simultaneously. Dunlop et al. (2003) found that both increased enterochromaffin cell counts and depression were equally important predictors of developing post-infectional IBS. However most studies of biological factors in our umbrella review fail to include a psychological or social measure in addition to the biological one. Another example concerned fibromyalgia, for which numerous somatic symptoms are a risk factor (Creed, 2022a). However, a study by Creed (2022a) demonstrated that there are new onset cases of FMS with only few somatic symptoms and that this subgroup of FMS patients shows, compared to individuals with FMS and numerous other somatic symptoms, a unique pattern of risk factors. These results are in accordance with Kendler’s concerns that we finally have to withdraw from the dualistic or dichotomous thinking within psychiatry and have to acknowledge that biological, psychological, and social cultural domains are inter-twined with each other in aetiological pathways (Kendler, 2012).

Strength and Weaknesses of the Review

Our review has a few limitations. First, we only included the most common FSS. Other intensively investigated syndromes or symptoms, such as chronic low back pain, might have added other well-documented aetiological factors (Vlaeyen et al., 2018). Second, systematic reviews of chronic pain not always provided specific results for primary pain as opposed to mixed pain, or secondary pain, and it was often difficult to extract specific details for FMS and IBS from those reviews. Third, our quality rating was done using a tool that was constructed for the evaluation of reviews of intervention trials. Fourth, we only included reviews, meaning that the most novel aetiological factors as investigated in empirical studies may not have been covered. Finally, the reviews included in our umbrella review are mainly based on studies implementing cross-sectional designs. We did not include animal research that would allow experimental designs and conclusions about causal factors.

Our review has also important strengths. First, it is the first comprehensive overview that covers aetiological factors of the most well-known FSS together. Second, both medical and psychiatric definitions of these syndromes were used, thereby avoiding bias. Third, we restricted our analysis to reviews which are typically the primary sources for guidelines that affect daily clinical practice. Finally, we were interested in aetiological factors on various levels. Therefore, we regard this review a very first step to unravel the "dappled nature of causes" of these syndromes (Kendler, 2012).

Implications for Research and Clinical Management

Our results have important implications for future research: First, we showed that the FSS are largely studied separately, with only a minority of reviews including more than one syndrome, despite the empirical overlap in symptoms and shared non-symptom characteristics of the patients (Wessely et al., 1999) suggesting that they constitute a family of disorders (Fink & Schröder, 2010; Janssens et al., 2015). Future studies should investigate them together, since this could facilitate the identification of both syndrome-specific and generic aetiological and pathogenic factors at different levels, which would critically inform the discussion between “splitters” and “lumpers” (Fink, 2017).

Second, current explanatory models promote a biopsychosocial approach to diseases in general (Rief & Broadbent, 2007; Witthöft & Hiller, 2010). However, since the appearance of IBS, FMS and CFS/ME as MESH terms in MEDLINE from 1989 onwards, very few systematic reviews have used comprehensive disease models to guide the selection of aetiological factors in FSS research. The still widely acknowledged dualistic "hardware versus software" rationale likely has slowed down scientific progress and might continue to do so (Fink & Schröder, 2010; Rosmalen, 2010) until it is replaced by empirically based pluralism (Kendler, 2012). It would be a major step forward if different groups of aetiological factors on different levels would be combined into one longitudinal, multidisciplinary study, in order to examine their interrelations (Rosmalen, 2010). This is currently done in a number of large epidemiological studies, e.g., DanFunD (Dantoft et al., 2017) and LifeLines (Scholtens et al., 2015).

Our review also has important implications for clinical management: The knowledge about aetiological factors that has been gained from our review has to be translated into explanatory models for single patients. For each individual case the contribution of biological, psychosocial, and healthcare factors has to be weighted, acknowledged, and negotiated with the patient. After all, it is the individual patient's history of risk and protective factors as well as his/her/their needs and wishes that is the foundation on which personalised care is built, not the theoretical preferences of the clinician (Gask, 2018).

Conclusions and Future Research

In summary, our umbrella review reveals that the literature on aetiological factors in FSS and SSD is predominantly characterised by a diagnosis-specific perspective with a focus on biological factors, based on a purely biomedical conceptualisation of FMS, IBS and CFS/ME as distinct disease entities. SSD, or the previously used diagnostic category somatoform disorders, is only sparsely investigated. The majority of reviews provide expert views rather than firm results, and overall the reviews are very often of low quality and mostly implement only cross-sectional designs. We however identified 25 systematic reviews, partly including meta-analyses, that provide information of a variety of biological factors and some psychological factors that function as potential mechanisms. The information gained from these high-quality studies should be translated into explanatory models for patients.

We believe that future research should strive for higher scientific standards and more interdisciplinary research collaboration. We recommend that more research work should focus on examining the different FSS together. Examining differences as well as similarities of specific FSS could be reached by an approach that uses the same data set gained in the same population (Monden et al., 2022). Recognising these health conditions as closely related and including all relevant factors that potentially play a role may lead to distinguishing evidence-based subtypes or syndromes that may benefit from person-centred approaches. It is our hope that this review contributes to the development of a commonly accepted and evidence-based conceptualisation of FSS and SSD in both medicine and psychiatry.

Funding: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Acknowledgments: We thank librarian Helene Sognstrup for literature search and import of references and papers, our web designers Nicolaj Knudsen and Ruben Engrob Nielsen for assistance with REDCap, our MSc student Linda Jolink for extraction of data regarding biological mechanisms, and Michael Cathro and Nathan Nelson for helping us with the extraction of journal-related data.

Competing Interests: Omer Van den Bergh is a Subject Editor for *Clinical Psychology in Europe* but played no editorial role in this particular article or intervened in any form in the peer review process.

Author Contributions: AS, JR, and MK invented the review. AS, MK, JR and CUR designed the final review, with contribution from OVdB and KJP. MK and AS wrote the review protocol, with contribution from all authors. AS and CUR finalized the search protocols together with the librarian HS. HS conducted the searches. AS choose technical platforms. All authors contributed to screening of titles, abstracts and full texts and to the extraction of data. AS, MK, and JR did the analyses, with contribution of CUR. AS, MK, and MWP created the figures. All authors were involved in the interpretation of data. AS, MK and JR drafted the manuscript, and all authors critically reviewed, edited and approved the final manuscript.

Data Availability: Data are available on reasonable request from the corresponding author.

Supplementary Materials

The Supplementary Materials (Kleinstäuber et al., 2023) contain the following items:

- **Supplementary Material 1.** Results of the search of electronic literature database
- **Supplementary Material 2.** References of included reviews (k = 452)
- **Supplementary Material 3.** References of reviews excluded after fulltext review (k = 526)
- **Supplementary Material 4.** Characteristics of included systematic reviews without meta-analysis (k = 123) and systematic reviews with meta-analyses (k = 132)
- **Supplementary Material 5.** Characteristics of included narrative reviews (k = 197)
- **Supplementary Material 6A.** Specific biological factors investigated in diagnosis-specific systematic reviews (FMS, IBS, CFS/ME and SSD) or systematic reviews that investigated at least two of these diagnoses simultaneously (combinations of FSS/SSD)
- **Supplementary Material 6B.** Specific psychosocial factors investigated in diagnosis-specific systematic reviews (FMS, IBS, CFS/ME and SSD) or systematic reviews that investigated at least two of these diagnoses simultaneously (combinations of FSS/SSD)
- **Supplementary Material 7.** Characteristics of systematic reviews with meta-analysis with an overall rating ‘high’ of confidence in the results of the review according to AMSTAR-2 (k = 25)
- **Supplementary Material 8.** PRISMA Checklist
- **Supplementary Material 9.** Study selection process (PRISMA flow chart)
- **Supplementary Material 10.** Frequency of systematic reviews per year of publication since 1990, divided into reviews of critically low, low, moderate and high quality, according to the

Assessment of Multiple SysTemAtic Reviews (AMSTAR-2). The grey vertical lines indicate the median (Md) publication year within each quality stratum.

Index of Supplementary Materials

Kleinstäuber, M., Schröder, A., Daehler, S., Pallesen, K. J., Rask, C. U., Sanyer, M., Van den Bergh, O., Weinreich Petersen, M., & Rosmalen, J. G. M. (2023). *Supplementary materials to "Aetiological understanding of fibromyalgia, irritable bowel syndrome, chronic fatigue syndrome and classificatory analogues: A systematic umbrella review"* [Additional information]. PsychOpen GOLD. https://doi.org/10.23668/psycharchives.13273

References

Albusoda, A., Ruffle, J. K., Friis, K. A., Gysan, M. R., Drewes, A. M., Aziz, Q., & Farmer, A. D. (2018). Systematic review with meta-analysis: Conditioned pain modulation in patients with the irritable bowel syndrome. *Alimentary Pharmacology & Therapeutics, 48*(8), 797–806. https://doi.org/10.1111/apt.14965

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). American Psychiatric Press.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). American Psychiatric Press.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Amiri, M., Esmaili, H., Hamad, A. H., Alavinia, M., Masani, K., & Kumbhare, D. (2021). Nociceptive flexion reflex threshold in chronic pain patients: A needed update for the current evidence. *American Journal of Physical Medicine & Rehabilitation, 100*(8), 750–759. https://doi.org/10.1097/PHM.0000000000001626

Andrés-Rodríguez, L., Borràs, X., Feliu-Soler, A., Pérez-Aranda, A., Angarita-Osorio, N., Moreno-Peral, P., Montero-Marin, J., García-Campayo, J., Carvalho, A. F., Maes, M., & Luciano, J. V. (2020). Peripheral immune aberrations in fibromyalgia: A systematic review, meta-analysis and meta-regression. *Brain, Behavior, and Immunity, 87*, 881–889. https://doi.org/10.1016/j.bbi.2019.12.020

Barsky, A. J., & Borus, J. F. (1999). Functional somatic syndromes. *Annals of Internal Medicine, 130*(11), 910–921. https://doi.org/10.7326/0003-4819-130-11-199906010-00016

Bernstein, J. (2016). Not the last word: Fibromyalgia is real. *Clinical Orthopaedics and Related Research, 474*(2), 304–309. https://doi.org/10.1007/s11999-015-4670-6

Budtz-Lilly, A., Fink, P., Ornbol, E., Vestergaard, M., Moth, G., Christensen, K. S., & Rosendal, M. (2015). A new questionnaire to identify bodily distress in primary care: The 'BDS checklist'. *Journal of Psychosomatic Research, 78*(6), 536–545. https://doi.org/10.1016/j.jpsychores.2015.03.006

Clauw, D. J. (2014). Fibromyalgia: A clinical review. *Journal of the American Medical Association, 311*(15), 1547–1555. https://doi.org/10.1001/jama.2014.3266

Cohen, H. (2017). Controversies and challenges in fibromyalgia: A review and a proposal. *Therapeutic Advances in Musculoskeletal Disease, 9*(5), 115–127. https://doi.org/10.1177/1759720X17699199

Creed, F. (2006). Should general psychiatry ignore somatization and hypochondriasis? *World Psychiatry: Official Journal of the World Psychiatric Association (WPA), 5*(3), 146–150. https://www.ncbi.nlm.nih.gov/pubmed/17139341

Creed, F. (2022a). The risk factors for self-reported fibromyalgia with and without multiple somatic symptoms: The Lifelines cohort study. *Journal of Psychosomatic Research, 155*, Article 110745. https://doi.org/10.1016/j.jpsychores.2022.110745

Creed, F. (2022b). Risk factors for self-reported irritable bowel syndrome with prior psychiatric disorder: The Lifelines cohort study. *Journal of Neurogastroenterology and Motility, 28*(3), 442–453. https://doi.org/10.5056/jnm21041

Dantoft, T. M., Ebstrup, J. F., Linneberg, A., Skovbjerg, S., Madsen, A. L., Mehlsen, J., Brinth, L., Eplov, L. F., Carstensen, T. W., Schroder, A., Fink, P. K., Mortensen, E. L., Hansen, T., Pedersen, O., & Jorgensen, T. (2017). Cohort description: The Danish study of functional disorders. *Clinical Epidemiology, 9*, 127–139. https://doi.org/10.2147/CLEP.S129335

Deary, V., Chalder, T., & Sharpe, M. (2007). The cognitive behavioural model of medically unexplained symptoms: A theoretical and empirical review. *Clinical Psychology Review, 27*(7), 781–797. https://doi.org/10.1016/j.cpr.2007.07.002

De Waal, M. W., Arnold, I. A., Eekhof, J. A., & van Hemert, A. M. (2004). Somatoform disorders in general practice: Prevalence, functional impairment and comorbidity with anxiety and depressive disorders. *The British Journal of Psychiatry, 184*(3), 470–476. https://doi.org/10.1192/bjp.184.6.470

Donnachie, E., Schneider, A., & Enck, P. (2020). Comorbidities of patients with functional somatic syndromes before, during and after first diagnosis: A population-based study using Bavarian routine data. *Scientific Reports, 10*(1), Article 9810. https://doi.org/10.1038/s41598-020-66685-4

Dunlop, S. P., Jenkins, D., Neal, K. R., & Spiller, R. C. (2003). Relative importance of enterochromaffin cell hyperplasia, anxiety, and depression in postinfectious IBS. *Gastroenterology, 125*(6), 1651–1659. https://doi.org/10.1053/j.gastro.2003.09.028

Engel, G. L. (1977). The need for a new medical model: A challenge for biomedicine. *Science, 196*(4286), 129–136. https://doi.org/10.1126/science.847460

Escobar, J. I., Rubio-Stipec, M., Canino, G., & Karno, M. (1989). Somatic Symptom Index (SSI): A new and abridged somatization construct. Prevalence and epidemiological correlates in two large community samples. *The Journal of Nervous and Mental Disease, 177*(3), 140–146. https://doi.org/10.1097/00005053-198903000-00003

Fink, P. (2017). Syndromes of bodily distress or functional somatic syndromes – Where are we heading. Lecture on the occasion of receiving the Alison Creed award 2017. *Journal of Psychosomatic Research, 97*, 127–130. https://doi.org/10.1016/j.jpsychores.2017.04.012

Fink, P., & Schröder, A. (2010). One single diagnosis, bodily distress syndrome, succeeded to capture 10 diagnostic categories of functional somatic syndromes and somatoform disorders. *Journal of Psychosomatic Research, 68*(5), 415–426. https://doi.org/10.1016/j.jpsychores.2010.02.004

Ford, A. C., Moayyedi, P., Chey, W. D., Harris, L. A., Lacy, B. E., Saito, Y. A., & Quigley, E. M. M. (2018). American College of Gastroenterology monograph on management of irritable bowel syndrome. *The American Journal of Gastroenterology, 113*(Suppl 2), 1–18. https://doi.org/10.1038/s41395-018-0084-x

Gask, L. (2018). In defence of the biopsychosocial model. *The Lancet Psychiatry, 5*(7), 548–549. https://doi.org/10.1016/S2215-0366(18)30165-2

Hamilton, W. T., Gallagher, A. M., Thomas, J. M., & White, P. D. (2009). Risk markers for both chronic fatigue and irritable bowel syndromes: A prospective case-control study in primary care. *Psychological Medicine, 39*(11), 1913–1921. https://doi.org/10.1017/S0033291709005601

Haney, E., Smith, M. E., McDonagh, M., Pappas, M., Daeges, M., Wasson, N., & Nelson, H. D. (2015). Diagnostic methods for myalgic encephalomyelitis/chronic fatigue syndrome: A systematic review for a National Institutes of Health Pathways to Prevention workshop. *Annals of Internal Medicine, 162*(12), 834–840. https://doi.org/10.7326/M15-0443

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research Electronic Data Capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics, 42*(2), 377–381. https://doi.org/10.1016/j.jbi.2008.08.010

Henningsen, P., Gündel, H., Kop, W. J., Löwe, B., Martin, A., Rief, W., Rosmalen, J. G. M., Schröder, A., van der Feltz-Cornelis, C., & Van den Bergh, O. (2018). Persistent physical symptoms as perceptual dysregulation: A neuropsychobehavioral model and its clinical implications. *Psychosomatic Medicine, 80*(5), 422–431. https://doi.org/10.1097/PSY.0000000000000588

Henningsen, P., Zipfel, S., & Herzog, W. (2007). Management of functional somatic syndromes. *Lancet, 369*(9565), 946–955. https://doi.org/10.1016/S0140-6736(07)60159-7

Henningsen, P., Zipfel, S., Sattel, H., & Creed, F. (2018). Management of functional somatic syndromes and bodily distress. *Psychotherapy and Psychosomatics, 87*(1), 12–31. https://doi.org/10.1159/000484413

Janssens, K. A., Zijlema, W. L., Joustra, M. L., & Rosmalen, J. G. (2015). Mood and anxiety disorders in chronic fatigue syndrome, fibromyalgia, and irritable bowel syndrome: Results from the LifeLines Cohort Study. *Psychosomatic Medicine, 77*(4), 449–457. https://doi.org/10.1097/PSY.0000000000000161

Joustra, M. L., Janssens, K. A., Bultmann, U., & Rosmalen, J. G. (2015). Functional limitations in functional somatic syndromes and well-defined medical diseases: Results from the general population cohort LifeLines. *Journal of Psychosomatic Research, 79*(2), 94–99. https://doi.org/10.1016/j.jpsychores.2015.05.004

Joustra, M. L., Minovic, I., Janssens, K. A. M., Bakker, S. J. L., & Rosmalen, J. G. M. (2017). Vitamin and mineral status in chronic fatigue syndrome and fibromyalgia syndrome: A systematic review and meta-analysis. *PLoS One, 12*(4), Article e0176631. https://doi.org/10.1371/journal.pone.0176631

Kendler, K. S. (2012). The dappled nature of causes of psychiatric illness: Replacing the organic-functional/hardware-software dichotomy with empirically based pluralism. *Molecular Psychiatry, 17*(4), 377–388. https://doi.org/10.1038/mp.2011.182

Kendler, K. S., Rosmalen, J. G. M., Ohlsson, H., Sundquist, J., & Sundquist, K. (2022). A distinctive profile of family genetic risk scores in a Swedish national sample of cases of fibromyalgia, irritable bowel syndrome, and chronic fatigue syndrome compared to rheumatoid arthritis and major depression. *Psychological Medicine, 53*(9), 3879–3886. https://doi.org/10.1017/S003329172200349X

Kitselaar, W. M., van der Vaart, R., Perschl, J., Numans, M. E., & Evers, A. W. M. (2023). Predictors of persistent somatic symptoms in the general population: A systematic review of cohort studies. *Psychosomatic Medicine, 85*(1), 71–78. https://doi.org/10.1097/PSY.0000000000001145

Klem, F., Wadhwa, A., Prokop, L. J., Sundt, W. J., Farrugia, G., Camilleri, M., Singh, S., & Grover, M. (2017). Prevalence, risk factors, and outcomes of irritable bowel syndrome after infectious enteritis: A systematic review and meta-analysis. *Gastroenterology, 152*(5), 1042–1054.e1041. https://doi.org/10.1053/j.gastro.2016.12.039

Li, C., Shuai, Y., Zhou, X., & Chen, H. (2020). Association between Helicobacter pylori infection and irritable bowel syndrome: A systematic review and meta-analysis. *Medicine, 99*(50), Article e22975. https://doi.org/10.1097/MD.0000000000022975

Martinez-Calderon, J., Jensen, M. P., Morales-Asencio, J. M., & Luque-Suarez, A. (2019). Pain catastrophizing and function in individuals with chronic musculoskeletal pain: A systematic review and meta-analysis. *The Clinical Journal of Pain, 35*(3), 279–293. https://doi.org/10.1097/AJP.0000000000000676

Mikocka-Walus, A., Knowles, S. R., Keefer, L., & Graff, L. (2016). Controversies revisited: A systematic review of the comorbidity of depression and anxiety with inflammatory bowel diseases. *Inflammatory Bowel Diseases, 22*(3), 752–762. https://doi.org/10.1097/MIB.0000000000000620

Monden, R., Rosmalen, J. G. M., Wardenaar, K. J., & Creed, F. (2022). Predictors of new onsets of irritable bowel syndrome, chronic fatigue syndrome and fibromyalgia: The lifelines study. *Psychological Medicine, 52*(1), 112–120. https://doi.org/10.1017/S0033291720001774

Mozhgani, S.-H., Rajabi, F., Qurbani, M., Erfani, Y., Yaslianifard, S., Moosavi, A., Pourrostami, K., Baradaran Bagheri, A., Soleimani, A., Behzadian, F., Safavi, M., & Rezaei, F. (2022). Human herpesvirus 6 infection and risk of chronic fatigue syndrome: A systematic review and meta-analysis. *Intervirology, 65*(1), 49–57. https://doi.org/10.1159/000517930

Murray, A. M., Toussaint, A., Althaus, A., & Löwe, B. (2016). The challenge of diagnosing non-specific, functional, and somatoform disorders: A systematic review of barriers to diagnosis in primary care. *Journal of Psychosomatic Research, 80*, 1–10. https://doi.org/10.1016/j.jpsychores.2015.11.002

Nimnuan, C., Rabe-Hesketh, S., Wessely, S., & Hotopf, M. (2001). How many functional somatic syndromes? *Journal of Psychosomatic Research, 51*(4), 549–557. https://doi.org/10.1016/S0022-3999(01)00224-0

Núñez-Fuentes, D., Obrero-Gaitán, E., Zagalaz-Anula, N., Ibáñez-Vera, A. J., Achalandabaso-Ochoa, A., López-Ruiz, M. C., Rodríguez-Almagro, D., & Lomas-Vega, R. (2021). Alteration of postural balance in patients with fibromyalgia syndrome—A systematic review and meta-analysis. *Diagnostics (Basel), 11*(1), Article 127. https://doi.org/10.3390/diagnostics11010127

Paras, M. L., Murad, M. H., Chen, L. P., Goranson, E. N., Sattler, A. L., Colbenson, K. M., Elamin, M. B., Seime, R. J., Prokop, L. J., & Zirakzadeh, A. (2009). Sexual abuse and lifetime diagnosis of somatic disorders: A systematic review and meta-analysis. *Journal of the American Medical Association, 302*(5), 550–561. https://doi.org/10.1001/jama.2009.1091

Rask, M. T., Rosendal, M., Fenger-Gron, M., Bro, F., Ornbol, E., & Fink, P. (2015). Sick leave and work disability in primary care patients with recent-onset multiple medically unexplained symptoms and persistent somatoform disorders: A 10-year follow-up of the FIP study. *General Hospital Psychiatry, 37*(1), 53–59. https://doi.org/10.1016/j.genhosppsych.2014.10.007

Rief, W., & Broadbent, E. (2007). Explaining medically unexplained symptoms-models and mechanisms. *Clinical Psychology Review, 27*(7), 821–841. https://doi.org/10.1016/j.cpr.2007.07.005

Rosmalen, J. G. (2010). The way forward: A case for longitudinal population-based studies in the field of functional somatic syndromes. *Journal of Psychosomatic Research, 68*(5), 399–401. https://doi.org/10.1016/j.jpsychores.2009.12.003

Sadowski, A., Dunlap, C., Lacombe, A., & Hanes, D. (2021). Alterations in heart rate variability associated with irritable bowel syndrome or inflammatory bowel disease: A systematic review and meta-analysis. *Clinical and Translational Gastroenterology, 12*(1), Article e00275. https://doi.org/10.14309/ctg.0000000000000275

Saha, S., Sehgal, K., Singh, S., Grover, M., Pardi, D., & Khanna, S. (2022). Postinfection irritable bowel syndrome following clostridioides difficile infection: A systematic review and meta-analysis. *Journal of Clinical Gastroenterology, 56*(2), e84–e93. https://doi.org/10.1097/MCG.0000000000001536

Salmon, P., Humphris, G. M., Ring, A., Davies, J. C., & Dowrick, C. F. (2007). Primary care consultations about medically unexplained symptoms: Patient presentations and doctor responses that influence the probability of somatic intervention. *Psychosomatic Medicine, 69*(6), 571–577. https://doi.org/10.1097/PSY.0b013e3180cabc85

Scholtens, S., Smidt, N., Swertz, M. A., Bakker, S. J., Dotinga, A., Vonk, J. M., van Dijk, F., van Zon, S. K., Wijmenga, C., Wolffenbuttel, B. H., & Stolk, R. P. (2015). Cohort Profile: LifeLines, a three-generation cohort study and biobank. *International Journal of Epidemiology, 44*(4), 1172–1180. https://doi.org/10.1093/ije/dyu229

Schröder, A., & Fink, P. (2011). Functional somatic syndromes and somatoform disorders in special psychosomatic units: Organizational aspects and evidence-based treatment. *The Psychiatric Clinics of North America, 34*(3), 673–687. https://doi.org/10.1016/j.psc.2011.05.008

Schwille-Kiuntke, J., Mazurak, N., & Enck, P. (2015). Systematic review with meta-analysis: Post-infectious irritable bowel syndrome after travellers' diarrhoea. *Alimentary Pharmacology & Therapeutics, 41*(11), 1029–1037. https://doi.org/10.1111/apt.13199

Shea, B. J., Grimshaw, J. M., Wells, G. A., Boers, M., Andersson, N., Hamel, C., Porter, A. C., Tugwell, P., Moher, D., & Bouter, L. M. (2007). Development of AMSTAR: A measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology, 7*, Article 10. https://doi.org/10.1186/1471-2288-7-10

Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Welch, V., Kristjansson, E., & Henry, D. A. (2017). AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ (Clinical Research Ed.), 358*, Article j4008. https://doi.org/10.1136/bmj.j4008

Smith, S. M., Dworkin, R. H., Turk, D. C., Baron, R., Polydefkis, M., Tracey, I., Borsook, D., Edwards, R. R., Harris, R. E., Wager, T. D., Arendt-Nielsen, L., Burke, L. B., Carr, D. B., Chappell, A., Farrar, J. T., Freeman, R., Gilron, I., Goli, V., Haeussler, J., . . . Witter, J. (2017). The potential role of sensory testing, skin biopsy, and functional brain imaging as biomarkers in chronic pain clinical trials: IMMPACT considerations. *The Journal of Pain, 18*(7), 757–777. https://doi.org/10.1016/j.jpain.2017.02.429

Spandler, H., & Allen, M. (2018). Contesting the psychiatric framing of ME/CFS. *Social Theory & Health, 16*(2), 127–141. https://doi.org/10.1057/s41285-017-0047-0

Svendsen, A. T., Bytzer, P., & Engsbro, A. L. (2019). Systematic review with meta-analyses: Does the pathogen matter in post-infectious irritable bowel syndrome? *Scandinavian Journal of Gastroenterology, 54*(5), 546–562. https://doi.org/10.1080/00365521.2019.1607897

Tak, L. M., Cleare, A. J., Ormel, J., Manoharan, A., Kok, I. C., Wessely, S., & Rosmalen, J. G. M. (2011). Meta-analysis and meta-regression of hypothalamic-pituitary-adrenal axis activity in functional somatic disorders. *Biological Psychology, 87*(2), 183–194. https://doi.org/10.1016/j.biopsycho.2011.02.002

Tak, L. M., Riese, H., de Bock, G. H., Manoharan, A., Kok, I. C., & Rosmalen, J. G. (2009). As good as it gets? A meta-analysis and systematic review of methodological quality of heart rate variability studies in functional somatic disorders. *Biological Psychology, 82*(2), 101–110. https://doi.org/10.1016/j.biopsycho.2009.05.002

Tracy, L. M., Ioannou, L., Baker, K. S., Gibson, S. J., Georgiou-Karistianis, N., & Giummarra, M. J. (2016). Meta-analytic evidence for decreased heart rate variability in chronic pain implicating parasympathetic nervous system dysregulation. *Pain, 157*(1), 7–29. https://doi.org/10.1097/j.pain.0000000000000360

Van Kerkhoven, L. A., Laheij, R. J., & Jansen, J. B. (2007). Meta-analysis: A functional polymorphism in the gene encoding for activity of the serotonin transporter protein is not associated with the irritable bowel syndrome. *Alimentary Pharmacology & Therapeutics, 26*(7), 979–986. https://doi.org/10.1111/j.1365-2036.2007.03453.x

Varjú, P., Gede, N., Szakács, Z., Hegyi, P., Cazacu, I. M., Pécsi, D., Fábián, A., Szepes, Z., Vincze, A., Tenk, J., Balaskó, M., Rumbus, Z., Garami, A., Csupor, D., & Czimmer, J. (2019). Lactose intolerance but not lactose maldigestion is more frequent in patients with irritable bowel syndrome than in healthy controls: A meta-analysis. *Neurogastroenterology and Motility, 31*(5), Article e13527. https://doi.org/10.1111/nmo.13527

Vlaeyen, J. W. S., Maher, C. G., Wiech, K., Van Zundert, J., Meloto, C. B., Diatchenko, L., Battie, M. C., Goossens, M., Koes, B., & Linton, S. J. (2018). Low back pain. *Nature Reviews: Disease Primers, 4*(1), Article 52. https://doi.org/10.1038/s41572-018-0052-1

Wang, C., Yin, Y., Wang, L., Guo, X., Liu, L., & Qi, X. (2023). Association between Helicobacter pylori infection and irritable bowel syndrome: A systematic review and meta-analysis. *Postgraduate Medical Journal, 99*(1169), 166–175. https://doi.org/10.1136/postgradmedj-2021-141127

Wang, L., Alammar, N., Singh, R., Nanavati, J., Song, Y., Chaudhary, R., & Mullin, G. E. (2020). Gut microbial dysbiosis in the irritable bowel syndrome: A systematic review and meta-analysis of case-control studies. *Journal of the Academy of Nutrition and Dietetics, 120*(4), 565–586. https://doi.org/10.1016/j.jand.2019.05.015

Wessely, S., Nimnuan, C., & Sharpe, M. (1999). Functional somatic syndromes: One or many? *Lancet, 354*(9182), 936–939. https://doi.org/10.1016/S0140-6736(98)08320-2

Witthöft, M., & Hiller, W. (2010). Psychological approaches to origins and treatments of somatoform disorders. *Annual Review of Clinical Psychology, 6*, 257–283. https://doi.org/10.1146/annurev.clinpsy.121208.131505

Wolfe, F. (2009). Fibromyalgia wars. *The Journal of Rheumatology, 36*(4), 671–678. https://doi.org/10.3899/jrheum.081180

Yon, K., Nettleton, S., Walters, K., Lamahewa, K., & Buszewicz, M. (2015). Junior doctors' experiences of managing patients with medically unexplained symptoms: A qualitative study. *BMJ Open, 5*(12), Article e009593. https://doi.org/10.1136/bmjopen-2015-009593

Zhu, Y., Zheng, G., & Hu, Z. (2018). Association between SERT insertion/deletion polymorphism and the risk of irritable bowel syndrome: A meta-analysis based on 7039 subjects. *Gene, 679*, 133–137. https://doi.org/10.1016/j.gene.2018.08.059

EACLIPT

Clinical Psychology in Europe (CPE) is the official journal of the European Association of Clinical Psychology and Psychological Treatment (EACLIPT).

leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.