Statistical analysis plan

”Exposure Therapy vs Standardized Education for Distress Related to Somatic Symptoms (SOMEX1)” (ClinicalTrials.gov identifier: NCT04942028)

A rudimentary statistical analysis plan (SAP) formed part of the study protocol, approved 2021-05-26, published on ClinicalTrials.org (NCT04942028). This is an amendment based on the study protocol version 2, approved 2021-12-08, and according to Gamble et al (1). This latest and more detailed version of the SAP was completed on 2023-10-05; prior to last recruitment, and prior to the extraction of efficacy data for the primary publication.

Roles and responsibilities

Erland Axelsson[1 2], principal investigator.
Jonna Hybelius[1,2], project manager. Drafted this document under the supervision of EA.
Steven Nordin[3], consultative role.
Sigrid Salomonsson[4], consultative role.
John Wallert[4], consultative role.
Caroline Wachtler[1], consultative role.
Eva Toth-Pal[1 2], consultative role, representative of primary care clinic.
Sandra af Winklerfelt-Hammarberg[1 2], consultative role, representative of primary care clinic.

[1] Division of Family Medicine and Primary Care, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden
[2] Liljeholmen Primary Health Care Center, Region Stockholm, Stockholm, Sweden
[3] Department of Psychology, Umeå University, Umeå, Sweden
[4] Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet & Stockholm Health Care Services, Region Stockholm, Stockholm, Sweden

Background and objectives

If not stated otherwise, the background and research questions below have been derived from the main study protocol.

Approximately one fifth of primary care patients seek care for symptoms that cannot readily be given a medical explanation. In addition, primary care is tasked with offering care for prevalent chronic diseases such as asthma and diabetes, where somatic symptoms often lead to distress and pervasive behavioral changes. Exposure-based treatment has shown promise in reducing somatic symptom load and increasing quality of life in several conditions where patients report distress related to persistent somatic symptoms. In routine care, however, access to such treatment is limited. One reason might be that there exists no flexible exposure-based treatment protocol that can easily be tailored to suit a wide spectrum of patient groups who suffer from persistent somatic symptoms. We have evaluated a unified exposure-based treatment for somatic symptom distress in a previous pilot study and found it feasible to proceed with a randomized controlled trial. Therefore, in this randomized controlled trial (N=160), we aim to test if the same tailored internet-delivered exposure-based treatment is more efficacious than a standardized education control for adult patients with clinically significant distress related to somatic symptoms in a primary care setting.

Primary research question

- Compared to the control condition, does flexible internet-delivered exposure-based treatment lead to a larger improvement in self-rated somatic symptom burden as assessed using the Patient Health Questionnaire 15 (PHQ-15)? Hypothesis: Yes.

Secondary research questions of relevance

- Compared to the control condition, does flexible internet-delivered exposure-based treatment lead to a larger improvement in symptom preoccupation, psychiatric symptom burden, and functional impairment? Hypothesis: Yes.
- In flexible internet-delivered exposure-based treatment, are effects maintained up to 12 months after treatment? Hypothesis: Yes.
- Is the controlled effect of the flexible exposure-based treatment on self-rated somatic symptoms moderated by baseline symptoms and preoccupation? Hypothesis: Yes.

Expanding on the question of moderators, this amendment adds secondary tests of the following potential moderators of the controlled effect of exposure-based treatment on somatic symptom burden:

- Overall somatic symptom burden (pre-treatment[1] PHQ-15).
    - Hypothesis: Positive relation with between-group effect.
        - Secondary test of the PHQ-15 pain subscale
- Overall symptom preoccupation (pre-treatment[1] Somatic Symptom Disorder 12).[2]
    - Hypothesis: Positive relation with between-group effect.
        - Sensitivity analysis based on the 14-item Health Anxiety Inventory
- Depression symptoms (pre-treatment first 2 items of Patient Health Questionnaire 9).
    - Hypothesis: Negative relation with between-group effect.
- Disability (pre-treatment 12-item WHO Disability Assessment Schedule 2).
    - Hypothesis: Positive relation with between-group effect.
- Medically explained vs unexplained somatic symptoms.
    - Hypothesis: Not a moderator.
- Recruited via routine care (referred or listed at clinic) vs other sampling methods.
    - Hypothesis: Not a moderator.
- Age (years at eligibility assessment).[3]
    - Hypothesis: Not a moderator.
- Gender (female vs. male).[3]
    - Hypothesis: Not a moderator.
- Post-secondary education vs. no post-secondary education. Exploratory.
- Years bothered by distress related to persistent somatic symptoms. Exploratory.
- Treatment completer (initiated ≥ 5 modules) vs not treatment completer.
    - Hypothesis: Positive relation with between-group effect.
        - Sensitivity analysis: requirement: ≥ 5 modules + ≥ 3 exposure exercises

[1] Listed as screening in the main protocol and changed to pre-treatment in this amendment, because this is more indicative of the symptoms of the patient just before treatment.

[2] We also intend to analyze symptom-related avoidance behaviors as a potential moderator. This analysis will be based on an instrument under development and will be presented in a secondary publication.

[3] Listed in the main protocol, but not under research questions.

Overall study methods

The study design and main power calculation below are identical to the main study protocol. More detailed a priori power analyses have been added. The paragraph describing the time plan has been updated, and thus constitutes an amendment to the original study protocol.

Trial design

Randomized controlled trial based at Liljeholmen academic primary care clinic, Stockholm, Sweden. Patients are randomized (1:1), in consecutive even numbered cohorts, to internet-delivered exposure-based treatment or an internet-delivered standardized education control.

Sample size and statistical power

The trial is powered to enable the study of moderate between-group effects (d=0.50) as based on mean tests with 80% power, given alpha=0.05 and up to 20% data loss. Based on Monte Carlo simulations using parameter estimates informed by the pilot study (2), power in the primary test will be sufficient to study effects in the lower moderate range (approx. d=0.40; see Table 1).

Table 1. Estimated power of primary test of PHQ-15, based on 1000 Monte Carlo simulations (1000 for each of 12 combinations of hypothetical parameter values, i.e., 12 000 runs in total).

Missing data scenario	Between-group effect		Power
	Difference at post	Cohen's d	
0% missing (N=160 at all assessments)	4.32	0.80	1.000
20% random measurements missing (N=160 only at pre-treatment)	4.32	0.80	1.000
20% completely missing (de facto N=128)	4.32	0.80	1.000
0% missing (N=160 at all assessments)	2.70	0.50	0.978
20% random measurements missing (N=160 only at pre-treatment)	2.70	0.50	0.965
20% completely missing (de facto N=128)	2.70	0.50	0.940
0% missing (N=160 at all assessments)	2.16	0.40	0.870
20% random measurements missing (N=160 only at pre-treatment)	2.16	0.40	0.853
20% completely missing (de facto N=128)	2.16	0.40	0.804
0% missing (N=160 at all assessments)	1.62	0.30	0.665
20% random measurements missing (N=160 only at pre-treatment)	1.62	0.30	0.611
20% completely missing (de facto N=128)	1.62	0.30	0.550

Approximate standardized effects (Cohen's d) based on a presumed standard deviation of 5.4.

Table 2. Estimated power of secondary test of SSD-12, based on 1000 Monte Carlo simulations (1000 for each of 12 combinations of hypothetical parameter values, i.e., 12 000 runs in total).

Missing data scenario	Between-group effect		Power
	Difference at post	Cohen's d	
0% missing (N=160 at all assessments)	10.00	0.80	0.990
20% random measurements missing (N=160 only at pre-treatment)	10.00	0.80	0.990
20% completely missing (de facto N=128)	10.00	0.80	0.969
0% missing (N=160 at all assessments)	6.25	0.50	0.789
20% random measurements missing (N=160 only at pre-treatment)	6.25	0.50	0.786
20% completely missing (de facto N=128)	6.25	0.50	0.722
0% missing (N=160 at all assessments)	5.00	0.40	0.615
20% random measurements missing (N=160 only at pre-treatment)	5.00	0.40	0.601
20% completely missing (de facto N=128)	5.00	0.40	0.503
0% missing (N=160 at all assessments)	3.75	0.30	0.399
20% random measurements missing (N=160 only at pre-treatment)	3.75	0.30	0.392
20% completely missing (de facto N=128)	3.75	0.30	0.326

Approximate standardized effects (Cohen's d) based on a presumed standard deviation of 12.5.

Table 3. Estimated power of continuous moderator tests (three-way interaction of time, condition, and potential moderator) focusing on the SSD-12, based on 1000 Monte Carlo simulations (1000 for each of 20 combinations of hypothetical parameter values, i.e., 20 000 runs in total).

Missing data scenario	Effect of moderator	Incremental d		Power
		Per unit	Per SD	
0% missing (N=160 at all assessments)	Very large effect	0.11	1.00	1.00
20% completely missing (de facto N=128)	Very large effect	0.11	1.00	1.00
0% missing (N=160 at all assessments)	Large effect	0.09	0.80	1.00
20% completely missing (de facto N=128)	Large effect	0.09	0.80	1.00
0% missing (N=160 at all assessments)	Larger moderate effect	0.05	0.50	0.97-0.98
20% completely missing (de facto N=128)	Larger moderate effect	0.05	0.50	0.94
0% missing (N=160 at all assessments)	Smaller moderate effect	0.04	0.40	0.89
20% completely missing (de facto N=128)	Smaller moderate effect	0.04	0.40	0.81-0.82
0% missing (N=160 at all assessments)	Small effect	0.03	0.30	0.66
20% completely missing (de facto N=128)	Small effect	0.03	0.30	0.56-0.57

Approximate standardized effects (Cohen's d) expressed as expected change in the between-group effect on the PHQ-15 (as based on a standard deviation of 5.4), per increase in SSD-12.

Table 4. Estimated power of dichotomous moderator tests (three-way interaction of time, condition, and potential moderator), based on 1000 Monte Carlo simulations (1000 for each of 240 combination of hypothetical parameter values, i.e., 240 000 runs in total).

Missing data scenario	Effect of moderator	Distribution	Incr. d	Power
0% missing (N=160 at all assessments)	Very large effect	50% / 50%	1.00	0.96-0.98
20% completely missing (de facto N=128)	Very large effect	50% / 50%	1.00	0.93-0.95
0% missing (N=160 at all assessments)	Very large effect	33% / 67%	1.00	0.96-0.99
20% completely missing (de facto N=128)	Very large effect	33% / 67%	1.00	0.92-0.95
0% missing (N=160 at all assessments)	Large effect	50% / 50%	0.80	0.85-0.89
20% completely missing (de facto N=128)	Large effect	50% / 50%	0.80	0.77-0.81
0% missing (N=160 at all assessments)	Large effect	33% / 67%	0.80	0.86-0.90
20% completely missing (de facto N=128)	Large effect	33% / 67%	0.80	0.78-0.82
0% missing (N=160 at all assessments)	Larger moderate effect	50% / 50%	0.50	0.46-0.54
20% completely missing (de facto N=128)	Larger moderate effect	50% / 50%	0.50	0.40-0.45
0% missing (N=160 at all assessments)	Larger moderate effect	33% / 67%	0.50	0.49-0.53
20% completely missing (de facto N=128)	Larger moderate effect	33% / 67%	0.50	0.40-0.45
0% missing (N=160 at all assessments)	Smaller moderate effect	50% / 50%	0.40	0.33-0.36
20% completely missing (de facto N=128)	Smaller moderate effect	50% / 50%	0.40	0.28-0.32
0% missing (N=160 at all assessments)	Smaller moderate effect	33% / 67%	0.40	0.32-0.38
20% completely missing (de facto N=128)	Smaller moderate effect	33% / 67%	0.40	0.26-0.32
0% missing (N=160 at all assessments)	Small effect	50% / 50%	0.30	0.20-0.23
20% completely missing (de facto N=128)	Small effect	50% / 50%	0.30	0.16-0.21
0% missing (N=160 at all assessments)	Small effect	33% / 67%	0.30	0.21-0.24
20% completely missing (de facto N=128)	Small effect	33% / 67%	0.30	0.17-0.21

Approximate standardized effects (Cohen's d) expressed as expected change in the between-group effect on the PHQ-15 (as based on a standard deviation of 5.4).

Time plan

We expect the last follow-up assessment to take place around the end of 2024 or early 2025. The analysis of most primary and secondary outcomes for the primary publication is scheduled to take place in mid 2024 to early 2025. SAPs for the cost-effectiveness analysis and mediation analysis will be published separately. This document concerns the primary publication only.

Statistical principles

The paragraph outlining the criteria for treatment adherence and the planned methods for reporting adherence to treatment in the scientific article is an amendment to the original study protocol. Apart from this, the statistical principles stated below are in accordance with the original protocol.

Confidence intervals and p-values

All applicable statistical tests will be performed using a 5% significance level. As is common practice in the field, whenever directional hypotheses are formulated, two-tailed 95% tests and confidence intervals will nevertheless be employed - thus effectively corresponding to one-sided 97.5% equivalents.

Adherence to the protocol

Adherence will be reported as the mean (SD) and median number of initiated treatment modules in the exposure and control group. Completion of the treatment or the control (dichotomous outcome: yes vs. no) is defined as having initiated at least 5 out of 10 modules. In sensitivity analyses, a more strict criterion for completion of the main treatment will also be explored in terms of having initiated at least 5 modules and also conducted at least 3 exposure exercises.

Analysis populations

The primary analysis will adhere to the intention-to-treat principle, meaning that data from all patients including dropouts will be used. In other words, the outcome will be analyzed in accordance with assigned groups. See the paragraph concerning handling of missing data below. As is described above, a secondary analysis will be concerned with completion as a moderator, and based on this model per-protocol estimates will also be reported.

Trial population

Here follows an expanded description of how data will be utilized to characterize the study flow and sample characteristics, intended to be read in conjunction with the original protocol.

Recruitment data and study flow

The time period of recruitment will be reported as the dates of first and last enrollment. The date of the last follow-up will also be reported. A CONSORT flowchart will summarize the number of applicants who provided informed consent and completed the screening, who completed the eligibility assessment, who met all eligibility criteria and were enrolled and randomized, who completed each respective allocated condition (see above), who dropped out (either explicit or at least 3 weeks of inactivity without resuming treatment), and who completed the post- and follow-up assessments. Reasons will be listed for why applicants were not enrolled post-screening.

Eligibility

Patients (N=160) who report being bothered by somatic symptoms (and who may or may not meet criteria for a diagnosis where distress related to symptoms is common, such as fibromyalgia or asthma), who express interest in psychological treatment, and who's medical status does not make exposure therapy unsuitable are potentially eligible for treatment. Please consult the study protocol for a complete report of inclusion and exclusion criteria.

Dropout rates
Dropouts are reported in terms of patients who stopped replying for at least 3 weeks without resuming treatment, or explicitly expressed a wish to discontinue their treatment and then did so. The primary reasons for dropouts will be reported, as rated post treatment by the therapist.

Baseline patient characteristics
Dichotomous and categorical variables will be summarized by frequencies and percentages. Continuous variables will be summarized using the mean, SD, median and range. Tests of statistical significance will not be undertaken for baseline characteristics. Rather, the clinical importance of any imbalance will be noted. The following baseline characteristics will be reported: age, gender, level of education, occupational status, duration of somatic symptom distress in years, whether symptoms are considered medically explained or not, somatic diagnoses, psychiatric diagnoses, recruitment path (routine care vs. not routine care), previous experience of psychological treatment (yes vs. no), overall somatic symptom burden (PHQ-15) and subdomains of somatic symptoms, symptom preoccupation (SSD-12), depression (PHQ-9), functional impairment (WD2-12), and psychotropic medication.

Therapist characteristics
Therapists are balanced over the two treatment arms. Therapist characteristics are reported in terms of profession, patients treated, years of working with psychological treatment, and years of working with persistent somatic symptom distress on a regular basis.

Analysis
The rudimentary statistical analysis plan outlined in the main study protocol has been expanded upon below. In cases where the main protocol is not referenced, the information provided is novel and specific to this amendment.

Efficacy outcomes
As stated in the main study protocol, the primary outcome is the effect on somatic symptom burden up to the post-treatment assessment, as measured using the Patient Health Questionnaire 15 (PHQ-15) (3). Secondary efficacy outcomes are listed in the main protocol and include symptom preoccupation, general anxiety, depression, and disability. Exhaustive mediation analyses will be presented as part of a secondary publication, and will be described elsewhere.

Method of analysis
Inferential analyses will be conducted by a person who is blind to treatment condition. As previously stated in the main study protocol, change in the primary outcome measure and other continuous scales will be analyzed using multilevel regression models (4) with patient at level 2, using an autoregressive (AR(1)) covariance structure, fitted by maximum likelihood. Longitudinal outcomes will be modelled intention-to-treat, with time, group (exposure vs. control), and the time×group interaction as predictors (α=5%). Because the feasibility trial was indicative of a curvilinear effect of time (unusual in the field), the inclusion of a time×time interaction will also be considered based on model fit (Bayesian Information Criterion, BIC).

Individual differences in baseline scores will be incorporated into the multilevel regression models as a random effect (i.e., random intercept). A random effect of change over time (i.e., a random slope) will also be included if convergence can be achieved.

As described in the main protocol, the primary statistical test is that of the coefficient for the time×group interaction pertaining to a group difference in average change in the PHQ-15 over the main phase of the trial, as fitted on multiply imputed data. Standardized mean effects are conceptualized as the time×group interaction divided by the endpoint standard deviation. Long-term effectiveness is modeled using piecewise regression models, with a spline at post-treatment.

In the original analysis plan, we specified that response rates would be based on the criterion for clinically significant improvement (CSI) (5). Because the CSI is intended to signify change to a reliable extent, and from a clinical pre-treatment score to a healthy post-treatment score, the CSI will be reported for the subsample of patients who scored at least in the moderate range on the PHQ-15 (≥10) before treatment. This will be based on test-retest reliabilities of r=0.93 (6) and r=0.65 (7), and scores below 10 (in the low range) will be considered healthy (3). Because the CSI is not relevant for patients scoring low in the PHQ-15 before treatment, and also because the proportion of reliable change, which is a requirement for CSI, is a systematic underrepresentation of the proportion of *actual* change (8), response rates will also be reported as the proportion of all patients who saw improvement equal to or beyond the minimally important difference (MID) (9).

In accordance with the main protocol, auxiliary analyses will concern moderators (see page 2) and test the three-way interaction between time, group, and the potential moderator. Pearson correlations < 0.60 will be considered indicative of the moderators being relatively distinct and suitable for analysis. Models will include a random intercept and slope if convergence can be achieved.

In additional exploratory secondary analyses, treatment effects will also be reported for substrata:

- As stated in the main protocol, the primary models for symptom burden (PHQ-15) and symptom preoccupation (SSD-12) will be refitted for each somatic diagnosis present in at least 5 patients.
- As stated in the main protocol, the PHQ-15 will be scored in terms of cardiopulmonary, fatigue, gastrointestinal, and pain subscales based on Witthöft et al. (10). Separate analyses will then be conducted for each of these subscales, first including all patients, and then including all who scored an average of at least 1 on a 0-2 scale, on the items belonging to that scale.
- The primary models for symptom burden (PHQ-15) and symptom preoccupation (SSD-12) will be refitted for patients with vs. without somatic symptom disorder (SSD).

There will be a test of the hypothesis that men are more inclined than women to seek health care for somatic symptoms when they are in fact suffering from a non-somatoform psychiatric condition, as based on a 2x2 $\chi 2$ test of exclusions due to other principal psychiatric conditions vs. other factors, men vs. women.

Missing data

As stated in the main study protocol, missing data will be imputed using hierarchical multiple imputation by chained equations in 100 datasets (revised from 20) using the predictive mean matching method (11,12). Multiple imputation will be based on the following predictors: time, age, gender, pre-treatment somatic symptoms, pre-treatment symptom preoccupation,

treatment adherence, and missing data rate. In addition, the following now specified based on pre-treatment data: disability, depression, general anxiety, pain intensity, post-secondary education, years bothered by distress related to persistent somatic symptoms, medically explained vs unexplained symptoms, and recruitment strategy (routine care vs. not routine care). Imputation will be done separately for each condition (exposure vs. control) in order to maintain condition-specific relationships in the data.

Harms and unwanted effects

The following description of how to report adverse events and deterioration constitutes an amendment to the original SAP. Adverse event (AE) data are collected at the post-treatment assessment. Patients report whether they have experienced any form of AE that potentially could be related to participation in the study. Up to three AEs can be reported, and for each AE, patients are encouraged to provide details in free-form text. For the purpose of the primary publication, answers are read by a person who is blind to treatment condition, and categorized according to the six common categories of AEs identified in a previously published factor analysis (13). The number and proportion of patients having experienced at least one AE, and at least one AE per type, will be reported separately for each treatment arm. The incidence of serious AEs defined as warranting immediate medical attention, or leading to hospitalization or death will also be reported. The proportion of minimal important deterioration is also presented, based on the MID.

Statistical software

The analysis will be carried out using R version 4.3.1 or later, and Stata version 15 or later.

References

1. Gamble C, Krishan A, Stocken D, Lewis S, Juszczak E, Doré C, et al. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. JAMA. 2017;318(23):2337-43.

2. Hybelius J, Gustavsson A, af Winklerfelt Hammarberg S, Toth-Pal E, Johansson R, Ljótsson B, Axelsson E. A unified Internet-delivered exposure treatment for undifferentiated somatic symptom disorder: single-group prospective feasibility trial. Pilot Feasibility Stud. 2022;8(1):149.

3. Kroenke K, Spitzer RL, Williams JBW. The PHQ-15: Validity of a New Measure for Evaluating the Severity of Somatic Symptoms. Psychosom Med. 2002;64(2):258–66.

4. Gueorguieva R, Krystal JH. Move Over ANOVA: Progress in Analyzing Repeated-Measures Data andIts Reflection in Papers Published in the Archives of General Psychiatry. Arch Gen Psychiatry. 2004;61(3):310–7.

5. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol. 1991;59(1):12-9.

6. Liao SC, Huang WL, Ma HM, Lee MT, Chen TT, Chen IM, et al. The relation between the patient health questionnaire-15 and DSM somatic diagnoses. BMC Psychiatry. 2016;16(1):351.

7. Han C, Pae CU, Patkar AA, Masand PS, Kim KW, Joe SH, et al. Psychometric Properties of the Patient Health Questionnaire-15 (PHQ-15) for Measuring the Somatic Symptoms of Psychiatric Outpatients. Psychosomatics. 2009;50(6):580–5.

8. Axelsson E, Hedman-Lagerlöf E. Unwanted outcomes in cognitive behavior therapy for pathological health anxiety: a systematic review and a secondary original study of two randomized controlled trials. Expert Rev Pharmacoecon Outcomes Res. 2023;1–15.

9. Toussaint A, Kroenke K, Baye F, Lourens S. Comparing the Patient Health Questionnaire - 15 and the Somatic Symptom Scale - 8 as measures of somatic symptom burden. J Psychosom Res. 2017;101:44–50.

10. Witthöft M, Hiller W, Loch N, Jasper F. The latent structure of medically unexplained symptoms and its relation to functional somatic syndromes. Int J Behav Med. 2013;20(2):172–83.

11. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. J Stat Softw. 2011;45(3).

12. Rubin, D. B. Multiple imputation for nonresponse in surveys. 1987.

13. Rozental A, Kottorp A, Boettcher J, Andersson G, Carlbring P. Negative Effects of Psychological Treatments: An Exploratory Factor Analysis of the Negative Effects Questionnaire for Monitoring and Reporting Adverse and Unwanted Events. Plos One. 2016 Jun 22;11(6):e0157503.