Study protocol and statistical analysis plan
Date: Apr 3, 2025

Study title:
Hypnotherapy With Audiofiles for Children and Adolescents With Disorders of Gut-brain Interactions in Sweden -a Feasibility Study

Scientific Summary

Background:

Functional gastrointestinal disorders (FGID) with abdominal pain, such as Irritable Bowel Syndrome (IBS), affect 10–25% of all children and adolescents in Sweden. These conditions are often associated with high school absenteeism, reduced quality of life, and frequent healthcare use. For children and adolescents with IBS, various psychosocial treatments are the only interventions that have shown good effect in scientific studies, while pharmacological treatments have had limited or no effect. In Sweden, the most commonly used and studied treatment for this patient group is exposure-based Cognitive Behavioral Therapy (CBT), which is also available as internet-delivered CBT (iCBT) in standard healthcare. However, iCBT requires therapist support from a psychologist, resulting in long wait times, and not all patients benefit. Therefore, additional evidence-based and preferably more accessible alternatives are in demand.

Hypnotherapy involving gut-directed meditation and relaxation exercises has shown good treatment effects for FGID in children and adolescents in international studies, but this form of treatment is currently not available in Sweden. Studies have also shown that home-based hypnotherapy with self-guided exercises via audio files is as effective as individual hypnotherapy, making it a cost-effective treatment. Audio files allow treatment to begin immediately after diagnosis by a physician, avoiding long wait times like those for iCBT. If hypnotherapy proves to be as effective in Swedish standard care as in international studies, it would represent a major advancement for the large and suffering group of children and adolescents with FGID.

Aim:

In an international research collaboration, we intend to translate and adapt hypnotherapy audio files — proven highly effective in a Dutch context — to Swedish conditions so that they can be used to treat children and adolescents in Sweden. The feasibility study described in this project plan is required to evaluate effect sizes and the practical aspects of the treatment. Our experience from previous treatment studies is that the insights gained from this type of feasibility study provide the knowledge needed to conduct large and successful randomized controlled trials (RCTs) in the next phase — for example, comparing hypnotherapy and iCBT in terms of effectiveness, cost-efficiency, and which patients respond best to which treatment. The overarching goal of the project is to help children and adolescents with FGID by ultimately introducing a new treatment option into standard care.

Feasibility Study Research Questions:

1. Can children and adolescents with pain-dominant FGID experience reduced gastrointestinal symptoms through hypnotherapy with audio files?
2. Can children and adolescents with pain-dominant FGID experience improved daily functioning through hypnotherapy with audio files?
3. Can children and adolescents with pain-dominant FGID experience improved quality of life through hypnotherapy with audio files?
4. What practical challenges may arise in patient recruitment, data collection, and the treatment itself (that need to be addressed before a larger RCT can be conducted)?
5. **Feasibility Study Method:**
   A feasibility study of the hypnotherapy intervention without a control group to evaluate feasibility, acceptability, and effect sizes for the primary outcome measures.

Children (N=20–30) of school age (8–17 years) with pain-dominant FGID (defined according to the Rome IV criteria) will be recruited from a pediatric outpatient clinic in Stockholm. Measurements of, among other things, gastrointestinal symptoms, functional level, and quality of life will be conducted using validated instruments before, during (every three weeks), and after treatment.

6. **Significance of the Feasibility Study:**
   The feasibility study will provide estimates of effect sizes for the outcome measures (required for power analysis) and offer invaluable practical experience that will make it possible to later carry out a successful large-scale clinical trial.

Project Description:

The first part of the project (which this application concerns) consists of a feasibility study in which the applicability, preliminary effectiveness, and mechanisms of action of the hypnotherapy treatment will be evaluated in a smaller patient group. The goal is to recruit 20–30 patients with a confirmed FGID diagnosis from the pediatric gastroenterology clinic at Sachsska Children's Hospital and/or other hospital-based specialist clinics at Sachsska. These patients will undergo a 12-week treatment involving hypnotherapy self-exercises using audio files at least 5 times per week. Patients and their parents will complete self-report assessments on symptom severity, quality of life, school absenteeism, etc., using standardized questionnaires via an online platform every three weeks. A table listing the outcome measures is attached.

All patients will be contacted by a gastroenterology-focused physician or nurse at the clinic or by phone (if referred from another unit) before inclusion to assess inclusion and exclusion criteria, inform the family about the study, and go over the treatment principles.

Inclusion criteria:

- Age 8–17 years
- Confirmed diagnosis of functional dyspepsia, functional abdominal pain, or IBS according to Rome IV criteria (including evaluation for organic differential diagnoses)
- Reading and writing proficiency in Swedish
- Less than 40% school absenteeism
- Stable dose (for over 1 month) if on psychopharmacological medication

Exclusion criteria:

- Non-functional medical condition
- More than 40% school absenteeism
- Ongoing psychotherapeutic treatment
- Severe psychosocial or psychiatric problems requiring more intensive care than hypnotherapy

These criteria will be assessed using questionnaires and individual interviews before treatment begins.

Data Analysis:
For the feasibility study, 20–30 patients will be recruited — the number is based on prior experience with similar pilot studies. Data will be collected every three weeks from the start to the end of treatment via a secure online service with identity verification. The same platform will be used to distribute treatment files and instructions to patients and their parents. Data will be stored in a secure database protected by two-factor authentication and access control. Once the data is no longer needed for the study, the key to the code will be destroyed (no later than 10 years after data collection).

An improvement of >30% on the primary outcome is defined as clinically significant, based on recommendations from previous studies. Although data will be collected across multiple outcomes and time points, a preliminary mediation analysis can also be conducted to assess what drives potential improvements and in which phase of treatment this occurs. Statistical analyses will be conducted by a statistician with experience in similar studies and analyses.

Data Analysis Plan:

1 Methods

The planned RCT shares many similarities with Bonnert et al. (2017) and Bonnert et al. (2014).

1.1 Mixed models

We let $\boldsymbol{y}_i=(y_{i1},...,y_{id})$ denote the response vector of d repeated measurements for individual $i=1,...,n$. For example, when modeling Peds-QL gastro the observed value y_{ij} in vector $\boldsymbol{y}_i$ represents the observed value of Peds-QL gastro at week j for patient i. In a GLMM, the conditional expected value of y_{ij} is given by

$$g[E(y_{ij}|\boldsymbol{u}_i)]=\boldsymbol{x}_{ij}\boldsymbol{\beta}+\boldsymbol{z}_{ij}\boldsymbol{u}_i,$$

where g is the link function, $\boldsymbol{x}_{ij}$ and $\boldsymbol{z}_{ij}$ are vectors of observed values of explanatory variables, the parameters β are the fixed effects, and u_i are the random effects. The LMM is a special case of GLMM for which g is the identity link function

$$E(y_{ij}|\boldsymbol{u}_i)=\boldsymbol{x}_{ij}\boldsymbol{\beta}+\boldsymbol{z}_{ij}\boldsymbol{u}_i.$$

1.1.1 Model specification

For each outcome variable with more than two measurements per patient (i.e., not only at week 0 and at week 12), we estimate the pre-post treatment effect using an LMM fitted by restricted maximum likelihood (REML). The model can be written as

$$y_{ij}=\beta_0+\beta_1\mathrm{WEEK}_{ij}+u_i+\epsilon_{ij},$$

where y_{ij} represents the value of the outcome for patient i at week j, β_0 is the fixed intercept, β_1 is the fixed slope coefficient that captures the average change of the outcome variable given one more week of treatment, and u_i is the patient specific random intercept. By using random intercepts we allow patients to differ at baseline. For example, some patients might have much higher Peds-QL gastro than others at baseline.
Due to the small sample size the decision was made to not include quadratic effects or random slopes as it cause convergence issues. However, normally these should be included if they

improve the goodness of fit of the model (based on i.a. the distribution of residuals, likelihood ratio tests, or the Akaike information criterion). A model that includes both quadratic effects and random slopes can be written as

$$y_{ij}=\beta_0+\beta_1\text{WEEK}_{ij}+\beta_2\text{WEEK2}_{ij}+u_{0i}+u_{1i}\text{WEEK}_{ij}+\epsilon_{ij}=\beta_0+u_{0i}+(\beta_1+u_{1i})\text{WEEK}_{ij}+\beta_2\text{WEEK2}_{ij}+\epsilon_{ij}.$$

By including a quadratic effect for weeks of treatment we account for nonlinear trends. For example, it might be that the effect of the treatment is decreasing over time. Furthermore, by including random slopes we allow for the weekly effect of the treatment to differ between patients.

For parent rated outcomes, we should also consider using a 3-level mixed model were we cluster the data by both parent and adolescent.

1.1.2 LMM estimate of Cohen's d

To obtain a standardized pre-post treatment effect (Cohen's d), the LMM estimate of the 12-week treatment effect is divided (standardized) by the standard deviation at baseline. That is, the LMM estimate of Cohen's d is defined as

$$d=12\times\beta_1 SD_{week=0}.$$

The confidence interval for Cohen's d is based on the LMM estimated standard errors (SE), that is

$$SE_d=12\times SE_{\beta_2} SD_{week=0}.$$

Standard errors can also be computed using bootstrap. This is particularity useful in case of a complex model for which Cohen's d is estimated using more than one parameter For example, if the model includes a quadratic treatment effect, then Cohen's d is estimated as follows

$$d=12\times\beta_1+12_2\times\beta_2 SD_{week=0}.$$

The standard error of Cohen's d with a quadratic treatment effect can also be calculated using the formula $Var(aX+bY)=a_2Var(X)+b_2Var(Y)+2abCov(X,Y)$, where $X=\beta_1$ and $Y=\beta_2$. However, since multicollinearity can be a problem in polynomial regression which can lead to inflated standard errors, the variable for weeks of treatment should first be centralized.

1.2 Cohen's d for paired data

Since the outcome variables for depression (CDI-S), anxiety (SCAS-S-C), quality of life of parents, and school absence only have two measurements per patient at week 0 and week 12, the decision was made to estimate the effect size for these variables using Cohen's adjusted d for paired data together with Hedges' small-sample bias correction, see Borenstein et al. (2009) . It is important to note that as this estimator requires paired data only patients with an observed value at both week 0 and at week 12 can be included.

We let $(y_{i,week=0},y_{i,week=12})$ denote a pair of observed values for some outcome variable Y at week 0 and 12 for patient i among n pairs with $d_i=y_{i,week=0}-y_{i,week=12}$. Then Cohen's d is defined as

$$d=\bar{Y}_{diff}S_{within}=\bar{Y}_{diff}S_{diff}/2(1-r)\sqrt{}$$

where Y_{diff} and S_{diff} is the mean and standard deviation of d_i, and r is the estimated correlation between weeks 0 and 12. The variance of Cohen's d is given by

$$V_d=(1n+d_2 2n)2(1-r).$$

To get the Hedges' corrected d, denoted g, we use the correction factor J as follows

$$g=J\times dV_g=J_2\times V_d$$

where

$$J=1-34df-1$$

with the degrees of freedom $df=2n-2$.

1.3 Paired T-test

The paired T-test can be used to test the null hypothesis that the pre-post mean difference is equal to zero. The test requires paired data and shares strong similarities with Cohen's adjusted d for paired data. Letting Y^{-}_{diff} denote the mean difference and S_{diff} denote the standard deviation of the difference, the test statistic is defined as

$$t=Y^{-}_{diff}S_{diff}/n\sqrt{}.$$

1.4 Unpaired Two-sample T-test

To compare the effect of the hypnotherapy treatment between two age groups at week 12 we can perform a two-sample T-test assuming unequal variances. The test statistic is calculated as

$$t=y^{-}1-y^{-}2s21n1+s22n2\overline{\qquad}\sqrt{}$$

where $y^{-}1$ and $y^{-}2$ represent the mean values, $s21$ and $s22$ represent the estimated variances, and $n1$ and $n2$ the sample sizes for age group 1 and 2 at week 12. Note that the T-test only uses data at week 12. Another option is to include an interaction term between age group and weeks of treatment in the mixed model in order to compare the pre-post treatment effect between the groups.
Seeing that our sample size is very small, we should avoid making to many tests as it could be viewed as P-value hacking.

1.5 Imputation of missing values

The pilot study includes 32 patients at baseline (week 0). Of these patients, there are 21 patients with complete follow-up (no missing data), 25 patients with no missing data at week 12, and 2 patients with no data after week 0. Data is rarely missing completely at random (MCAR). It is more likely to assume that there are systematic differences between the non-participating patients (i.e. the patients who at some point decided to drop out) and the participating patients, which could bias the results. For example, some patients might have discontinued the treatment due to insufficient treatment effect. If that is the case, there is a risk that the estimated effect of the treatment could be overestimated if the missing values are not handled correctly. For this reason, a sensitivity analysis was made, comparing the results of using no imputation to the results of using Last Observation Carried Forward (LOCF). A drawback of LOCF is that it underestimates the variability in the data, which can lead to too narrow confidence intervals. Another alternative is to use multiple imputation, which account for the uncertainty of estimating the missing data. However, if the observed data cannot account for the systematic differences, that is if the data is missing not at random (MNAR), multiple imputation may still lead to biased results.

1.6 Limitations

No control group: It is important to note that since we do not have a control group to compare with, the risk of receiving biased results is high, especially due to the phenomenon known as regression toward the mean. That is, many of the patients who were in poor health at baseline (week 0) would probably have felt better at week 12 even without treatment.

No analysis of outliers: No analysis of outliers has been performed.

No comparison of models: Due to the small sample size, the decision was made to only fit an LMM with patient specific random intercepts. However, a larger sample size would enable us to us more complex models that can provide better fit to the data. In a more complex model we can for example include quadratic effects, random slopes, or interaction terms. For parent rated outcomes, we should also consider using a 3-level mixed model were we cluster the data by both parent and adolescent.

Modeling discrete variables as continuous variables: The outcome variable for pain frequency represents the number of days during a week that the individual felt pain, which means that it is a discrete variable with a minimum value of zero and a maximum value of seven. In this analysis we will model pain frequency as if it was a continuous variable using LMM for ease of interpretation. However, count data is usually modeled using Poisson regression.