Modeling and Sample Size Planning in Intensive Longitudinal Intervention Study Design–Based on Dynamic Structural Equation Models

Authors: Liu Yue, He Yueling, Liu Hongyun, Liu Hongyun

Date: 2025-10-17T21:01:25+00:00

Abstract

Intensive longitudinal intervention research possesses advantages such as high ecological validity and the capability to provide real-time and personalized interventions. However, currently prevalent data analysis methods fail to adequately reflect the characteristics of intensive longitudinal data, while advanced data analysis models lack matched sample size planning methodologies, which greatly limits the widespread adoption and application of this paradigm. This paper conducts sample size planning based on dynamic structural equation modeling under two typical intensive longitudinal intervention experimental designs –single-arm design and randomized controlled design–by incorporating statistical power and effect size estimation accuracy through simulation research methods, performs a comprehensive comparison of the two designs from perspectives including Type I error rates, and finally proposes recommendations for experimental design and sample size planning.

Full Text

Data Analyses and Sample Size Planning for Intensive Longitudinal Intervention Studies with Dynamic Structural Equation Modeling

LIU Yue[1], HE Yueling[1], LIU Hongyun[2],[3]

[1]Institute of Brain and Psychological Sciences, Sichuan Normal University, Chengdu 610066, China
[2]Beijing Key Laboratory of Applied Experimental Psychology, Beijing Normal University, Beijing 100875, China
[3]Faculty of Psychology, Beijing Normal University, Beijing 100875, China

Abstract

Intensive longitudinal interventions (ILIs) offer high ecological validity and enable real-time, personalized treatment delivery. However, conventional data analysis methods fail to capture the distinctive features of intensive longitudinal data, while advanced analytical models lack corresponding sample size planning procedures—severely limiting the widespread adoption of this paradigm. This study addresses this gap by examining sample size planning for two prevalent ILI designs—single-arm trials and randomized controlled trials—using dynamic structural equation modeling (DSEM). Through Monte Carlo simulation, we evaluate statistical power and effect size estimation accuracy, present sample size recommendations via credible interval width contour plots, and compare the two designs in terms of Type I error rates. Finally, we provide practical guidance for experimental design and sample size determination.

Keywords: intensive longitudinal intervention, dynamic structural equation modeling, power analysis, effect size, sample size planning

---

1. Introduction

1.1 The Development of Intensive Longitudinal Intervention Research

As psychological research questions have become increasingly sophisticated, traditional laboratory-based intervention studies face challenges such as low ecological validity (Balaskas et al., 2021). Consequently, intensive longitudinal intervention has emerged as a prominent paradigm in mental health research (Chen & Zhou, 2017; Schueller et al., 2017). Intensive longitudinal intervention, also known as ecological momentary intervention (EMI), refers to the delivery of real-time psychological or behavioral disorder treatments during daily activities to help participants increase the frequency of healthy psychological states and behaviors (Heron & Smyth, 2010). In terms of intervention content, ILIs can target both discrete behaviors (e.g., addictive behaviors) and continuous daily emotions (e.g., depression) or physiological states. Regarding data collection protocols, ILIs can be implemented as standalone interventions without concurrent intensive tracking measurements (e.g., Reininghaus et al., 2024), or they can integrate intensive tracking measurements to provide continuous real-time monitoring and therapeutic support (Bell, 2018).

Intensive longitudinal intervention originated from intensive longitudinal studies (ILS), which commonly employ ecological momentary assessment (EMA), experience sampling methods (ESM), or daily diaries to evaluate life experiences in naturalistic settings. This approach enables repeated, real-time assessment of individuals' daily cognition, emotion, and behavior, offering high ecological validity and minimal recall bias. As a result, the number of ILS has grown dramatically in recent years (e.g., Wilhelm et al., 2012). Researchers subsequently extended the ILS paradigm to clinical intervention, using data collected from in-

tensive tracking measurements to deliver real-time, more targeted interventions in daily life. Heron and Smyth (2010) first explicitly proposed the concept of ecological momentary intervention (EMI). With the proliferation of smartphones, mobile applications have become the most common medium for implementing intensive longitudinal interventions (Schueller et al., 2017).

Internationally, this intervention approach has been widely applied to treat anxiety, depression, and other psychological disorders, as well as to promote healthy behaviors such as physical exercise and to support cessation of addictive behaviors like smoking (Smith & Juarascio, 2019), all demonstrating promising effectiveness. Although intensive longitudinal intervention research in China is still in its early stages, it shows a positive development trajectory (Zhang et al., 2021). Furthermore, with increasing demands for personalized interventions and rapid advancements in optimization algorithms, intensive longitudinal interventions have evolved into just-in-time adaptive interventions (JITAI; Schueller et al., 2017), emphasizing dynamic adjustment of intervention content based on real-time data in natural contexts. This development trend necessitates collecting sufficient intensive tracking measurement data before and after interventions to reflect intervention effects multidimensionally and provide support for continuous optimization and individualized feedback (Cuijpers et al., 2021). Overall, despite growing practical demands for intensive longitudinal intervention research, systematic methodological investigations on how to achieve study design through sample size planning and how to conduct multidimensional evaluation of intervention effects based on data analysis are lacking both domestically and internationally, significantly limiting the scientific development of this field.

The two typical experimental designs in traditional intervention research—single-arm trials (SAT) and randomized controlled trials (RCT)—can be directly extended to intensive longitudinal intervention research (Yi, 2020). SAT refers to a design where all participants receive the same intervention simultaneously, with intervention effects evaluated by comparing intensive tracking measurements before and after the intervention (pre- and post-intervention phases). This design is primarily used for pilot studies with small samples before large-scale RCTs (Baey & Le Deley, 2011) or for studies with hard-to-recruit populations (e.g., elderly individuals, see Mair et al., 2022; patients with severe depression, see Rauschenberg et al., 2021; problem gambling populations, see Hawker et al., 2021; sexual risk behavior populations, see Shrier & Spalding, 2017). RCT refers to a design where participants are randomly assigned to either an intervention group (experimental group) or a control group, with the intervention group receiving the treatment and the control group receiving no treatment or a standard treatment (e.g., placebo). Intervention effects are evaluated by comparing differences in intensive tracking measurements between the two groups during the pre- and post-intervention phases (e.g., Bell, 2018). This design offers two clear advantages: first, the control group effectively separates natural changes from external confounding factors, demonstrating rigorous causal inference of intervention effects; second, random assignment reduces selection bias and ensures comparability between intervention and control groups during the

pre-intervention phase, better satisfying the prerequisites for causal inference.

summarizes the basic characteristics of 36 empirical intensive longitudinal intervention studies in psychology from 2015-2025 (detailed search procedures are provided in the appendix at https://doi.org/10.57760/sciencedb.psych.00506). The results show that single-arm and randomized controlled designs are the two primary experimental designs, with the majority of studies (69.4%) adopting randomized controlled designs.

Table 1 Summary of Psychology Intensive Longitudinal Intervention Studies (2015-2025, $N = 36$)

Design Type	Count (%)	Sample Size Determination Method
Single-arm design	9 (25.0%)	Variance analysis, t-test, chi-square test, regression analysis
Randomized controlled design	25 (69.4%)	Thematic analysis, linear mixed-effects model, dynamic structural equation model, power analysis, previous studies or pilot studies, practical resource considerations
Microrandomized trials	2 (5.6%)	

Note: Microrandomized trials refer to experimental designs where participants are randomly assigned to receive or not receive intervention at each repeated measurement occasion.

1.2 Modeling Intensive Longitudinal Intervention Research

Existing data analysis methods in intensive longitudinal intervention research primarily include linear mixed-effects models, ANOVA, t-tests, and regression analysis (see Table 1). However, these methods suffer from three significant limitations. First, they ignore individual differences in intervention effects. Traditional methods such as ANOVA and t-tests fail to account for the hierarchical structure where measurement time points are nested within participants, overlooking potential individual differences in intervention effects and consequently leading to biased statistical conclusions (Hoffman & Walters, 2022). Additionally, the homogeneity of residual variance assumption in mixed-effects models is likely violated in practice, and heterogeneity in within-person variability poses greater challenges for data analysis (Hedeker et al., 2008). Second, they neglect the autocorrelation structure of time series data. Outcome variables of interest are measured repeatedly, exhibiting time series characteristics, yet most analytical methods do not incorporate autocorrelation, resulting in biased parameter estimates and erroneous statistical inferences (Kenny & Judd, 1986). Third, they focus exclusively on intervention effects reflected in the mean of outcome variables. Existing methods use the outcome variable itself as the dependent

variable in regression models, failing to adequately incorporate the dynamic developmental characteristics of variables and making it difficult to provide evidence for intervention effectiveness from multiple perspectives.

Therefore, to integrate the distinctive features of intensive longitudinal intervention research and conduct comprehensive, multidimensional evaluations of intervention effects, it is necessary to analyze data using dynamic structural equation modeling (DSEM), a commonly employed method in intensive longitudinal research. DSEM integrates time series modeling, structural equation modeling, and linear mixed-effects modeling (Asparouhov et al., 2018). In analyzing intensive longitudinal intervention data, DSEM offers two main advantages: first, it fully captures the nested structure and time series characteristics inherent in the data; second, it enables comprehensive evaluation of intervention effects through model parameters, including not only changes in the mean of outcome variables but also changes in inertia and intra-individual variability (IIV) (Sherwood, 2022). For example, reduced inertia in negative emotions reflects individuals' ability to disengage from negative emotional states, while increased inertia in positive emotions reflects individuals' capacity to maintain positive emotional states—these can be examined through changes in autocorrelation coefficients between pre- and post-intervention phases (Hamaker et al., 2021). Moreover, IIV can reflect cognitive functioning and serves as an important predictor of many physical and mental disorders (Aschenbrenner et al., 2024). Changes in IIV can be examined through changes in within-person residual variance between pre- and post-intervention phases. Consequently, researchers have extended DSEM models for intensive longitudinal intervention research. For instance, Hamaker et al. (2021) developed a DSEM model within the RCT framework to evaluate intervention effects across three dimensions: mean, inertia, and IIV; Li et al. (2024) combined autoregressive models with latent growth curve models to examine changes in mean, autoregression, and IIV across different phases. However, their methods impose high demands on data collection designs, requiring more than two phases of intensive tracking data to obtain parameter estimates related to growth trends. Furthermore, the strong assumption of linear growth (or decline) may not align with reality.

1.3 Sample Size Planning for Intensive Longitudinal Intervention Research

Currently, only a handful of studies (see Table 1, only 2) have applied DSEM to analyze intensive longitudinal intervention data. The application of advanced statistical modeling methods, as well as the replicability and generalizability of results, depends on the development of matching sample size planning methods. As academic discussions about research misconduct and the reproducibility crisis in psychology intensify, an increasing number of journals and institutions advocate for the rational application of sample size planning methods to scientifically determine design elements such as participant numbers and measurement time points. This approach avoids problematic practices like p-hacking

—where researchers continue collecting data until significance is achieved without controlling Type I error rates—thereby promoting transparency in research processes and results and fostering a better academic environment (Hu et al., 2016; Nosek et al., 2022). However, only a minority (30.6%) of existing intensive longitudinal intervention studies have determined sample sizes based on power analysis (see Table 1), with half (50.0%) not even mentioning the basis for sample size determination. This has resulted in substantial variation in participant numbers and intervention occasions (or corresponding intensive measurement occasions) across studies (see Table 2).

Table 2 Sample Sizes in Psychology Intensive Longitudinal Intervention Studies Using SAT and RCT Designs (2015-2025, $N = 34$)

Design	Participants per Group	Pre-Phase Measurement Occasions (Days)	Intervention Phase Measurement Occasions (Days)	Intervention Phase Intervention Occasions (Days)
SAT	84 (21)	60 (14)	126 (28)	70 (20)
RCT	24 (7)	14 (7)	12 (7)	61 (20)

Note: The table summarizes sample sizes for the two most widely used designs (SAT and RCT) across 34 studies. SAT = single-arm trial; RCT = randomized controlled trial. Since RCT designs typically use two equal-sized groups, the table reports participants per group. A few studies did not conduct intensive tracking measurements during the intervention phase, hence the minimum value of 0 for intervention phase measurement occasions.

An increasing number of studies advocate for determining sample size based on both power analysis and effect size estimation accuracy (Liu et al., 2023, 2024; Arend & Schäfer, 2019; Maxwell et al., 2008). A scientifically sound sample size should simultaneously ensure: (1) correct rejection of the null hypothesis with accurate estimation of effect direction, and (2) high precision in effect size estimation. On one hand, power analysis based on null hypothesis significance testing (NHST) requires that sample size achieve a predetermined power standard. In power analysis for DSEM, deriving power analytically is extremely difficult due to the model' s multiple random parameters; consequently, existing studies predominantly use Monte Carlo simulation to conduct power analysis and determine sample size (Fang & Wang, 2024; Lafit et al., 2022; Schultzberg & Muthén, 2018). However, few studies have conducted power analysis for DSEM models applicable to intensive longitudinal intervention research. On the other hand, the core of effect size accuracy analysis is controlling the width of the confidence interval (CI) for the effect size—narrower intervals indicate more accurate estimation (Maxwell et al., 2008). Additionally, metrics such as relative bias of parameter estimates and CI coverage of true values can provide further information about effect size estimation accuracy. Therefore, Liu et

al. (2024) proposed confidence interval width contour plots to help determine sample sizes that simultaneously meet requirements for power and effect size accuracy. However, no study has yet applied this approach to DSEM models for intensive longitudinal intervention research, combining power analysis and effect size accuracy requirements to provide sample size planning recommendations.

1.4 Problem Statement

In summary, DSEM represents an effective method for analyzing intensive longitudinal intervention data, yet few methodological studies have focused on DSEM models applicable to such analyses. The DSEM model proposed by Hamaker et al. (2021) is not suitable for single-arm designs, while the model used by Yi (2020) only examined intervention effects reflected in the mean of outcome variables. This study extends these models to enable comprehensive evaluation of intervention effects across mean, inertia, and IIV for the two most common intensive longitudinal intervention designs—single-arm and randomized controlled designs—providing more detailed information feedback for personalized interventions.

Currently, few studies have investigated sample size planning for intensive longitudinal intervention experimental designs using DSEM. Although Yi (2020) conducted sample size planning based on DSEM under three intensive longitudinal intervention designs, that study' s model focused only on intervention effects reflected in the mean of outcome variables, neglecting that changes in autoregression and IIV are equally important for evaluating intervention effectiveness. Moreover, the basis for sample size planning referenced only power analysis without incorporating effect size estimation results (e.g., Maxwell et al., 2008). Therefore, it is necessary to explore sample size planning that integrates both power analysis and effect size accuracy analysis based on extended DSEM models that simultaneously reflect intervention effects on mean, autoregression, and IIV, providing a more scientific basis for sample size determination in actual research.

This study first extends and summarizes DSEM models applied to single-arm and randomized controlled designs based on the more scientifically rigorous design that includes intensive tracking measurements both before and after intervention. Second, using Monte Carlo simulation methods, we examine sample size planning for DSEM models under both designs and visually present results combining power analysis and effect size accuracy analysis through credible interval width contour plots. Third, through simulation studies, we compare differences between the two designs from the perspective of Type I error rates under varying sample size conditions. Finally, we illustrate how to implement sample size planning based on DSEM models using an empirical intensive longitudinal intervention study. All Mplus code used in this study is available in the appendix (https://doi.org/10.57760/sciencedb.psych.00506).

2. Dynamic Structural Equation Models for Intensive Longitudinal Intervention Research

Before analysis, data collected from intensive longitudinal intervention studies are organized into long-format data suitable for Mplus analysis (see Table 3). ID represents participant number (from 1 to N), represents participant i' s outcome variable values measured during the pre-intervention phase, represents participant i' s outcome variable values measured during the post-intervention phase, Time represents measurement time points (pre-intervention from 1 to , post-intervention from 1 to ; i.e., when measurement time points differ between pre- and post-phases, the maximum measurement time point T is used, with unmeasured time points recorded as missing), pre- and post-intervention phases are coded separately, and Group represents the grouping variable (0 = control group, 1 = intervention group). Data from single-arm designs include columns 1-4, while data from randomized controlled designs include columns 1-5.

Table 3 Example Data Structure for Intensive Longitudinal Intervention Studies

ID	Time	Group
1	1	0
1	2	0
...	...	...
1	1	1
1	2	1
...	...	...

2.1 Dynamic Structural Equation Modeling Under Single-Arm Design

Building upon Yi (2020)' s model for examining differences in outcome variable means between pre- and post-intervention phases, this study extends the DSEM model to examine changes in autoregression and intra-individual variability (Model 1, illustrated in Figure 1 [Figure 1: see original paper]).

Figure 1 Schematic Diagram of Dynamic Structural Equation Model for Single-Arm Design

First, the variance of outcome variable y in pre- and post-intervention phases is decomposed into between-person and within-person components.

Decomposition Model:

Pre-intervention phase:

Post-intervention phase:

where represents participant i' s outcome variable value at pre-intervention time point t, represents participant i' s outcome variable value at post-intervention time point t (time points in pre- and post-phases are coded separately starting from 1), represents the between-person component, and represents the within-person component.

Second, within-person models are defined separately for pre- and post-intervention phases.

Within-Person Model:

where and represent random autoregressive coefficients for pre- and post-intervention phases, respectively; within-person residuals follow a multivariate normal distribution. In practice, residuals are typically assumed to be independent across phases, i.e., . Since variance is always positive, following previous research, the logarithm of within-person residual variance (i.e.,) is used to facilitate further modeling of intra-individual variability (Hamaker et al., 2018; Schultzberg & Muthén, 2018).

Between-Person Model:

The between-person model constructs both a pre-intervention phase model and a model reflecting differences between post- and pre-intervention phases.

Pre-intervention phase model:

Post-pre difference model:

where are fixed parameters for the pre-intervention phase model, representing between-person means of the outcome variable, autoregressive coefficient, and residual variance, respectively; are fixed parameters representing changes in mean, autoregressive coefficient, and within-person residual variance from pre- to post-intervention—i.e., intervention effects across three dimensions. These are the key parameters targeted in power analysis and effect size accuracy analysis.

Notably, although we can theoretically define the between-person models in equations (5)-(7) simultaneously, during parameter estimation we can only estimate the sum of fixed and random effects for intercepts in the pre-intervention model and the difference model (e.g., equation (8) can only estimate). Therefore, a two-step approach can be used for parameter estimation in practice. Step 1: Use Bayesian methods to estimate parameters in the pre-intervention model and save their posterior distributions. Step 2: Draw several parameter values (e.g., 100) from the posterior distribution obtained in Step 1, treat them as fixed values, and estimate the difference parameters in the pre-post difference model to obtain their posterior distributions.

In Monte Carlo simulation-based power analysis, fixed effects in the pre-intervention portion ( ) can be set to 0 to test fixed effects reflecting intervention effects ( ). Random parameters in the between-person model follow a multivariate normal distribution: .

It is worth noting that the single-arm design assumes no natural change in outcome variables across phases, attributing all changes to intervention effects—a potentially overly stringent assumption in practice. Additionally, the between-person model has an equivalent alternative formulation where the pre-post difference equations (8)-(10) can be restructured to model the post-intervention phase directly (the form used in Mplus estimation). For example, equation (8) becomes . When all random effects are independent, the random effects for post-intervention results include both pre-intervention random effects ( ) and difference random effects ( ).

2.2 Dynamic Structural Equation Modeling Under Randomized Controlled Design

The dynamic structural equation model under randomized controlled design (Model 2) adopts the extension developed by Hamaker et al. (2021), illustrated in Figure 2 [Figure 2: see original paper].

Figure 2 Schematic Diagram of Dynamic Structural Equation Model for Randomized Controlled Design (Without Covariates)

In the between-person model for post-pre differences (equations (8)-(10)), a grouping variable *Group* is added (e.g., *Group* = 1 for intervention group, *Group* = 0 for control group). Intervention effects are reflected through regression coefficients predicting pre-post differences in mean, autoregression, and intra-individual variability (i.e., Δlog) from the grouping variable.

Model 2' s decomposition model and within-person model are identical to those under single-arm design (equations (1)-(4)). The pre-intervention phase model for the between-person portion is also identical to equations (5)-(7). The difference lies in the addition of the grouping variable to the pre-post difference model under randomized controlled design:

where is a dummy variable indicating whether participant i belongs to the intervention or control group, and represent the effects of the grouping variable on pre-post differences in mean, autoregression, and within-person residual variance—i.e., intervention effects. These are the key parameters targeted in power analysis and effect size accuracy analysis. As in the single-arm design, random parameters in the between-person model follow a multivariate normal distribution.

Notably, this model assumes no pre-post differences for the control group. Similar to Model 1, when all random effect components are independent, post-intervention random effects include both pre-intervention random effects and difference random effects.

In practice, outcome variables may undergo natural changes across pre- and post-intervention phases over time (Hamaker et al., 2021). If a single-arm design is used, such natural changes become confounded with experimental treatment effects and cannot be separated. In contrast, randomized controlled de-

signs can separate natural changes through control group pre-post differences and obtain purer intervention effects based on regression coefficients for the grouping variable. This study further extends the DSEM model under randomized controlled design to accommodate natural changes between pre- and post-intervention phases (Model 3). The difference from the previous model lies in the pre-post difference model specification:

where terms with subscript *change.i* represent natural changes between phases for control group participant i—i.e., differences occurring between the two phases without any intervention. The meanings of remaining parameters are identical to Model 2.

---

3. Simulation Study 1: Sample Size Planning Under Single-Arm Design

Simulation Study 1 is based on SAT. Under the model framework constructed in Section 2.1, we examine statistical power and parameter estimation accuracy for parameters reflecting intervention effects ( ) across different sample size conditions, and provide sample size recommendations through credible interval width contour plots.

3.1 Simulation Study Design

This study considers two scenarios. Scenario 1: Equal numbers of measurement occasions in pre- and post-intervention phases. Scenario 2: Unequal numbers of measurement occasions in pre- and post-intervention phases. In practice, the number of measurement occasions in the pre-intervention phase is generally smaller than in the post-intervention phase (e.g., Mair et al., 2022). However, considering that the pre-intervention phase also requires a sufficient number of measurement occasions to obtain accurate and stable parameter estimates for autoregression and intra-individual variability, this study uses a 1:3 ratio of pre- to post-intervention measurement occasions as an example to examine how unbalanced designs affect sample size planning.

3.1.1 Sample Size Levels Referencing empirical intensive longitudinal intervention study designs (Table 1) and DSEM-related simulation study designs, we set sample size levels (Fang & Wang, 2024; Schultzberg & Muthén, 2018). Total number of measurement time points (T) includes 6 levels: 20, 40, 80, 120, 160, and 200. Number of participants (N) includes 7 levels: 30, 60, 100, 150, 200, 300, and 400. These two factors are fully crossed, forming 42 sample size combinations.

3.1.2 Data Generation Referencing parameter settings from similar previous studies (Fang & Wang, 2024; Schultzberg & Muthén, 2018; Yi, 2020), data were generated based on Model 1. Fixed effects were set as . For simplicity, all

random components were constrained to be independent, assuming random effects . Random effects for the pre-intervention phase were set based on previous research (Fang & Wang, 2024; Yi, 2020). Referencing comparisons of random effect estimates from two components in empirical intensive longitudinal intervention studies (see Section 6) and combining them with pilot studies using different parameter settings, random effects for the pre-post difference model were set as . Following similar sample size planning studies (Arend & Schäfer, 2019), intervention effect values representing pre-post differences were set at a medium level according to Cohen' s (1988) standards, with standardized parameter values (referencing Cohen' s $d = 0.2, 0.5, 0.8$). Since this study focuses only on average intervention effects across all individuals, standardization is based on between-person level random effect variance (Arend & Schäfer, 2019): , where *var* is the total random effect variance at the between-person level. Based on our assumption of independent random effects, , where *var* is the sum of variances of all random effects at the between-person level. Therefore, . Data were simulated 500 times for each sample size combination.

3.1.3 Model Fitting The same model used for data generation (Model 1) was fitted using Bayesian estimation, the commonly employed method for DSEM parameter estimation. Note that simulation studies assume stationarity within each intensive tracking measurement phase, thus detrending is unnecessary (Zhou et al., 2021). However, actual research still requires detrending within each phase (see Section 6). All data simulation and analysis were completed using Mplus 8.10 (Muthén & Muthén, 2023). Specifically, the MONTECARLO command was used to define data generation conditions (e.g., sample size, number of replications), the MODEL POPULATION command defined the data-generating model and true parameter values, and the ANALYSIS command defined the fitted model. In Bayesian estimation, Mplus default non-informative priors were used (i.e., regression coefficient fixed parts , random effect variances). Minimum iterations were set to 10,000, with the first 5,000 as burn-in and 2 chains as default. Convergence criteria adopted Mplus default settings, where a PSR (potential scale reduction) value less than 1.1 for each parameter indicates convergence.

3.1.4 Evaluation Criteria Evaluation metrics include four aspects: (1) **Convergence rate**: the proportion of converged parameter estimation runs out of total replications. All subsequent evaluation metrics are calculated based on converged runs. (2) **Statistical power**: the proportion of replications where the 95% credible interval (CI) for parameters reflecting intervention effects ( ) excludes 0. The preset power standard is ≥ 0.8. (3) **Effect size estimation accuracy**[1]: includes relative parameter estimation bias (rbias), root mean squared error (RMSE), credible interval width, and coverage probability (CP) for parameters of interest. (4) **Standard error estimation accuracy**: bias of estimated standard errors relative to the standard deviation of parameter estimates (SE-SD bias). If effect size estimation is accurate, rbias

should be within [-0.1, 0.1], RMSE should be small, width should be narrow, and CP should be between 0.925 and 0.975 (Bradley, 1978). If standard error estimation is accurate, SE-SD bias should be close to 0 (Schultzberg & Muthén, 2018).

[1]Generally, accuracy refers to the closeness of estimates to true values (including rbias, RMSE, etc.), while precision refers to the stability of estimates (primarily standard error, reflected in CI width). However, when the model is correctly specified (as in this study where the fitted model matches the data-generating model) and parameter estimation methods are unbiased, the two are approximately equivalent. Therefore, following Maxwell et al. (2008), we use the umbrella term "accuracy in parameter estimation (AIPE)."

3.2 Results

Under balanced design scenarios, model convergence rates reached over 95% across all conditions, with 81% of conditions achieving 100% convergence. Under unbalanced design scenarios, convergence rate was only 49% under the $N = 30$, $T = 20$ condition, but exceeded 90% in all other conditions, with 48% achieving 100% convergence (results in Appendix Table 2). Balanced designs showed superior convergence rates compared to unbalanced designs.

3.2.1 Statistical Power Power results for each parameter under different sample size conditions in balanced designs are presented in Table 4 . As shown, power increases with sample size. Once participant number reaches a certain level (e.g., $N \geq 200$), the influence of measurement time points on power diminishes. Power levels are generally consistent across the three parameters. Under equivalent sample size conditions, power for the autoregression intervention effect parameter is higher than the other two parameters, with the largest differences observed when N is small. Power for the IIV intervention effect parameter is the lowest. Power results for unbalanced designs (Appendix Table 3) are slightly lower than for balanced designs.

Table 4 Power, Relative Bias, and Credible Interval Width Results for Balanced Single-Arm Design

N	T	Power	rbias	Width
...	...	...	...	...

Note: N = number of participants; T = total number of measurement time points. Bolded power values indicate results < 0.8. Bolded rbias values indicate results outside [-0.1, 0.1].

3.2.2 Effect Size and Standard Error Estimation Accuracy Table 4 presents relative bias and credible interval width results for the three parameters under different sample size conditions in balanced designs. Additional

results (RMSE, coverage, and SE bias) are in Appendix Table 4. All relative bias values fall within acceptable ranges. However, when the number of measurement time points is small (e.g., $T = 20$ or 40), relative bias for the other two parameters falls below -0.1, underestimating intervention effect sizes. Credible interval width decreases as sample size increases. RMSE results are consistent with relative bias. Coverage mostly falls between 0.925 and 0.975, indicating good coverage and moderate standard errors. SE bias fluctuates around 0, suggesting accurate effect size standard error estimation.

Effect size and standard error estimation results for unbalanced designs are in Appendix Tables 3 and 5 . Relative bias is slightly smaller in unbalanced designs compared to balanced designs, with showing small overall bias and showing bias outside acceptable ranges when $T = 20$. Credible interval widths are slightly larger in unbalanced designs when participant numbers are small ($N = 30, 60$). RMSE, coverage, and SE bias are comparable between unbalanced and balanced designs.

3.2.3 Sample Size Recommendations Following Liu et al. (2024), we provide sample size recommendations using credible interval width contour plots. Since this study uses Bayesian estimation to obtain posterior distribution credible intervals, we refer to them as credible interval width contour plots. Figure 3 [Figure 3: see original paper] and Appendix Figures 1-2 show credible interval width contour plots for the three parameters under balanced design. Appendix Figures 3-5 show corresponding plots for unbalanced design. Shaded areas represent conditions meeting the power ≥ 0.8 criterion, while different colored contour lines correspond to different credible interval widths. Following Liu et al. (2024), we define acceptable maximum credible interval width based on Cohen' s (1988) small and large effect size standards of 0.2 and 0.8. Under small effect size conditions, ; under large effect size conditions, . Therefore, the maximum acceptable credible interval width is 0.179 - 0.045 = 0.134.

Figure 3 Credible Interval Width Contour Plot for Under Balanced Single-Arm Design

Note: Shaded area represents conditions with power $\geq$ 0.8. Contour lines from 0.08 to 0.50 represent 95% CI widths. For example, the contour line at 0.134 indicates that the area above this line has 95% CI width $\leq$ 0.134.

In balanced designs, participant number and measurement time points show compensatory effects: increasing either enhances power and reduces credible interval width. However, participant number must meet certain requirements (e.g., $N \geq 60$ for) to achieve power > 0.8. The edge of the shaded region nearly coincides with the 0.134 contour line, indicating high consistency between sample sizes determined by power and effect size accuracy criteria. Comparing contour plots across the three parameters reveals that has the largest shaded area, with the other two being similar, suggesting that requires relatively smaller sample sizes. Based on the contour plots, appropriate sample sizes can be

selected at turning points along the edge of shaded regions, considering specific research costs and credible interval width contours.

Specifically, under medium effect size conditions, if simultaneously adopting power ≥ 0.8 and credible interval width < 0.134 criteria: for , recommended $N = 100$ with $T = 80$ (8,000 data points) or $N = 60$ with $T = 160$ (9,600 data points); for , recommended $N = 60$ with $T = 80$ (4,800 data points); for , recommended $N = 60$ with $T = 160$ (9,600 data points). Overall, to correctly identify intervention effects across all three parameters and accurately estimate effect sizes, the maximum required sample size across parameters should be taken. Under balanced design, at least $N = 60$ participants and $T = 160$ measurement time points are needed; under unbalanced design, at least $N = 150$ participants and $T = 80$ measurement time points are needed, requiring larger N and smaller T compared to balanced design.

---

4. Simulation Study 2: Sample Size Planning for Randomized Controlled Design

Simulation Study 2 is based on RCT. Under the model framework constructed in Section 2.2, we examine statistical power and parameter estimation accuracy for three parameters reflecting intervention effects ( ) across different sample size conditions, and provide sample size recommendations through credible interval width contour plots. As in Study 1, we consider both equal (Scenario 1, 1:1) and unequal (Scenario 2, 1:3) numbers of measurement occasions between pre- and post-intervention phases.

4.1 Simulation Study Design

Simulation Study 2 uses the same sample size levels as Simulation Study 1, comprising 42 sample size combinations. A balanced design with equal group sizes was adopted (participants per group $= N/2$), so N in the sample size levels represents total participants across both groups. Data were generated based on Model 2, where the grouping variable is a binary categorical variable. Since this study uses Model 2 with the grouping variable as a Level-2 predictor, we follow Rights and Sterba' s (2018) approach for linear mixed-effects models (multilevel models) by calculating the proportion of variance explained by the grouping variable' s fixed slope at each level . This was set at a medium level of 0.09 (Cohen, 1988). For the mean intervention effect (equation (11)), the model is transformed to use post-intervention mean as the dependent variable: . Assuming independent random effects, the proportion of variance explained by this regression equation is . Since this study uses a design with equal group sizes, the variance of the grouping variable . Similarly, . All random components were constrained to be independent as in Simulation Study 1, with . Remaining true parameter values for data generation were identical to Study 1. The fitted

model matched the data-generating model. Bayesian estimation settings and evaluation criteria in Mplus were identical to Simulation Study 1.

4.2 Results

Model convergence rates exceeded 95% across all conditions, with 88% achieving 100% convergence.

4.2.1 Statistical Power Power results for each parameter under different sample size conditions in balanced designs are shown in Table 5. Similar to Simulation Study 1, power increases with sample size. Once participant number reaches a certain level (e.g., $N \geq 200$), the influence of measurement time points diminishes. Under equivalent sample size conditions, power for the IIV intervention effect parameter is lower than the other two parameters. Overall, power obtained in Simulation Study 2 is lower than in Simulation Study 1, likely because RCT designs split participants into two groups, reducing per-group sample size. When considering conditions with equal single-group sample sizes (e.g., $N = 60$ in Study 2 vs. $N = 30$ in Study 1), power levels are roughly comparable. Under small T conditions ($T = 20$), unbalanced designs show greater power than balanced designs, particularly when N is large ($N \geq 150$) and $T = 20$ (Appendix Table 6).

Table 5 Power, Relative Bias, and Credible Interval Width Results for Balanced Randomized Controlled Design

N	T	Power	rbias	Width
...	...	...	...	...

Note: N = total number of participants (N/2 per group); T = total measurement time points. Bolded power values indicate results < 0.8. Bolded rbias values indicate results outside [-0.1, 0.1].

4.2.2 Effect Size and Standard Error Estimation Accuracy Table 5 presents relative bias and credible interval width results for the three parameters under different sample size conditions in balanced designs. Additional results (RMSE, coverage, SE bias) are in Appendix Table 7 . Relative bias falls within acceptable ranges across all conditions except when $N = 30$ or 60 with $T = 20$, where the mean intervention effect parameter is underestimated. Credible interval width decreases as sample size increases. RMSE results are consistent with relative bias. Coverage mostly falls between 0.925 and 0.975, though slightly above 0.975 occurs more frequently when measurement time points are limited (e.g., $T = 20$), suggesting potentially larger standard errors in these conditions. SE bias fluctuates around 0, indicating accurate effect size standard error estimation.

Effect size and standard error estimation results for unbalanced designs are in Appendix Tables 6 and 8 . Credible interval widths are slightly smaller in unbalanced designs compared to balanced designs, particularly when both N and T are small. Relative bias, RMSE, coverage, and SE bias are comparable between unbalanced and balanced designs.

4.2.3 Sample Size Recommendations Following the method in Simulation Study 1, we calculated maximum acceptable credible interval widths for the three parameters. Based on small and large effect size standards of 0.01 and 0.25 (Liu et al., 2024), the maximum acceptable credible interval width is 0.258 - 0.045 = 0.213. Figure 4 [Figure 4: see original paper] and Appendix Figures 6 [Figure 6: see original paper]-7 show credible interval contour plots for the three parameters under balanced design. Appendix Figures 8 [Figure 8: see original paper]-10 show corresponding plots for unbalanced design.

Figure 4 Credible Interval Width Contour Plot for Under Balanced Randomized Controlled Design

Note: Shaded area represents conditions with power $\geq$ 0.8.

In balanced designs, similar to Simulation Study 1, compensatory effects exist between participant number and measurement time points once participant number reaches a certain threshold. For example, to achieve adequate power for , at least 100 participants are required; for , at least 80 participants are needed –higher than the minimum requirements in Simulation Study 1. Additionally, the autoregression intervention effect parameter shows the largest shaded area, requiring relatively smaller sample sizes. Unlike Simulation Study 1, the shaded area in Simulation Study 2' s credible interval contour plots is reduced, indicating that larger participant numbers are needed to meet power requirements. The maximum acceptable credible interval width contour line falls below the lower edge of the shaded area, suggesting this criterion requires smaller sample sizes than power criteria.

Specifically, under medium effect size conditions, if simultaneously adopting power $\geq$ 0.8 and credible interval width $<$ 0.213 criteria, we recommend using balanced designs with equal group sizes: for , total $N = 100$ with $T = 160$ (16,000 data points) or total $N = 150$ with $T = 120$ (18,000 data points); for , total $N = 60$ with $T = 160$ (9,600 data points) or total $N = 100$ with $T = 80$ (8,000 data points); for , total $N = 120$ with $T = 160$ (19,200 data points). Overall, to correctly identify intervention effects across all three parameters and accurately estimate effect sizes, balanced designs require at least $N = 100$ participants and $T = 160$ measurement time points; unbalanced designs require at least $N = 100$ participants and $T = 120$ measurement time points, with unbalanced designs requiring fewer measurement time points.

5. Simulation Study 3: Type I Error Rate Comparison Between Two Designs

Combining results from both simulation studies reveals that SAT requires smaller sample sizes than RCT to achieve adequate power and effect size accuracy. However, this does not necessarily mean SAT is recommended in practice. Since SAT cannot separate natural changes from intervention effects, it may lead to erroneous estimation of intervention effects in actual intervention research. Simulation Study 3 considers scenarios with natural changes between pre- and post-intervention phases. Under equal measurement occasion conditions between phases, we analyze data using DSEM models for both designs and compare Type I error rates for intervention effect parameters across different sample size conditions to further illustrate the applicability conditions for each intensive longitudinal intervention design.

5.1.1 Data Generation

Data were generated based on Model 3. Intervention effects of the grouping variable on post-intervention mean, autoregression, and intra-individual variability were set to 0, representing no intervention effect. Random effect components used the same settings as Model 2. Parameters representing natural changes ( ) were set at two levels: (1) a small effect size level according to Cohen (1988) ( ), i.e., ; and (2) an even smaller level ( ), i.e., .

5.1.2 Simulation Factors

Considering the study purpose, three simulation factors were examined. Natural change effect size (ES) includes 2 levels: 0.1 and 0.2. Since this study does not focus on sample size planning, it does not need to cover as many sample size levels as Simulation Studies 1 and 2. Therefore, Study 3 uses fewer sample size levels: total pre-post measurement time points (T) includes 5 levels (40, 80, 120, 160, 200) and participant number (N) includes 3 levels (60, 100, 200). These three factors are fully crossed, forming 30 sample size combinations. Data were simulated 500 times for each combination.

5.1.3 Model Fitting

For RCT, Model 3 was used to fit generated data. For SAT, intervention group participants were selected from simulated data and Model 1 was applied. Consequently, participant numbers under SAT were half those under RCT. Bayesian estimation settings in Mplus were identical to Simulation Studies 1 and 2.

5.1.4 Evaluation Metrics

Type I error rate was the primary evaluation metric for comparing results between the two designs—the proportion of converged replications where credible

intervals for intervention effect parameters (SAT: ; RCT:) exclude 0. Ideal Type I error rates should fall within [0.025, 0.075].

5.2 Results

Type I error rate results for both designs are shown in Table 6. When natural changes exist between pre- and post-intervention phases, RCT yields Type I error rates mostly within acceptable ranges, whereas SAT produces inflated Type I error rates when both participant number and measurement time points are large. This inflation increases with larger natural change effect sizes. Among the three intervention effects, Type I error rate inflation is most severe for the autoregression intervention effect, consistent with its relatively highest power results (Simulation Studies 1 and 2). Specifically, when natural change effect size = 0.2, SAT shows inflated Type I error rates even with $N = 30$. When $N = 100$ and $T = 200$, all three parameters show Type I error rates above 0.5. When natural change effect size = 0.1, SAT shows inflated Type I error rates when $N \geq 50$. When $N = 100$ and $T = 200$, all three parameters show Type I error rates above 0.14.

Table 6 Type I Error Rate Results for Both Designs

Natural ES	Design	N (SAT)	T	Mean	Autoregression	IIV
0.1	RCT	60 (30)	60	···	···	···
···	···	···	···	···	···	···

Note: 0.1 and 0.2 represent natural change effect sizes between pre- and post-intervention phases. Mean = mean intervention effect; Autoregression = autoregressive intervention effect; IIV = intra-individual variability; SAT = single-arm trial; RCT = randomized controlled trial. N = number of participants (values in parentheses indicate SAT sample size, which is half of RCT). T = total measurement time points. Bolded values indicate Type I error rates outside [0.025, 0.075].

---

6. Empirical Study: Sample Size Planning for an Ecological Momentary Intervention for Social Media Appearance Anxiety

This empirical study demonstrates how to apply our methods to guide sample size planning based on a pilot ecological momentary intervention for social media appearance anxiety. The process includes: (1) conducting a pilot study and analyzing results to determine true values for the simulation; (2) applying Monte Carlo simulation based on DSEM and set true values to repeatedly generate and analyze data; (3) calculating evaluation metrics including power and

effect size accuracy; (4) synthesizing these metrics to determine final sample size recommendations.

6.1 Pilot Study and Results Analysis

Research shows that using highly visual social media platforms based on visual content can trigger transient appearance-related anxiety, termed social media appearance anxiety (Hawes et al., 2020). We conducted an ecological momentary intervention for female college students' social media appearance anxiety using self-compassion statement reading, obtaining preliminary results. A subsequent study plans to validate these findings with a broader sample including different genders, necessitating scientific sample size planning for the expanded sample. The pilot study process and results are detailed in the appendix.

6.2 Sample Size Planning Based on Pilot Study

Using pilot study results (Appendix Table 11) as true parameter values, we generated data, fitted models, and conducted analyses following Simulation Study 2. Note that sample size planning based on pilot studies can proceed with trial-and-error, setting appropriate sample size levels in simulations. Since simulations found that even with $N = 100$ and $T = 200$, power for all intervention effect parameters remained below 0.8, we set N at 4 levels (200, 250, 300, 400) and T at 5 levels (40, 80, 120, 160, 200). Power and effect size accuracy results for the three intervention effect parameters are in Appendix Tables 9 and 10.

The results show that when measurement time points are limited, relative bias for falls below -0.1, tending to underestimate this parameter. Relative bias for other parameters falls within [-0.1, 0.1]. RMSE results are consistent with relative bias. Coverage values fall within [0.925, 0.975], indicating good coverage and moderate standard errors. SE bias fluctuates around 0, suggesting accurate effect size standard error estimation. Since the true effect size for is very small, even increasing sample size to $N = 400$ and $T = 200$ cannot achieve power > 0.8, preventing generation of a credible interval width contour plot for this parameter.

Credible interval width contour plots are shown in Figures 5 [Figure 5: see original paper] and 6. For the mean intervention effect, power > 0.8 is achieved when $N = 250$ with $T = 80$ (20,000 data points) or $N = 300$ with $T = 40$ (12,000 data points). For the IIV intervention effect, power > 0.8 is achieved when $N = 200$ with $T = 120$ (24,000 data points) or $N = 250$ with $T = 40$ (10,000 data points). Finally, researchers can select appropriate sample sizes within shaded regions based on desired credible interval widths.

Figure 5 Credible Interval Width Contour Plot for Based on Pilot Study Results

Note: Shaded area represents conditions with power $\geq$ 0.8.

Figure 6 Credible Interval Width Contour Plot for Based on Pilot Study Results

Note: Shaded area represents conditions with power ≥ 0.8.

---

7. Discussion

7.1 Main Evaluation Metrics for Sample Size Planning

The study found that DSEM models generally converge successfully. Under SAT, sample sizes meeting credible interval width requirements are slightly larger than those meeting power requirements; under RCT, sample sizes meeting credible interval width requirements are smaller than those meeting power requirements. This indicates that sample sizes determined by power analysis and effect size accuracy analysis may differ across designs. Researchers can choose one or both criteria based on specific needs. Additionally, the recommended sample sizes from this study suggest that some empirical studies have generally small sample sizes (see Table 2), indicating that many empirical studies lack scientifically justified sample size planning, possibly because their data analysis models are simpler and require smaller samples.

From a sample size planning perspective, both designs require minimum participant numbers to achieve adequate power. Parameter estimation accuracy is generally good. For SAT, intervention effects on autoregression and IIV are underestimated when measurement time points are limited (e.g., $T = 20$ or 40). For RCT, the mean intervention effect is underestimated when both participant number and measurement time points are small (e.g., $N = 30$ or 60, $T = 20$). Coverage of credible intervals is good, and standard errors are moderate. Standard error estimation is accurate. Note that the scales differ for mean, autoregression, and residual variance, making power and accuracy results not directly comparable across intervention effects. For example, achieving medium or large intervention effects for autoregressive coefficients may be difficult under DSEM' s stationarity assumptions in practice. Therefore, comparisons of sample size planning results across different intervention effects should be made cautiously.

Regarding balanced vs. unbalanced designs: For SAT, unbalanced designs show slightly lower power and slightly larger credible interval widths when participant numbers are small. For RCT, unbalanced designs show greater power than balanced designs when measurement time points are limited ($T = 20$), with slightly smaller credible interval widths. Overall, unbalanced designs are advantageous when measurement time points are scarce, allowing more measurements to be allocated to the more important intervention phase.

7.2 Comparison of Sample Size Planning Between Two Designs

This study explored sample size planning under both designs and compared their Type I error rates. Results show that under equal intervention effect sizes, SAT requires smaller sample sizes than RCT. However, if natural changes in outcome variables occur across pre- and post-intervention phases, SAT cannot control for these differences and will overestimate Type I error rates for intervention effects, with more severe inflation at larger sample sizes.

Each design has distinct advantages, disadvantages, and applicable contexts. Although SAT requires smaller sample sizes, improvements in psychological symptoms during post-intervention intensive tracking may result from natural developmental changes rather than pure intervention effects, carrying higher risk of Type I error inflation. RCT is more scientifically rigorous, using a control group to separate effects of other factors over time and obtain purer intervention effects—particularly important for longer-duration intensive longitudinal interventions (Hamaker et al., 2021). Consequently, this design is most widely used in intensive longitudinal intervention research (74.3%, see Table 1). However, RCT requires larger sample sizes, especially participant numbers, substantially increasing research costs and reducing feasibility.

7.3 Practical Recommendations

This study examined sample size planning under two typical intensive longitudinal intervention designs and compared Type I error rates across sample size conditions. Applied researchers can refer to the Mplus code provided in our appendix to generate simulated data under different sample size combinations, analyze data, calculate relevant evaluation metrics, and conveniently plot credible interval width contours using code from Liu et al. (2024) (see appendix) to obtain recommended sample sizes. Based on our findings, we offer the following recommendations:

First, select appropriate intervention experimental designs based on specific contexts. Comparisons of sample size planning across designs reveal that SAT and RCT have distinct advantages, disadvantages, and applicable contexts. Researchers should choose designs based on characteristics of outcome variables, participant populations, measurement frequency, etc. When participant populations are readily available or when multiple (two or more) control groups are needed, the more scientifically rigorous RCT is recommended (Wright et al., 2018). When participant populations or symptoms are rare, or when natural changes in outcome variables across phases are known to be absent, SAT with lower sample size requirements may be used. Regardless of design choice, sample size planning should be scientifically justified.

Second, select scientifically appropriate data analysis methods based on intensive longitudinal data characteristics. Although DSEM has become increasingly mainstream for intensive longitudinal research, no empirical studies have applied this method to intensive longitudinal intervention data,

with only a few methodological studies validating DSEM for examining intervention effects (Hamaker et al., 2021; Yi, 2020). To fully exploit intensive longitudinal data and accurately capture time series and nested structure characteristics while evaluating intervention effects multidimensionally across mean, autoregression, and IIV, we recommend using the DSEM models employed in this study. Sample size planning should also be based on these models to facilitate adoption of advanced analytical methods. Note that our sample size planning method focuses only on parameters related to intervention effects; if other parameters or concerns (e.g., model fit, predictive accuracy) are of interest, corresponding evaluation metrics should be integrated into sample size planning.

Third, carefully set data-generating model parameters when using simulation-based sample size planning. Simulation-based power and accuracy analysis requires specifying all parameter values for the data-generating model. Since intervention effects are individual-level variables, only fixed effects can be set at medium levels following sample size planning conventions (Arend & Schäfer, 2019). However, the three outcome variables (mean, autoregression, residual variance) have different possible value ranges, and small/medium/large effect size standards vary across effect size indicators, complicating fixed effect parameter settings. Moreover, DSEM models contain multiple random effects, with few studies available for reference on their values. Pilot studies found that random effect variance settings directly affect accuracy of intervention effect parameter estimation, making appropriate random effect variance specification a key concern in most DSEM simulation studies (e.g., Fang & Wang, 2024). We recommend referencing fixed and random effects from previous intensive longitudinal studies involving outcome variables when setting data-generating model parameters. When no reference studies exist, pilot experiments can be conducted to set values based on estimated fixed and random effects. However, some research notes that biased point estimates from pilot studies may not yield reasonable sample sizes (e.g., Albers & Lakens, 2018). Therefore, consider drawing true values from pilot study effect size distributions, incorporating minimum effect sizes of interest (SESOI), referencing meta-analyses, or directly applying parameter settings from this study. Regardless of approach, the basis for model parameter settings should be clearly stated in formal studies to standardize sample size planning procedures.

Fourth, comprehensively consider time trend issues in sample size planning. Intensive longitudinal intervention studies may involve two types of time trend effects: (1) Within each pre- and post-intervention phase, outcome variables may change over time. This trend violates stationarity assumptions (Zhou et al., 2019) and causes parameter estimation bias. Intensive longitudinal simulation studies generally assume stationarity and omit detrending (e.g., Asparouhov et al., 2018; Fang & Wang, 2024). However, empirical studies typically require detrending (e.g., Wang & Maxwell, 2015). For example, the pilot study showed significant negative time effects in both phases (pre-intervention time effect CI = [-0.016, -0.008]; post-intervention time effect CI = [-0.011, -0.003]),

indicating decreasing social media appearance anxiety over time. (2) Natural changes between pre- and post-intervention phases—i.e., differences in mean, autoregression, and IIV that would occur without any intervention (control group). We recommend separating this confounding effect through experimental design. In summary, we recommend using RCT designs when possible and omitting detrending when applying simulation-based sample size planning, but detrending within each phase when analyzing actual data.

Finally, we recommend the following flowchart sequence for determining appropriate sample sizes in actual intensive longitudinal intervention studies. If considering medium effect sizes for all intervention effects and referencing other parameter settings from this study, to simultaneously meet power and effect size accuracy requirements, we recommend: SAT—minimum $N = 60$, $T = 160$ (balanced) or $N = 150$, $T = 80$ (1:3 unbalanced); RCT—minimum $N = 100$, $T = 160$ (balanced) or $N = 100$, $T = 120$ (1:3 unbalanced).

Figure 7 [**Figure 7: see original paper**] Flowchart for Sample Size Planning in Intensive Longitudinal Intervention Studies

7.4 Future Directions

This study has limitations suggesting three directions for future research. **First**, following previous work (Liu et al., 2024; Usami, 2020), we derived maximum acceptable credible interval width from desired CI limits. However, when effect sizes are small or large (e.g., Cohen' s $d = 0.2$ or 0.8) or at non-critical values (e.g., $d = 0.4$ or 0.6), this method cannot determine maximum width. Recent work by Kowialiewski (2025) proposed determining sample size based on effect size stabilization criteria—incrementally increasing sample size via simulation until parameter estimate changes fall below a preset threshold. Future research could reference this approach to determine maximum acceptable credible interval width for stable effect size estimates.

Second, this study adopted frequentist sample size planning principles, conducting Monte Carlo simulations based on fixed effect size values while ignoring effect size uncertainty (e.g., Pek & Park, 2019). Future research could employ Bayesian frameworks, generating effect sizes from specific distributions to obtain more scientific sample size recommendations (Kruschke & Liddell, 2018).

Third, as model complexity increases, computational demands of existing simulation-based sample size planning methods grow exponentially, reducing efficiency. Moreover, setting discrete sample size levels yields only limited condition combinations that may not identify globally optimal solutions. To address this, researchers have introduced surrogate model frameworks based on machine learning predictions for power analysis, using search algorithms to identify optimal sample size designs (Zimmer & Debelak, 2025). Future research could apply this approach to DSEM-based sample size planning for intensive longitudinal intervention studies and develop corresponding software to promote efficient, widespread research.

References

Albers, C., & Lakens, D. (2018). When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. *Journal of Experimental Social Psychology*, 74, 187–195.

Arend, M. G., & Schäfer, T. (2019). Statistical power in two-level models: A tutorial based on Monte Carlo simulation. *Psychological Methods*, 24(1), 1–19.

Aschenbrenner, A. J., & Jackson, J. J. (2024). High-frequency assessment of mood, personality, and cognition in healthy younger, healthy older and adults with cognitive impairment. *Aging, Neuropsychology, and Cognition*, 31(5), 914–931.

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(3), 359–388.

Baey, C., & Le Deley, M. C. (2011). Effect of a misspecification of response rates on type I and type II errors, in a phase II Simon design. *European Journal of Cancer*, 47(11), 1647–1652.

Balaskas, A., Schueller, S. M., Cox, A. L., & Doherty, G. (2021). Ecological momentary interventions for mental health: A scoping review. *PLOS ONE*, 16(3), e0248152.

Bell, I. H., Fielding-Smith, S. F., Hayward, M., Rossell, S. L., Lim, M. H., Farhall, J., & Thomas, N. (2018). Smartphone-based ecological momentary assessment and intervention in a coping-focused intervention for hearing voices (SAVVy): Study protocol for a pilot randomized controlled trial. *Trials*, 19, 1–13.

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, 31(2), 144–152.

Chen, M., & Zhou, P. (2017). Ecological momentary assessment and intervention of substance use. *Advances in Psychological Science*, 25(2), 247–252.

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

Cuijpers, P., Pineda, B. S., Quero, S., Karyotaki, E., Struijs, S. Y., Figueroa, C. A., ···& Muñoz, R. F. (2021). Psychological interventions to prevent the onset of depressive disorders: A meta-analysis of randomized controlled trials. *Clinical Psychology Review*, 83, 101955.

Fang, Y., & Wang, L. (2024). Dynamic structural equation models with missing data: Data requirements on N and T. *Structural Equation Modeling: A Multidisciplinary Journal*, 31(5), 891–908.

Hamaker, E. L., Asparouhov, T., & Muthén, B. (2021). Dynamic structural equation modeling as a combination of time series modeling, multilevel modeling, and structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (2nd ed., pp. 359–388). Guilford Press.

Hawes, T., Zimmer-Gembeck, M. J., & Campbell, S. M. (2020). Unique associations of social media use and online appearance preoccupation with depression, anxiety, and appearance rejection sensitivity. *Body Image*, 33, 66–76.

Hawker, C. O., Merkouris, S. S., Youssef, G. J., & Dowling, N. A. (2021). A smartphone-delivered ecological momentary intervention for problem gambling (GamblingLess: Curb Your Urge): Single-arm acceptability and feasibility trial. *JMIR Mental Health*, 8(3), e25786.

Hedeker, D., Mermelstein, R. J., & Demirtas, H. (2008). An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. *Biometrics*, 64(2), 627–634.

Heron, K. E., & Smyth, J. M. (2010). Ecological momentary interventions: Incorporating mobile technology into psychosocial and health behaviour treatments. *British Journal of Health Psychology*, 15(1), 1–39.

Hoffman, L., & Walters, R. W. (2022). Catching up on multilevel modeling. *Annual Review of Psychology*, 73, 659–689.

Hu, C., Wang, F., Guo, J., Song, M., Sui, J., & Peng, K. (2016). The reproducibility issue in psychological research: From crisis to opportunity. *Advances in Psychological Science*, 24(9), 1504–1518.

Kenny, D. A., & Judd, C. M. (1986). Consequences of violating the independence assumption in analysis of variance. *Psychological Bulletin*, 99(3), 422–431.

Kowialiewski, B. (2025). The power of effect size stabilization. *Behavior Research Methods*, 57(1), 1–8.

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, 25, 178–206.

Lafit, G., Sels, L., Adolf, J. K., Loeys, T., & Ceulemans, E. (2022). Power-LAPIM: An application to conduct power analysis for linear and quadratic longitudinal actor–partner interdependence models in intensive longitudinal dyadic designs. *Journal of Social and Personal Relationships*, 39(10), 3085–3115.

Li, Y., Williams, L., Muth, C., Heshmati, S., Chow, S. M., & Oravecz, Z. (2024). A growth of hierarchical autoregression model for capturing individual differences in changes of dynamic characteristics of psychological processes. *Structural Equation Modeling: A Multidisciplinary Journal*, 32(2), 237–250.

Liu, Y., Xu, L., Liu, H., Han, Y., You, X., & Wan, Z. (2024). Confidence interval width contours: Sample size planning for linear mixed-effects models.

Acta Psychologica Sinica, 56(1), 124–138.

Mair, J. L., Hayes, L. D., Campbell, A. K., Buchan, D. S., Easton, C., & Sculthorpe, N. (2022). A personalized smartphone-delivered just-in-time adaptive intervention (JitaBug) to increase physical activity in older adults: Mixed methods feasibility study. *JMIR Formative Research*, 6(4), e34662.

Maxwell, S. E., Kelley, K., & Rausch, J. R. (2008). Sample size planning for statistical power and accuracy in parameter estimation. *Annual Review of Psychology*, 59, 537–563.

Muthén, L. K., & Muthén, B. O. (1998–2024). *Mplus user' s guide*. Muthén & Muthén.

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Almenberg, A. D., ···& Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, 73, 719–748.

Pek, J., & Park, J. (2019). Complexities in power analysis: Quantifying uncertainties with a Bayesian-classical hybrid approach. *Psychological Methods*, 24(5), 590–605.

Rauschenberg, C., Boecking, B., Paetzold, I., Schruers, K., Schick, A., van Amelsvoort, T., & Reininghaus, U. (2021). A compassion-focused ecological momentary intervention for enhancing resilience in help-seeking youth: Uncontrolled pilot study. *JMIR Mental Health*, 8(8), e25650.

Reininghaus, U., Daemen, M., Postma, M. R., Schick, A., Hoes-van der Meulen, I., Volbragt, N., ···& van Amelsvoort, T. (2024). Transdiagnostic ecological momentary intervention for improving self-esteem in youth exposed to childhood adversity: The SELFIE randomized clinical trial. *JAMA Psychiatry*, 81(3), 227–239.

Rights, J. D., & Sterba, S. K. (2019). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological Methods*, 24(3), 309–338.

Schueller, S. M., Aguilera, A., & Mohr, D. C. (2017). Ecological momentary interventions for depression and anxiety. *Depression and Anxiety*, 34(6), 540–545.

Schultzberg, M., & Muthén, B. (2018). Number of subjects and time points needed for multilevel time-series analysis: A simulation study of dynamic structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(4), 495–515.

Sherwood, S. N. (2022). *Feasibility and efficacy of virtual darkness in reducing intra-individual sleep variability among young adults with insomnia* (Doctoral dissertation, University of Nevada, Las Vegas).

Shrier, L. A., & Spalding, A. (2017). "Just take a moment and breathe and think" : Young women with depression talk about the development of an ecolog-

ical momentary intervention to reduce their sexual risk. *Journal of Pediatric and Adolescent Gynecology*, 30(1), 116–122.

Smith, K. E., & Juarascio, A. (2019). From ecological momentary assessment (EMA) to ecological momentary intervention (EMI): Past and future directions for ambulatory assessment and interventions in eating disorders. *Current Psychiatry Reports*, 21, 1–8.

Usami, S. (2020). Confidence interval-based sample size determination formulas and some mathematical properties for hierarchical data. *British Journal of Mathematical and Statistical Psychology*, 73(S1), 1–31.

Wang, L. P., & Maxwell, S. E. (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Psychological Methods*, 20(1), 63–83.

Wilhelm, P., Perrez, M., & Pawlik, K. (2012). Conducting research in daily life: A historical review. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 3–14). Guilford Press.

Wright, C., Dietze, P. M., Agius, P. A., Kuntsche, E., Livingston, M., Black, O. C., Room, R., Hellard, M., & Lim, M. S. (2018). Mobile phone-based ecological momentary intervention to reduce young adults' alcohol use in the event: A three-armed randomized controlled trial. *JMIR mHealth and uHealth*, 6(7), e149.

Yi, Z. (2020). *Intensive longitudinal data analyses and sample size considerations in intervention studies with dynamic structural equation modeling* (Doctoral dissertation, University of South Florida).

Zhang, W., Xu, L., Yao, L., Zhong, W., & Li, J. (2024). Application progress of ecological momentary intervention in health behavior promotion. *Journal of Nursing Science*, 39(2), 116–121.

Zhou, L., Wang, M., & Zhang, Z. (2021). Intensive longitudinal data analyses with dynamic structural equation modeling. *Organizational Research Methods*, 24(2), 219–250.

Zimmer, F., & Debelak, R. (2025). Simulation-based design optimization for statistical power: Utilizing machine learning. *Psychological Methods*, 30(3), 513–536.

Appendices

Literature Search Procedures for 2015-2025 Intensive Longitudinal Intervention Studies

Literature searches were conducted using the X-Mol academic platform (https://www.x-mol.com/), a comprehensive academic search engine integrat-

ing and updating global scholarly resources in real-time from sources including PubMed, Web of Science Core Collection, Scopus, Crossref, arXiv, and major publishers (Elsevier, Springer, Nature, Wiley, Taylor & Francis). Since intensive longitudinal interventions are often defined as "ecological momentary interventions" in the literature, searches focused on this core concept using keywords "ecological momentary intervention" and its standard abbreviation "EMI." The time range was 2015–2025. Journal sources are listed in Appendix Table 1.

Appendix Table 1 Journal Sources for 2015–2025 Intensive Longitudinal Intervention (EMI) Studies

Journal	Impact Factor
Internet Interventions	3.6
Psychiatry Research	3.9
JAMA Psychiatry	22.5
...	...

Note: Complete table available in online appendix.

Appendix Table 2 Convergence Rates for Unbalanced Single-Arm Design

N	T	Convergence Rate
30	20	0.49
...	...	...

Appendix Table 3 Power, Relative Bias, and Credible Interval Width for Unbalanced Single-Arm Design

N	T	Power	rbias	Width
...	...	...	...	...

Appendix Table 4 RMSE, Coverage, and SE Bias for Balanced Single-Arm Design

N	T	RMSE	Coverage	SE-SD Bias
...	...	...	...	...

Appendix Table 5 RMSE, Coverage, and SE Bias for Unbalanced Single-Arm Design

N	T	RMSE	Coverage	SE-SD Bias
...	...	...	...	...

Appendix Table 6 Power, Relative Bias, and Credible Interval Width for Unbalanced Randomized Controlled Design

N	T	Power	rbias	Width
...	...	...	...	...

Appendix Table 7 RMSE, Coverage, and SE Bias for Balanced Randomized Controlled Design

N	T	RMSE	Coverage	SE-SD Bias
...	...	...	...	...

Appendix Table 8 RMSE, Coverage, and SE Bias for Unbalanced Randomized Controlled Design

N	T	RMSE	Coverage	SE-SD Bias
...	...	...	...	...

Appendix Table 9 Power, Relative Bias, and Credible Interval Width Based on Pilot Study Results

N	T	Power	rbias	Width
...	...	...	...	...

Appendix Table 10 RMSE, Coverage, and SE Bias Based on Pilot Study Results

N	T	RMSE	Coverage	SE-SD Bias
...	...	...	...	...

Appendix Table 11 Parameter Estimates from Pilot Study of Social Media Appearance Anxiety EMI

Parameter	Posterior Mean	95% CI	Width
...	...	...	...

Note: Bolded parameter estimates have 95% credible intervals excluding 0 (significant results).

Appendix Figure 1 Credible Interval Width Contour Plot for Under Balanced Single-Arm Design

Appendix Figure 2 Credible Interval Width Contour Plot for Under Balanced Single-Arm Design

Appendix Figure 3 Credible Interval Width Contour Plot for Under Unbalanced Single-Arm Design

Appendix Figure 4 Credible Interval Width Contour Plot for Under Unbalanced Single-Arm Design

Appendix Figure 5 Credible Interval Width Contour Plot for Under Unbalanced Single-Arm Design

Appendix Figure 6 Credible Interval Width Contour Plot for Under Balanced Randomized Controlled Design

Appendix Figure 7 Credible Interval Width Contour Plot for Under Balanced Randomized Controlled Design

Appendix Figure 8 Credible Interval Width Contour Plot for Under Unbalanced Randomized Controlled Design

Appendix Figure 9 [**Figure 9: see original paper**] Credible Interval Width Contour Plot for Under Unbalanced Randomized Controlled Design

Appendix Figure 10 [**Figure 10: see original paper**] Credible Interval Width Contour Plot for Under Unbalanced Randomized Controlled Design

Mplus Code for Single-Arm Design Data Generation and Analysis (Model 1, Balanced, $N = 100$, $T = 200$)

```
MONTECARLO:
  NAMES = y1 y2;
  NOBS = 10000;
  NREP = 500;
  CSIZES = 100(100);
  NCSIZES = 1;
  LAGGED = y1(1) y2(1);
  REPSAVE = ALL;
  SAVE = D:/mplus/samplesize/s2/100-100-0.5/model-100-100.rep*.dat;
  RESULTS = D:/mplus/samplesize/s2/100-100-0.5/output.sav;
```

```
  BPARAMETERS = D:/mplus/samplesize/s2/100-100-0.5/bayes.dat;

ANALYSIS:
  TYPE = twolevel random;
  ESTIMATOR = BAYES;
  PROCESSORS = 2;
  BITER = (10000);
  BSEED = 5240;

MODEL POPULATION:
  %within%
    phi1|y1 on y1&1;
    logv1|y1;
    phi2|y2 on y2&1;
    logv2|y2;
  %between%
    y1*0.04; [y1*0];
    phi1*0.04; [phi1*0.2];
    y2*0.01; [y2*0.112];
    y2 on y1@1;
    phi2 on phi1@1;
    phi2*0.01; [phi2*0.112];
    [logv1*0]; [logv2*0.112];
    logv1*0.04; logv2*0.01;
    logv2 on logv1@1;

MODEL:
  %within%
    phi1|y1 on y1&1;
    logv1|y1;
    phi2|y2 on y2&1;
    logv2|y2;
  %between%
    y1*0.04; [y1*0];
    phi1*0.04; [phi1*0.2];
    y2*0.01; [y2*0.112];
    y2 on y1@1;
    phi2 on phi1@1;
    phi2*0.01; [phi2*0.112];
    [logv1*0]; [logv2*0.112];
    logv1*0.04; logv2*0.01;
    logv2 on logv1@1;

OUTPUT:
  TECH1;
  TECH8;
```

Mplus Code for Randomized Controlled Design Data Generation and Analysis (Model 2, Balanced, $N = 100$, $T = 100$)

```
MONTECARLO:
  NAMES = y1 y2 group;
  NOBS = 10000;
  NREP = 500;
  CSIZES = 100(100);
  NCSIZES = 1;
  LAGGED = y1(1) y2(1);
  BETWEEN = group;
  CUTPOINTS = group(0);
  REPSAVE = ALL;
  SAVE = F:/LiuYue/EMI/RCT/100-100/model-100-100.rep*.dat;
  RESULTS = F:/LiuYue/EMI/RCT/100-100/output.sav;
  BPARAMETERS = F:/LiuYue/EMI/RCT/100-100/bayes.dat;

ANALYSIS:
  TYPE = twolevel random;
  ESTIMATOR = BAYES;
  PROCESSORS = 2;
  BITER = (10000);
  BSEED = 5240;

MODEL POPULATION:
  %within%
    phi1|y1 on y1&1;
    logv1|y1;
    phi2|y2 on y2&1;
    logv2|y2;
  %between%
    group*1; [group*0];
    y1*0.04; [y1*0];
    phi1*0.04; [phi1*0.2];
    y2*0.01; [y2*0];
    y2 on y1@1 group*0.14;
    phi2 on phi1@1 group*0.14;
    phi2*0.01; [phi2*0];
    [logv1*0]; [logv2*0];
    logv1*0.04; logv2*0.01;
    logv2 on logv1@1 group*0.14;

MODEL:
  %within%
    phi1|y1 on y1&1;
    logv1|y1;
```

```
    phi2|y2 on y2&1;
    logv2|y2;
  %between%
    y1*0.04; [y1*0];
    phi1*0.04; [phi1*0.2];
    y2*0.01; [y2*0];
    y2 on y1@1 group*0.14;
    phi2 on phi1@1 group*0.14;
    phi2*0.01; [phi2*0];
    [logv1*0]; [logv2*0];
    logv1*0.04; logv2*0.01;
    logv2 on logv1@1 group*0.14;

OUTPUT:
  TECH1;
  TECH8;
```

R Code for Generating Credible Interval Width Contour Plots (Randomized Controlled Design)

```r
# Data import
data <- read.csv("study2_{{{gamma41}}_{{ci}}}.csv")
data <- as.data.frame(data)
power <- read.csv("study2_{{{gamma41}}_{{power}}}.csv")
names(power)[1] <- "x"
pt <- read.csv("study2_{{{gamma41}}_{{pt}}}.csv")
names(pt)[1] <- "x"

# Set axis and contour scales
kd <- c(0.1, 0.15, 0.213, 0.5, 0.7, 0.9)
xbreak <- c(30, 60, 100, 150, 200, 300, 400)
ybreak <- c(20, 40, 80, 120, 160, 200)
label <- xbreak
xlimit <- c(0, 400)
ylimit <- c(0, 200)

# Plot contour
plot1 <- ggplot() +
  theme_{bw}() +
  xlab("Number of Participants") +
  ylab("Number of Time Points") +
  stat_{contour}(
    data = data,
    aes(x = level2, y = level1, z = ci, colour = ..level..),
    breaks = kd,
    linewidth = 1.08
```

```
) +
guides(color = guide_{colorbar}(
  title = "95% Credible Interval Width",
  title.theme = element_{text}(size = 12),
  draw.ulim = TRUE,
  draw.llim = TRUE,
  reverse = TRUE
)) +
scale_{{{color}}_{{gradientn}}}(
  colors = rev(c("#1822c7", "#c300a2", "#ff0073", "#ff714d", "#ffba43", "#f9f871")),
  breaks = kd
) +
theme(
  axis.text = element_{text}(size = 12),
  axis.title = element_{text}(size = 12),
  axis.line.x = element_{line}(size = 1),
  axis.line.y = element_{line}(size = 1),
  legend.text = element_{text}(size = 12),
  legend.key.height = unit(2, "cm")
) +
scale_x_{continuous}(
  limits = xlimit,
  breaks = xbreak,
  labels = label,
  expand = c(0, 0)
) +
scale_y_{continuous}(
  limits = ylimit,
  breaks = ybreak,
  expand = c(0, 0)
) +
theme(
  axis.ticks.length.y = unit(-0.1, 'cm'),
  axis.ticks.length.x = unit(-0.1, 'cm')
) +
geom_{polygon}(
  data = power,
  aes(x = x, y = y, group = 1),
  fill = "#edeed3",
  alpha = 0.3,
  color = "#edeed3",
  linewidth = 1.06
) +
geom_{point}(
  data = pt,
  aes(x = x, y = y),
```

```
    size = 1
  ) +
  geom_{line}(
    data = pt,
    aes(x = x, y = y),
    color = "grey",
    linewidth = 1
  )

plot1
```

Process and Results of Pilot Study on Social Media Appearance Anxiety EMI

The pilot study procedure was as follows. First, participants were recruited and completed eligibility screening and pre-test questionnaires. The screening questionnaire excluded participants with excessively high or low trait appearance anxiety, those who hadn' t used social media on their phones recently, etc. Pre-test questionnaires included individual-level variables such as self-compassion. A total of 237 participants met eligibility criteria and completed pre-test measures. They were evenly divided into intervention and control groups. All participants completed 20 days of intensive tracking measurements (once daily) before and after intervention, reporting daily social media appearance anxiety levels (1-100 scale). During the 20-day post-intervention period, the intervention group read self-compassion materials daily while the control group read neutral materials. After excluding dropouts, 200 participants remained (intervention group: $n = 105$; control group: $n = 95$).

Using social media appearance anxiety as the outcome variable, data were analyzed with Model 3. Note that simulation studies assume stationarity within each intensive tracking phase, thus detrending is unnecessary (Zhou et al., 2019). However, actual research requires detrending within each phase. Therefore, measurement time points were included as predictors for detrending in both pre- and post-intervention phases (Wang & Maxwell, 2015). Pilot study results are in Appendix Table 11. At the current sample size, the mean difference intervention effect parameter (CI = [-0.275, -0.031]), indicating the EMI significantly reduced social media appearance anxiety. The IIV intervention effect parameter (CI = [-0.586, -0.069]), showing reduced intra-individual variability. However, the autoregression intervention effect parameter (CI = [-0.090, 0.134]), indicating no change in autoregressive properties.

Appendix Table 11 Parameter Estimates from Social Media Appearance Anxiety EMI Pilot Study

Parameter	Posterior Mean	95% CI	Width
...	...	...	...

Note: Intervention materials were developed and evaluated based on relevant literature, demonstrating that self-compassion statements were significantly more related to self-compassion than neutral statements.

Note: Figure translations are in progress. See original paper for figures.