**Power computations in ESM designs**

In this paper we provide a tool for planning a sample an intensive longitudinal study for a given designs.

The reader who is planning his study and needs to know the optimal sample size for his study has two options. First, he may study the tables with recommended sample sizes given in this paper and select the table, which equals or is close to the design he actually intends to use. Second, he may use the R-function that we developed to run a simulation with the parameters of his choice. This function is freely available.

**Introduction about ESM/EMA research**

what is ESM, why important, missing values, attrition,

**Types of effects in ESM**

Synchronous versus longitudinal effects (lags)

direct versus indirect effects

conditional versus unconditional effects

with hand without auto-correlation

**Analyses tools**

Multilevel analysis

**Attrition and missing values**

After computing an estimate for the sample size, this estimate will almost invariable underestimate the sample size required for the study, unless you carefully address the following issues. First, participants often drop out of studies, a phenomenon called attrition in longitudinal studies. Longer and more intensive studies are likely to have higher attrition rates. Participants may also have more missing data than in cross-sectional studies, because of the intensive character of the study. Second, participants may exhibit more variation (i.e. be more different) than expected, which directly inflates the error variance and therefore the effective sample size. Third, participants sometimes provide data that is not useable (e.g. errors or irrealistic values), in which case they have to be excluded for some or all analyses. In fact, mistakes can be made at all levels during the data gathering process, which cause loss of data. Because this influences the actual required sample size, it is important to be aware of these issues. If no other guidelines are available, adding 20% to the raw estimate seems reasonable.

**The moderation model**

In a moderation model we assume that the effect of the predictor x on the dependent variable y is conditional on another variable z, which is called the moderator. In the statistical model (Hayes, 2013) this implies that there is an interaction term between x and z in the model.

*Effect sizes*

One of the crucial parameters in computing power analyses is the expected effect size. We assume that *x*, *z*, and ** are normally distributed and standardized with mean 0 and variance equal to 1. The effects sizes of the three effects are the standardized coefficients (beta), defined as:

.

If *x*, *z* and *xz* are uncorrelated, the variance of *y* is:

,

because the variance of *xz* is:

.

We choose b1, b2 and b3 such that their squares sum to 1. For example:

b1 = *sqrt*(.5), b2 = *sqrt*(.3) and b3 = *sqrt*(.2). If the variance of  is for example also 1, it follows that the = 1 + = 2. The expected *R* squared of this model is then:

.

The three effects sizes then become:

.

By changing the variance of the error the effect sizes can be manipulated.

*Correlated predictors*

When *x* and *z* are correlated the variance of the interaction term *xz* becomes:

.

Here *r* is the correlation between x and z. Assuming the interaction term is independent from x and z, the variance of y then becomes:

.

The effect size of x and z are computed with this variance term. For instance for *x*:

The effect size of the interaction term becomes:

The R2 of the model with correlated predictors can be obtained by:

Using these expressions we can construct expected effected sizes in a simulation study.

*Simulation study 1: moderation in the one-level model*

We simulate *x*, *z* and ** from a multivariate normal distribution using the R function *mvrnorm*(). The correlation between x and z has one of three values, representing no correlation (r=0), small correlation (r = .30), moderate correlation (r=.50), and high (r = .80). The *b* coefficients are chosen as the square roots from respectively, 0.5, 0.3, and 0.2.

The variance of  is one of three values, representing small (1), medium (3), and large levels (9) of random error. This corresponds with *R* squared values of respectively .50, .25, and .10.

With uncorrelated predictors the effect sizes (beta’s) of the three parameters are respectively, 0.50, 0.39, 0.32 in the small error condition, 0.35, 0.27, 0.22 in the medium error condition, and 0.22, 0.17, and 0,14 in the large error condition.

for samples of N = 1,000 the results are given in table 1. The power is always 100% for this table.

Table 1. Simulation results single level moderation model (N=1,000)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Var(e) | r(x,z) | Beta1 | Beta2 | Beta3 | R2 |
| 1 | 0.00 | 0.50 | 0.39 | 0.31 | 0.50 |
| 1 | 0.30 | 0.47 | 0.37 | 0.31 | 0.56 |
| 1 | 0.50 | 0.45 | 0.35 | 0.32 | 0.59 |
| 1 | 0.80 | 0.43 | 0.33 | 0.34 | 0.64 |
|  |  |  |  |  |  |
| 3 | 0.00 | 0.35 | 0.27 | 0.22 | 0.25 |
| 3 | 0.30 | 0.34 | 0.27 | 0.23 | 0.29 |
| 3 | 0.50 | 0.34 | 0.26 | 0.24 | 0.33 |
| 3 | 0.80 | 0.32 | 0.25 | 0.26 | 0.37 |
|  |  |  |  |  |  |
| 9 | 0.00 | 0.22 | 0.17 | 0.14 | 0.10 |
| 9 | 0.30 | 0.22 | 0.17 | 0.15 | 0.13 |
| 9 | 0.50 | 0.22 | 0.17 | 0.15 | 0.14 |
| 9 | 0.80 | 0.22 | 0.17 | 0.17 | 0.16 |

Because the relevant effect size for moderation (Beta 3) only slightly changes as the correlation changes, we report only one value of rho (rho = .30).

Table 2. Power results of single level moderation model. N=100

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Var(e) | N | Beta1 | Beta2 | Beta3 | R2 | Power |
| 1 | 100 | 0.47 | 0.37 | 0.31 | 0.56 | 0.988 |
| 3 |  | 0.34 | 0.26 | 0.23 | 0.31 | 0.705 |
| 9 |  | 0.22 | 0.17 | 0.15 | 0.15 | 0.293 |
|  |  |  |  |  |  |  |
| 1 | 150 | 0.47 | 0.37 | 0.31 | 0.56 | 0.999 |
| 3 |  | 0.34 | 0.26 | 0.23 | 0.31 | 0.876 |
| 9 |  | 0.22 | 0.17 | 0.15 | 0.15 | 0.466 |
|  |  |  |  |  |  |  |
| 1 | 200 | 0.47 | 0.37 | 0.31 | 0.56 | 1.000 |
| 3 |  | 0.34 | 0.26 | 0.23 | 0.31 | 0.950 |
| 9 |  | 0.22 | 0.17 | 0.15 | 0.15 | 0.577 |
|  |  |  |  |  |  |  |
| 1 | 250 | 0.47 | 0.37 | 0.31 | 0.56 | 1.000 |
| 3 |  | 0.34 | 0.26 | 0.23 | 0.31 | 0.983 |
| 9 |  | 0.22 | 0.17 | 0.15 | 0.15 | 0.657 |
|  |  |  |  |  |  |  |
| 1 | 300 | 0.47 | 0.37 | 0.31 | 0.56 | 1.000 |
| 3 |  | 0.34 | 0.26 | 0.23 | 0.31 | 0.992 |
| 9 |  | 0.22 | 0.17 | 0.15 | 0.15 | 0.744 |
|  |  |  |  |  |  |  |

*Multilevel moderation*

The single level moderation model can be extended to the multilevel situation.

Assume the following simple multilevel moderation model:

with as before *y* the dependent variable measured on subject *i*, *x* the predictor and *z* the moderator, *xz* the interaction term. The *b*’s are regression coefficients. Variables x and z are distributed as *N*(0,1) and the error term  as *N*(0,()). The correlation between x and z is r. There is no correlation between  with x and z.

In EMA designs there are (at least) two levels and we must specify at what level the moderation takes place. First the moderation can take place at the first level. At this level the synchronous model is written as:

with *t* the index that indicates the time or beep. Here y, x and z are measured at approximately the same moment. Taking the longitudinal nature of the data into account, it is natural to assume that y measured at *t* is predicted by x measured at *t*-1 and by the interaction (xz) measured at t-1. In addition, auto-regression (also with lag 1) can be incorporated in the model, that it is assumed that y at time *t* is predicted by y at time *t*-1:

This longitudinal model could be extended with predictors measured at earlier time points, in other words lagged 2 or lagged 3 variables could be added to the model.

For the multilevel analysis we assume that all slopes represent fixed effects. The intercept is assumed to be a random effect, which implies that the levels of the dependent variable may vary across subjects.

Results from ESM simulation:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| r(x,z) | Var(e) | Beta1 | Beta2 | Beta3 | R2 |
| 0.00 | 1 | 0.50 | 0.39 | 0.32 | 0.50 |
| 0.30 | 1 | 0.50 | 0.39 | 0.31 | 0.50 |
| 0.50 | 1 | 0.49 | 0.38 | 0.31 | 0.51 |
| 0.80 | 1 | 0.49 | 0.38 | 0.31 | 0.53 |
|  |  |  |  |  |  |
| 0.00 | 3 | 0.35 | 0.27 | 0.22 | 0.25 |
| 0.30 | 3 | 0.35 | 0.27 | 0.22 | 0.25 |
| 0.50 | 3 | 0.35 | 0.27 | 0.22 | 0.26 |
| 0.80 | 3 | 0.35 | 0.27 | 0.22 | 0.27 |
|  |  |  |  |  |  |
| 0.00 | 9 | 0.22 | 0.17 | 0.14 | 0.10 |
| 0.30 | 9 | 0.22 | 0.17 | 0.14 | 0.10 |
| 0.50 | 9 | 0.22 | 0.17 | 0.14 | 0.10 |
| 0.80 | 9 | 0.22 | 0.17 | 0.14 | 0.11 |

Changes in the correlation (r) between the predictors have a very small effect on the effect sizes. This is especially true for relatively small values of b3 (interaction term), which is usually the case in practical applications. For running a power simulation a single value of *r* (e.g. *r* = 0.3) is therefore sufficient.

**Mediation model**

We consider the following model. There is a pre-intervention measurement (PRE) and three post intervention and follow up measurements (POST, FU1, FU2, FU3). The effect van the intervention (condition) on the measurements is mediated by a mediator. Also a moderator moderates the effect of the intervention on the mediator. The parameter “a1” represents the main effect of the intervention on the mediator, “a3” represents the moderating (interaction) effect. effects b1, b2, b3 and b4 are the effects of the mediator on the measurements.

(no moderation)

(moderated mediation)

The effect sizes for the predictor, moderator and interaction on the mediator are represented by a1, a2 and a3, respectively. If auto-correlation is assumed this effect is represented by a4. The b’s represent the effects of respectively the mediator, predictor, and auto-correlation on the dependent variable.

The indirect effect of x on y through m is represented by “ind”.

In each condition a SEM analysis (using LAVAAN in R) was replicated 500 times. The results of the bias (true value minus the mean recovered parameters) and power are given in the following tables.

**BIAS (A1= .5, A3 = .5, B1 = .4, B2= .3, B3 = .2, B4 = .1)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| N | A1 (cond) | A3 (int) | B1 | B2 | B3 | B4 |
| 100 | .003 | -.011 | .044 | .033 | .022 | .011 |
| 150 | -.001 | -.003 | .042 | .031 | .021 | .010 |
| 200 | .002 | .009 | .043 | .034 | .022 | .011 |
| 250 | .008 | .003 | .043 | .033 | .021 | .011 |

**Power (A1=. 5, A3 = .5, B1 = .4, B2 = .3, B3 = .2, B4 = .1)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| N | A1 (cond) | A3 (int) | B1 | B2 | B3 | B4 |
| 100 | .73 | .71 | 1.00 | 1.00 | .46 | .00 |
| 150 | .88 | .89 | 1.00 | 1.00 | .86 | .02 |
| 200 | .97 | .94 | 1.00 | 1.00 | .99 | .05 |
| 250 | .98 | .98 | 1.00 | 1.00 | 1.00 | .11 |

**Attrition**

We assume 20%, 10% and 10% and 0% attrition at the consecutive measurements. Thus, at T1 we have left 80%, at T2 70%, at T3 60% and at T4 50% of the original sample. Under these assumptions the net sample size has to be 2\*N for T4, 1.67 \* N for T3, 1.43 \* N for T2 and 1.25 \* N for T1, where N is obtained from the tables presented above.

**Conclusions**

In all conditions the parameters are recovered quite well, the b’s are slightly underestimated. The power for b1 and b2 is always very good (100%) for b3 N=150 is the minimum sample size necessary to have power larger than 80%. For the weak effect of b4 the power is almost zero, this parameter is almost never significant.

The effect size of a1 and a3 need to be substantial (.5). In that case N = 150 seems sufficient. To obtain sufficient power for weaker effect sizes, samples sizes larger than 300 are presumably necessary.

If it is possible to assume that moderation and main effects of the intervention on the mediator are substantial, we start from N=150 and then considering the assumed attrition rates, we need a sample of N =250 (150\* 1.67) to have sufficient power to estimate the effect on T2.

Additional analyses showed that if we assume an ES of 0.2 for T4 and accounting for 50% attrition, then with N = 200 we obtain a power of 46%. With N= 300 (50% attrition accounted for) the

\*\*\*\*\*

Om de verschillende deelvragen te beantwoorden wordt multilevel analyse (Raudenbush & Bryk, 2002) gebruikt. Om een beeld te krijgen hoeveel kinderen nodig zijn om een voldoende hoge power te verkrijgen, is een simulatiestudie (Hedeker, Gibbons, & Waternaux, 1999) uitgevoerd in R (Team, 2013) met het pakket “lme4” (Bates, Maechler, Bolker, & Walker, 2013).

In de simulatie is uitgegaan van een middelgroot effect (.5) van training (in vergelijking met de controle groep) op cognities en vermijdingsgedrag. Voorts wordt er een zwak effect (.20) verwacht tussen de mediatoren enerzijds en angst anderzijds. Aangezien angststoornis de primaire uitkomstmaat is, is de primaire effectmaat het indirecte effect van de training op de angststoornis. Voor een power van 80% bij een type I fout van 5% gaf de simulatie (met 500 replicaties) aan dat bij 120 kinderen per conditie het indirecte pad een power van ca. 80% had.

In de simulatie is rekening gehouden met een verwachte uitval tussen de metingen. Er is aangenomen dat er tussen twee opeenvolgende metingen sprake is van 10% willekeurige uitval.

\*\*\*\*\*

To obtain an estimate of the power for a parameter estimate take the following steps:

Define which parameters are of primary interest (hypotheses).

Obtain realistic estimates for the parameters based on previous research or an educated guess.

Example

Treatment (X): Therapy (0=no, 1=yes)

5 repeated measures (M): mediators (Cognition)

5 repeated measures (W): mediators (Avoidance Behavior)

5 repeated measures (Y): outcomes (Anxiety)

Moderator (IQ) measured at T0

Design

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Treatment | T0 |  | T1 |  | T2 |  | T3 | T4 |
| therapy | O | X1 | O | X2 | O | X12 | O | O |
| control | O |  | O |  | O |  | O | O |

X1: therapy focus on cognition

X2: therapy focus on behaviour

X12: focus on cognition and behaviour

T4 = follow up

Observations include: mediators, outcome

Intelligence

d

Cognition

a1

b1

Anxiety

Therapy

c’

b2

Behaviour

a2

Figure 1. Research model: the effect of therapy on anxiety is mediated by Cognition and Behaviour. The effect of Therapy on cognition is moderated by intelligence.

Hypothesis 1. Cognition scores at T1 are more improved from T0 for the treatment group than for the control group (a1)

Hypothesis 2. Behaviour scores at T2 are more improved from T0 for the treatment group than for the control group (a2)

Hypothesis 3. Anxiety scores at T3 are more improved from T0 for the treatment group than for the control group (c)

Hypothesis 4. The effect of Therapy on Anxiety is mediated by Cognition (a1b1)

Hypothesis 5. The effect of Therapy on Anxiety is mediated by Behaviour (a2b2)

Hypothesis 6. The effect of Therapy on Cognition is moderated by Intelligence (interaction a1 x d)

Hypothesis 7. Anxiety scores at T4 (follow up) are the same as anxiety scores at T3

We assume the following attrition percentages across the time points: 0%, 10%, 20%, 30%, 40%.

**Multilevel Analysis**

The a1-step in de mediation model is defined as the difference between the treatment versus the control group at T1 for cognition (M). This is the coefficient of the interaction term Treatment x Time(1) with the mediator (M) as the dependent variable.

We expect a medium effect size (d = .5) for the first mediator (Hypothesis 1).

Hypo1 : ES(mMt1 - mMc1) = 0.5, where mMt1 is the mean of the Cognition mediator of the treatment group at t1 and mMc1 the mean of the control group at t1.

By definition: ES = (mMt1 - mMc1)) / SD(pooled)

The SD within all groups is by definition 1, because all variables are standardized, thus the difference in means is

mMt 1- mMc1 = 1 \* 0.5 = 0.50. We assume no effect on the mediators for the waiting list control group, thus (expected) mMc1 = 0, and mMt1 = .5.

The a2-step in de mediation model is defined as the difference between the treatment versus the control group at T2 for behaviour (W). This is the coefficient of the interaction term Treatment x Time(2) with W as the dependent variable.

We expect a medium effect size (d = .5) for the second mediator (Hypothesis 2).

Hypo2 : ES(mWt2 - mWct) = 0.5. Since the effect in the control group is expected to be 0, we get mWt2 = .5, where mWt2 is the mean of the Behaviour mediator of the treatment group at T2.

For the outcome (Y) we assume a small effect (0.2) for the waiting list control group, thus

ES(mYc1 - mYc0) = 0.2 , where mYc0 = 0 by definition.

Var(Yc1 - Yc0) = Var( Y1q) + Var(Yc0) - 2.cor(Yc1, Yc0) = 2 - 2.rho, where rho is the auto-correlation between the time points assumed to follow an AR(1) model.

With rho = .30, the variance of the difference becomes: 1.40 and the SD = 1.18, thus mY1c = 1.18 \* 0.2 + 0 (=mYc0) = 0.23. This effect remains during the rest of the measurements.

We expect a medium effect (0.5) for the anxiety variable at T1, it follows that mYt1 = 0.50 + 0.23 = .73. When the effect of W is present (T2), we expect a small additional effect (0.2), so mYt2 = 0.73 + 0.20 = .93, which remains at T3 and the follow-up T4.

Based on these computations we define the following patterns for the variables. The pattern in the treatment group for cognition is given by (0.00, 0.50, 0.50, 0.50, 0.50) and for behaviour by (0.00, 0.00, 0.50, 0.50, 0.50), whereas the control group for the mediators have a null-pattern (0, 0, 0, 0, 0). The dependent variable Y has pattern (0, .73, .93, .93, .93) versus the pattern in the control group of Y (0, 0.23, 0.23, 0.23, 0.23).

**Sample size computation**

Define N as the sample size per time point, with n participants in the control and n in the treatment group (thus N = 2n).

According to Gpower a single between groups effect (t-test) of ES= .5 with a power=.80 needs N=128 (thus N=142 with 10% missing, 160 with 20% missing, 183 with 30% missing).

Interaction between treatment and repeated (4) Y (or repeated M), requires a sample size according to Gpower for F-test (assuming no sphericity) with ES=.20, rho=.30 and Power=.80 of N=50 (thus N=71 with 30% missing).

However, comparing interaction with two repeated measurements N=72 (with missing N=103).

The effect m 🡪 y and w 🡪 y are assumed to be weak (.20), so standardized coefficients of b = .20 are expected. This implies that explained variance by m and w is at most 2 x 4%, leaving 92% remaining variance. The remaining variance consists of the therapy effect and error.

According to Gpower for the slope of .20 an N=191 (with 10%-40% missing pattern this implies n=24 per group per time point) is needed to obtain power=.80. Our effective sample size for the slope is 5N - missing = 4N.

With an attrition pattern: (10%, 20%, 30%, 40%) we have the following results from our Monte Carlo simulation:

Power of mediation design with 500 simulations (a1=.50, a2=.50, b1=.20, b2=.20, errors=.50, rho = .30)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| N | a1 | a2 | b1 | b2 | a1b1 | a2b2 | direct (cp) | total eff (c) |
| 100 | .45 | .41 | .98 | .98 | .38 | .33 | .29 | .66 |
| 150 | .66 | .58 | 1.0 | 1.0 | .46 | .46 | .46 | .83 |
| 200 | .77 | .71 | 1.0 | 1.0 | .65 | .62 | .54 | .90 |
| 250 | .87 | .80 | 1.0 | 1.0 | .78 | .72 | .62 | .95 |
|  |  |  |  |  |  |  |  |  |

Power of mediation design with 500 simulations (a1=.50, a2=.50, b1=.20, b2=.20, errors=.50, rho = .50)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| N | a1 | a2 | b1 | b2 | a1b1 | a2b2 | direct (cp) | total eff (c) |
| 100 | .55 | .49 | .97 | .95 | .42 | .41 | .33 | .75 |
| 150 | .75 | .65 | 1.0 | 1.0 | .64 | .57 | .43 | .87 |
| 200 | .86 | .79 | 1.0 | 1.0 | .77 | .71 | .56 | .92 |
| 250 | .94 | .86 | 1.0 | 1.0 | .85 | .79 | .87 | .98 |
|  |  |  |  |  |  |  |  |  |

**Literature**

Hayes, A. (2013). *Introduction to mediation, moderation, and conditional process analysis*. *New York, NY: Guilford*. http://doi.org/978-1-60918-230-4