

Paper 76-2010

Power to detect therapist effects in randomized controlled trials of interventions in clinical psychology

Doug Thompson, Assurant Health, Milwaukee, WI

Fary Cachelin, California State University, Los Angeles, CA

Ruth Striegel-Moore, Wesleyan University, Middletown, CT

Terry Wilson, Rutgers University, New Brunswick, NJ

ABSTRACT

In randomized controlled trials (RCTs) of interventions in clinical psychology, it is typical for multiple therapists to administer a given intervention. A therapist effect is present when treatment outcomes differ between therapists. Often, therapist effects have been viewed as a nuisance or an adjustment factor. However, therapist effects may be of practical importance; for example, they might indicate interventions that require extensive training to administer effectively. How many therapists are required to ensure adequate power to detect clinically meaningful therapist effects? Few studies have examined this issue. The present study examined how power is impacted by the number of therapists and the size of the therapist effect. The study utilized Monte Carlo simulations in SAS. This paper describes how to implement power analyses of therapist effects in SAS, as well as results from an initial set of analyses.

INTRODUCTION

In clinical psychology, when treatment outcomes differ between therapists, a therapist effect is in evidence. Past work has shown mixed results regarding therapist effects in clinical psychology interventions. The majority of studies that have examined the issue have found no therapist effects (e.g., Loeb, Wilson, Labouvie et al, 2005; Webb, DeRubeis & Barber, 2010) but some studies have reported a significant therapist effect (e.g., Jannoun, Munby, Catalan, & Gelder, 1980; Lewis et al, 2010; Mathews et al., 1976).

One issue is that most randomized controlled trials (RCTs) in clinical psychology are not specifically designed to have sufficient power to detect therapist effects. Therapist effects are often viewed as a nuisance or as an adjustment factor. However, therapist effects are important because they might indicate interventions that require extensive training, clinical aptitude or experience to administer effectively. Thus, it is worthwhile to consider what would be required to design an RCT with sufficient power to detect therapist effects.

The objective of the present study was to estimate the power to detect therapist effects, varying the number of therapists and the size of the therapist effect, while holding constant the study design and sample size. This paper also provides a detailed walk through an example of SAS code that could be used to design such studies.

METHODS

Study design

Sample size and power estimation are illustrated in the context of design of a randomized controlled trial (RCT) of therapy for eating disorders. Specifically, we designed a randomized controlled trial of guided self-help (GSH) compared with a no-treatment control, for treatment of binge eating disorder (BED) and bulimia nervosa (BN) in Mexican American women. The primary outcome was the count of episodes of binge eating during the past month. This outcome will be measured at baseline and post-treatment (i.e., after 12 weeks of treatment). The study also will include a follow-up visit at 6 months and additional outcome measurements, but these are disregarded for purposes of the present paper. One primary hypothesis is that treatment with GSH (compared with no GSH) will result in a greater reduction of binge eating at post-treatment (12 weeks after the beginning of treatment).

According to the study design, eligible participants will be randomized to either GSH (n=50) or no-treatment control (n=50). Prior analyses estimated that a sample size of 50 per group would provide sufficient power to test the primary hypotheses of the study. The assumptions used in sample size calculations were based on comparable past studies. The required sample size for the proposed study was estimated using simulations based on the following assumptions from the literature (this is the subset of assumptions relevant to the current paper):

- Generalized linear models will be the analytic technique;
- the significance of the treatment effect at post-treatment will be tested;
- equal numbers of participants will be randomly assigned to the GSH and control groups;
- in both groups, the raw mean (SD) of frequency of binge eating at baseline will be approximately 9.9 (14.4) per month;
- the outcome will be positively skewed and log-transformed prior to analysis;
- there will be a 10% reduction in binge eating in the control group (assuming some spontaneous recovery) and a 33% reduction in the GSH group (33% was the smallest reduction found in comparable past studies, and is used to be conservative); and
- there will be 30% attrition between baseline and post-treatment.

At the time that the study was originally designed, therapist effects were not considered. However, the issue of therapist effects came up during a kickoff meeting for the study. There was some leeway with regard to the number of therapists that would be used to conduct the study. The question arose, what number of therapists would be optimal? This question inspired the work described in the present paper. There are several different ways to answer the question.

This paper focuses exclusively on power to detect therapist effects. The therapist effect is defined as the estimated between-therapist difference in reduction of binge eating between baseline and post-treatment. Greater variation between therapists means that there is a greater estimated therapist effect. If some therapist achieves better outcomes than other therapists (in terms of reduction in binge

eating between baseline and post-treatment), a therapist effect will be in evidence. The null hypothesis is no difference among therapists.

Factors varied in the present study were 1) the number of therapists (between 2 and 5) and 2) the estimated therapist effect size. Three therapist effect sizes were considered, varying from larger to smaller. For purposes of this study, effect size was defined as (max Therapist mean) – (min Therapist mean) / common standard deviation (Grissom & Kim, 2005). The simulations examined 12 different scenarios, defined by crossing number of therapists with therapist effect size. The scenarios are described in Table 1.

Table 1. Average episodes of binge eating (BE) per month by visit (baseline vs. post-treatment), condition (control vs. treatment), therapist (ID #), number of therapists and therapist effect size.

Therapist effect size	Number of therapists	BE at baseline		BE at post-treatment					Max. difference among therapists
		(control and all therapists)		Treatment (therapist #)					
		Control (#0)		#1	#2	#3	#4	#5	
Largest	Five	9.9	8.9	4.7	5.6	6.6	7.6	8.6	4.0
	Four	9.9	8.9	4.7	5.6	7.6	8.6	n/a	4.0
	Three	9.9	8.9	4.7	6.6	8.6	n/a	n/a	4.0
	Two	9.9	8.9	4.7	8.6	n/a	n/a	n/a	4.0
Medium	Five	9.9	8.9	5.1	5.9	6.6	7.3	8.1	3.0
	Four	9.9	8.9	5.1	5.9	7.3	8.1	n/a	3.0
	Three	9.9	8.9	5.1	6.6	8.1	n/a	n/a	3.0
	Two	9.9	8.9	5.1	8.1	n/a	n/a	n/a	3.0
Smallest	Five	9.9	8.9	5.6	6.1	6.6	7.1	7.6	2.0
	Four	9.9	8.9	5.6	6.1	7.1	7.6	n/a	2.0
	Three	9.9	8.9	5.6	6.6	7.6	n/a	n/a	2.0
	Two	9.9	8.9	5.6	7.6	n/a	n/a	n/a	2.0

Table 1 describes the average tendencies in each scenario. Because the Monte Carlo simulations incorporate a realistic degree of sampling variability, means may vary widely among samples drawn from the population described in a given scenario.

A key consideration is whether the therapist effects will be modeled as random or fixed. A fixed effects model would be:

$$Y_{ij} = \alpha + \beta_0 * X_{0j} + \beta_1 * X_{1j} \dots \beta_n * X_{nj} + \epsilon_{ij}$$

$$\epsilon_{ij} \sim N(0, \sigma^2)$$

$$\text{Var}(Y_{ij}) = \sigma^2$$

Y_{ij} is the predicted log of binge eating at post-treatment for patient j within therapist i . α is an intercept term; as coded in this study, it is interpreted as the predicted log of binge eating at post-treatment for patients in the control group, assuming 0 episodes of binge eating at baseline (the intercept term would be more informative if baseline binge eating was mean- or median-centered prior to analysis). β_0 represents the effect of baseline binge eating. β_1 through β_n represent the effects of therapists i through n ($i=1$ to n therapists). $X_{1j} \dots X_{nj}$, are a set of n indicator variables for the therapists, coded 1 if patient j is

treated by therapist i and 0 otherwise. Control subjects are treated by none of the therapists, therefore they have a code of 0 for all of the therapist indicators. ϵ_{ij} is a residual term, assumed to be normally distributed with a mean of 0. A common residual variance is assumed for all patients, σ^2 .

Using a fixed effects approach, inferences are limited to the estimated therapist effects β_1 through β_n – the estimated therapist effects are not generalized to some larger population of therapists. In contrast, using a random effects approach, one assumes that the therapists in the study are a random sample from a larger population of therapists. A random effects approach enables the results to be generalized to the larger population of therapists. Using a fixed approach, the correlation among patients within therapists is typically not taken into account (although there are ways to do so using design effects adjustments or GEE, as implemented in PROCs SURVEYREG or GENMOD with a REPEATED statement). Homoscedasticity is assumed. In contrast, using a random effects approach, the correlation structure among patients within therapists (the “intra-therapist correlation”) can be modeled, appropriately accounting for the non-independence of patients within therapists. Random effects models also enable modeling of unequal variance for different patients (heteroscedasticity), in contrast to the constant variance assumed in the fixed effects approach. In other words, the assumptions of the fixed effects approach are more minimal, but potentially important aspects of the data (e.g., heteroscedasticity and the intra-therapist correlation) are not taken into account. The random effects approach has more assumptions but enables a more rich and potentially realistic representation of the data.

Random effects analyses of therapist effects are discussed by Walters (2010) and Walwyn and Roberts (2010). The design in the present study is referred to by Walwyn and Roberts as “partially nested,” meaning that only patients in the treatment arm have a therapist (control patients have no therapist). Walwyn and Roberts present several options for modeling such studies using a random effect approach.

For a definitive study of therapist effects, a random effects analysis is certainly ideal. However, a random effects analysis may not be plausible in small trials such as the one described in this paper. Brown and Prescott (2001) recommended that there be at least 5 clusters for a random effects analysis (in this paper, therapists are the clusters). Although this is only a rule of thumb, it has face validity – one would hesitate to generalize to a larger population of clusters (e.g., therapists) after having observed fewer than 5 or so clusters. Typically, one would want to observe many more than 5 clusters before feeling comfortable about generalizing to a larger population of clusters. Another problem is that the assumption of random sampling from a larger population of therapists was not clearly met in this study. The therapists were a carefully selected group of graduate students at a single university -- they were selected based on their interest in the study, their aptitude and their perceived ability to administer the treatment effectively. It is not clear what larger population would be appropriate for inference. Finally, based on a total sample of 100 patients (50 treated with GST, 50 control), it is not clear that the data are sufficient to support complex modeling of variance structures (e.g., heteroscedasticity, intra-therapist correlation). The variance estimates could easily be influenced by outliers/leverage points; thus complex modeling of variance structures could be misleading rather than informative. For these reasons, we

opted for a fixed effects approach as the primary method to analyze therapist effects, while recognizing the limitations of this approach.

Simulations

Monte Carlo simulations were used to estimate power to detect the therapist effect (Muthen & Muthen, 2002; Williams et al, 2007). Monte Carlo simulations enable power to be estimated with a great degree of flexibility and realism. Briefly, using this method, assumptions are made about the true characteristics of the population. In the present study, these assumptions were based on relevant past literature. Assumptions include the mean and standard deviation of the outcome, the size of the treatment effect, and the size of the therapist effect. Then repeated samples are drawn from the specified population. There is a realistic element of randomness that induces variation between the samples. This simulates what would actually be observed if one repeatedly kept doing studies of random samples of participants drawn from the population. In each sample, the statistical test of interest is performed (in this study, this is a test of the therapist effect). Power is the proportion of samples that have a significant test result.

In a Monte Carlo study, often the greatest effort is devoted to realistically specifying the characteristics of the population, including variability as well as average tendencies and intervention effect sizes. After this is done, it is relatively straightforward to draw repeated samples from the population, conduct the test in each sample and then summarize the test results across samples.

Here, we walk through the process of drawing a single random sample and performing the test on this sample. To estimate power (see Results section), we drew 500 random samples for each scenario (defined by a combination of number of therapists and therapist effect size) and summarized the results of the test of therapist effect across samples.

Illustrative simulation

The illustrative simulation uses the third scenario described in Table 1 (i.e., largest therapist effect, three therapists). First, we use a set of %let statements to specify key parameters that will be used in the simulation, namely the seed for random draws; the patient sample size per group (GSH and control); the average reduction in binge eating at post-treatment for patients in the treatment and control groups, relative to baseline; the number of therapists; and the effect of each therapist, relative to the average. The code below shows the %let statements for the example simulation.

```
* Specify seed for drawing random samples;  
%let seed = 8462;  
* Specify number of participants per group (GSH vs. control), ;  
* fixed in this study;  
%let n_per_group = 50;  
* Specify the average treatment effect (i.e., binge eating post-treatment;  
* = binge eating baseline * trtl_reduction).;
```

```

%let trt1_reduction = 0.67;
* Specify the average reduction in binge eating in the control group;
%let trt0_reduction = 0.90;
* Number of therapists;
%let n_therapists = 3;
* Size of the therapist effect;
* Binge eating for patients of therapist a at post treatment = ;
*   binge eating baseline * (trt1_reduction + therapist_effect a).;
%let therapist_effect = -0.20 0.00 0.20;

```

Next, the simulated complete dataset is drawn. It is assumed that each therapist will get an equal number of patients, if possible. If (n therapists / n patients) is not evenly divisible (i.e., there is a non-integer remainder), then the remaining patients will be randomly allocated to any of the therapists. In this situation, some therapists will end up getting more patients than others. Simulate baseline frequency of binge eating (called "ede0" in the SAS code) according to a probability distribution based on past literature. Ede0 has the same distribution for the GSH and control groups. For patients randomized to the control group, simulate post-treatment frequency of binge eating as the baseline frequency multiplied by 90%, plus a random deviation for each patient (this simulates a degree of spontaneous recovery for patients in the control group). For patients randomized to the GSH group, simulate post-treatment frequency of binge eating as the baseline frequency multiplied by 47% (therapist #1), 67% (therapist #2) or 87% (therapist #3), plus a random deviation for each patient. In other words, on average, patients in the GSH group will exhibit a post-treatment binge eating reduction of about 33%. However, therapist #1's patients have a tendency to improve by about 53%, therapist #2's patients have a tendency to improve by about 33%, and therapist #3's patients have a tendency to improve by 13% (which isn't much better than the control group's average improvement of 10%). Due to the random component, which simulates factors such as patient differences that are out of the control of the therapist, the outcomes for the 3 therapists will differ from simulation to simulation. The code below shows how the data are drawn in the example simulation.

```

data sim;
retain therapist patients_therapist;

effect_list = "&therapist_effect";

* Assume that the therapists will get (approximately) the same number of
patients.;
n_per_therapist=&n_per_group/&n_therapists;

* Initialize therapist as 1, at first therapist 1 starts with no patients
(until one is assigned).;
therapist=1;
patients_therapist=0;

do group=0 to 1;
  do _id=1 to &n_per_group;
    * This yields mean about 10, sd about 14, skewed.;
    * This baseline distribution is realistic based on past literature.;
    ede0= exp(1.81 + 1.03*rannor(&seed));
    if group=1 then do;

```

```

* Assign the next patient to the current therapist -- ;
* increment the therapists patient count by 1.;
patients_therapist+1;

* If the current therapist has too many patients, resolve;
if patients_therapist>n_per_therapist then do;
* Assign the patient to the next therapist, if there is;
* another therapist;
if therapist<&n_therapists then do;
therapist+1;
patients_therapist=1;
end;
* If all therapists have their allotment of patients,;
* randomly assign the patient to a therapist.;
else do;
therapist=int(1+(&n_therapists*ranuni(&seed)));
end;
end;

* Get the treatment effect for the current therapist;
treatment_effect_therapist =
(scan(effect_list,therapist,' '))*1;
* Percent change for the therapist is the average, plus the;
* therapist-specific effect.;
* Change for the patient post-treatment is a function of;
* the therapist effect, plus a;
* random component.;
change=(&trt1_reduction+treatment_effect_therapist) +
0.3*rannor(&seed);
if change<0 then change=0;
edel=ede0*change;
end;

* There are no therapists in control group;
if group=0 then do;
change=&trt0_reduction + 0.3*rannor(&seed);
if change<0 then change=0;
edel=ede0*change;
end;

* The outcome cannot be less than 0. Set to zero if less than 0.;
* (this condition very rarely applies).;
array edes{*} ede0-edel;
do i=1 to dim(edes);
if edes{i}^=. and edes{i}<0 then edes{i}=0;
end;

if group=1 then id=_id+&n_per_group;
else id=_id;
output;
end;
end;
label therapist = 'Therapist ID (0 = control)';
run;

```

For purposes of subsequent analyses, therapists are denoted using dummy-coding. For patients in the control group, the therapist ID value is assigned a code of 0 prior to the dummy-coding. This is accomplished in the macro `dummy_coding` shown below.

```
* Dummy-code therapists;
%macro dummy_coding;
data sim;
set sim;
if group=0 then therapist=0;
  %do trn=1 %to &n_therapists;
    if therapist=&trn then t&trn = 1;
    else t&trn= 0;
  %end;
run;
%mend dummy_coding;
%dummy_coding;
```

Next, we descriptively examine the distribution of the data resulting from the example simulation. Using PROC UNIVARIATE, histograms were created for frequency of binge eating by therapist (#1, #2 and #3; Therapist #0 is used to denote the control group) and visit (baseline vs. post-treatment). A table with mean, median and SD, by therapist and visit, was created as an additional way to summarize the data.

```
* Descriptive statistics for frequency of binge eating, by therapist;
* (as for control), at baseline and post treatment.;
ods html;
proc univariate data=sim noprint;
  class therapist;
  var ede0;
  histogram ede0 / nrows = 4
    midpoints=0 to 50 by 2;
  output out=baseline_stats mean=mean_baseline median=median_baseline
std=sd_baseline;
run;

ods html;
proc univariate data=sim noprint;
  class therapist;
  var ede1;
  histogram ede1 / nrows = 4
    midpoints=0 to 50 by 2;
  output out=post_stats mean=mean_post median=median_post std=sd_post;
run;

data baseline_and_post_stats;
merge baseline_stats post_stats;
by therapist;
run;
```

The resulting histograms (Figure 1 for baseline, Figure 2 for post-treatment) and summary table (Table 2) are shown below. At baseline, there is no marked difference in the distribution of frequency of binge eating (`ede0`) among the therapists (#1, #2, #3) and the control group (therapist #0). However, at post-treatment, the distribution for therapist #1 is markedly shifted toward 0 and there is also some

improvement for therapist #2, while therapist #3's patients show less improvement over the control group. This is apparent in the histograms as well as in the summary table.

Figure 1. Baseline distribution of binge eating (ede0) by therapist (#0 = control)

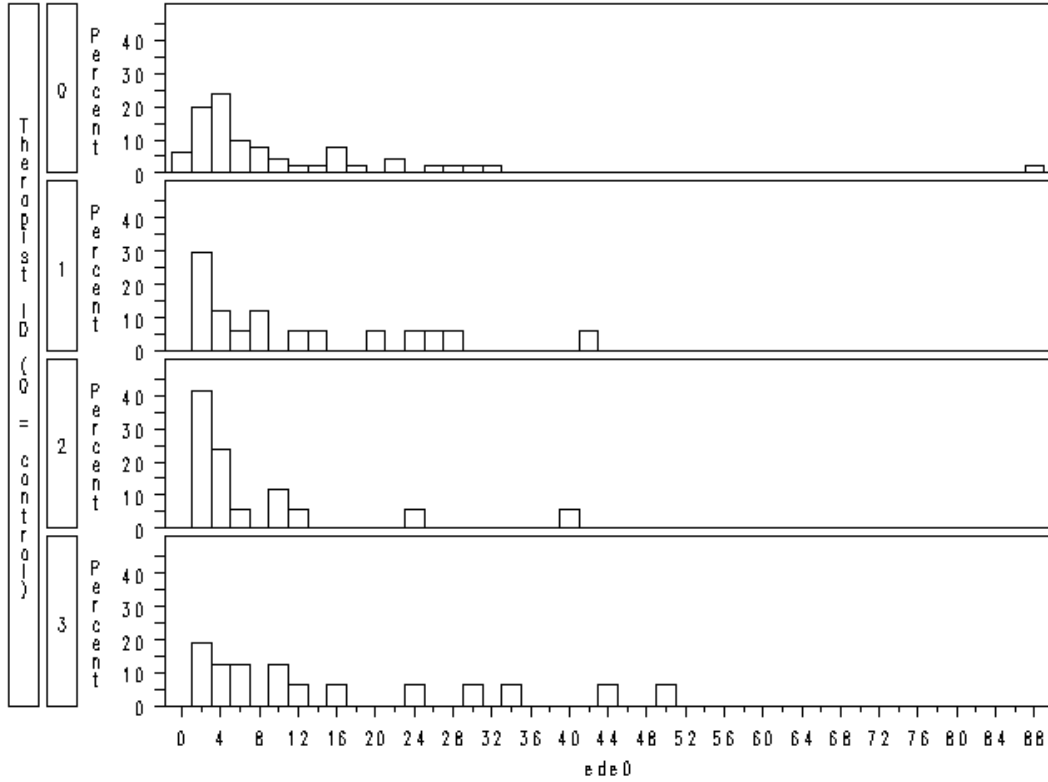


Figure 2. Post-treatment distribution of binge eating (ede1) by therapist

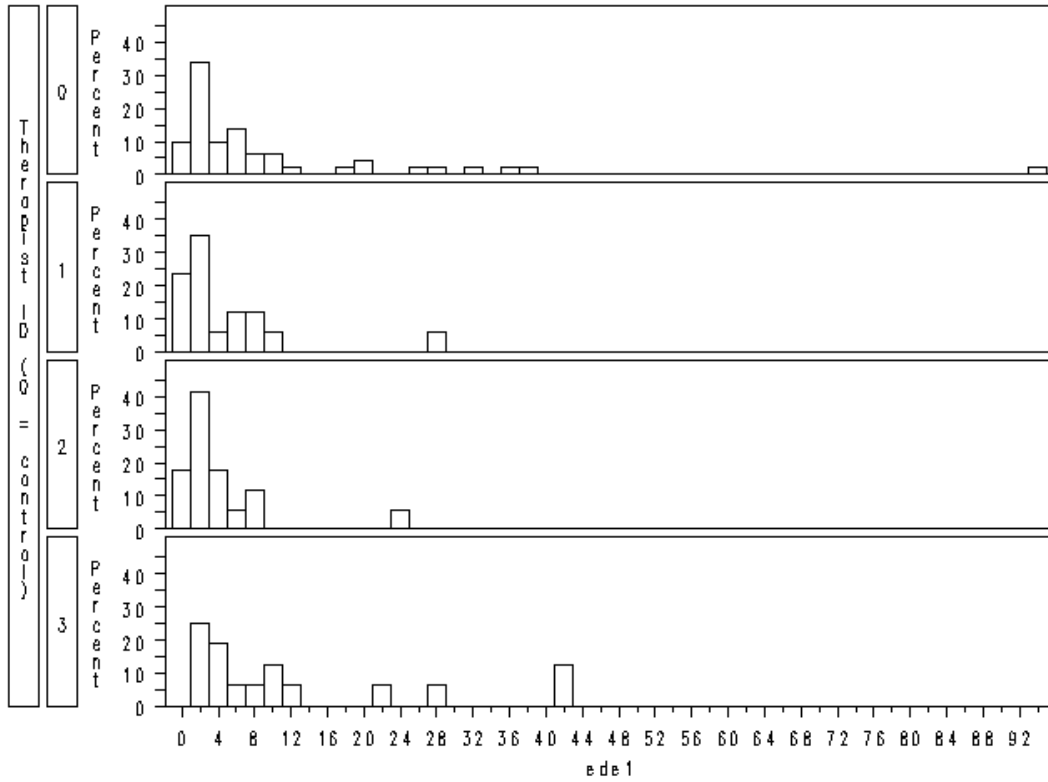


Table 2. Illustrative simulation: Mean, SD and median frequency of binge eating per month by visit (baseline vs. post-treatment) and therapist (therapists 1-3 denote treatment therapist, therapist 0 denotes the control condition).

Therapist ID	Baseline			Post-treatment			mean base-post
	Mean	SD	Median	Mean	SD	Median	
0 (control)	10.0	14.0	4.9	9.6	15.4	4.7	0.5
1	11.9	12.1	7.1	4.6	6.5	2.2	7.3
2	7.7	10.1	3.7	4.1	5.9	1.9	3.6
3	16.2	16.1	10.0	12.7	13.7	7.6	3.5

The information described above is for complete data, assuming no dropout (attrition) between baseline and post-treatment. However, when designing this study, we assumed a dropout rate of 30% between baseline and post-treatment (again, this is a conservative assumption based on the worst observed scenario in past studies). The SAS code below shows how dropout is simulated. For 30% of the patients (selected at random), post-treatment frequency of binge eating (ede1) was converted to missing, representing attrition between baseline and the post-treatment measurement. This is illustrated in the SAS code below for the example simulation.

```

* Simulate the assumed degree of dropout at the post-treatment visit;
* (visit 1);
data sim2;
set sim;
keep group c_group id ede visit therapist t1-t&n_therapists;
*****;
* Simulate 30% dropout at visit 1,;
*****;
drop=ranuni(&seed);
dropout=0;
drop1=0;

do visit=0 to 1;
  if visit=0 then ede=ede0;
  if visit=1 then do;
    if drop<=0.30 then do; drop1=1; dropout=1; end;
    if dropout=0 then ede=ede1;
    else ede=.;
  end;
  output;
end;
run;

proc sort data=sim2;
by group id;
run;

data base(keep=group id ede rename=(ede=base_ede)) otherv;
set sim2;
if visit=0 then output base;
else output otherv;
run;

```

In the next data step, baseline and post-treatment frequency of binge eating (including attrition) are log-transformed for subsequent statistical analysis.

```
* Log-transform frequency of binge eating for purposes of;  
* statistical testing;  
data monte;  
merge otherv(in=a) base;  
by group id;  
if a;  
ln_ede = log(1+ede);  
ln_base_ede = log(1+base_ede);  
run;
```

Prior to statistical test of the therapist effect, multiple imputation is used to handle the missing data, as shown in the PROC MI code below.

```
* Multiple imputation to handle missing data at post-treatment;  
* (The MAR/MCAR assumption is met due to the mechanism that generated;  
* the missing data at post-treatment).;  
ods listing;  
proc mi data=monte out=monte_mi nimpute=5 seed=64321;  
%let n_therapists_minus1 = %eval(&n_therapists-1);  
var ln_ede ln_base_ede group t1-t&n_therapists_minus1;  
mcmc nbiter=2000 niter=1000;  
run;  
  
proc sort data=monte_mi;  
by _Imputation_;  
run;
```

Finally, a statistical test of the therapist effect is conducted. The null hypothesis is: Δ therapist #1 = Δ therapist #2 = Δ therapist #3, where Δ is the difference in frequency of binge eating between baseline and control for a given therapist. Due to multiple imputation to handle missing data at post-treatment, PROC MIANALYZE is used for the statistical test. This is accomplished in the macro tests_with_multiple_imputation shown below.

```
* Statistical test of the therapist effect, taking into account the fact;  
* that multiple imputation was used to handle missing data post-treatment.;  
options mprint;  
%macro tests_with_multiple_imputation;  
ods listing;  
proc reg data=monte_mi(where=(visit=1)) outest=outreg covout;  
by _Imputation_;  
model ln_ede = ln_base_ede  
%do trns2= 1 %to &n_therapists;  
t&trns2  
%end;  
;  
run;  
quit;
```

```
ods listing;

ods listing;
proc mianalyze data=outreg;
modeleffects
%do trns4 = 1 %to &n_therapists;
t&trns4
%end;
;
%let n_v_minus1 = %eval(&n_therapists-1);
test1: test
%do trns3 = 1 %to &n_therapists_minus1;
t&trns3 =
%end;
t&n_therapists / mult;
ods output TestMultStat=TestMultStat;
run;
%mend tests_with_multiple_imputation;
%tests_with_multiple_imputation;
options nomprint;
```

In this example simulation, there was a significant therapist effect – therapists differed in reduction of binge eating of their patients between baseline and post-treatment.

```
* In this simulation, the therapist effect is significant.;
data _TestMultStat;
set TestMultStat;
if probf^=. then do;
  * Assume 2-tailed test;
  if probf<0.05 then sig=1;
  else sig=0;
end;
run;
```

For the Monte Carlo simulations, the procedure above was repeated 500 times for each scenario (varying seeds for random generation as well as the parameters of the scenario) and results were summarized across the simulations.

RESULTS

Power to detect therapist effects in each scenario is shown in the table below.

Power to detect therapist effects

Therapist effect size	Number of therapists			
	2	3	4	5
Large	92%	77%	70%	58%
Medium	75%	46%	45%	37%
Small	44%	28%	25%	23%

Only with the largest effect size, with two therapists, did power to detect the therapist effect exceed the conventional level of 80% (although it came close with 3 therapists and a large effect size, and 2 therapists and a medium effect size).

CONCLUSION

Power to detect therapist effects increases as the size of the therapist effect increases, and as the number of therapists decreases. The former finding is rather obvious, although the present study makes a contribution by quantifying the power under specific scenarios. The latter finding also makes sense – the smaller the number of therapists, the greater the number of patients within each (because the total patient sample size is fixed in this situation), thus the more information available to make conclusions about the effectiveness of the therapist.

Generally, if it is of interest to estimate therapist effects and the therapist effects are estimated as fixed (as opposed to random therapist effects), it is better to use a smaller number of therapists, in order to have greater precision to estimate the effectiveness of each therapist. In larger studies, it may be possible to adequately assess therapist effects across a greater number of therapists. In such studies, it may be possible to use a random effects approach. This would enable a broader scope of inference, especially if the therapists can be viewed as a random sample from a larger population of therapists. This would be the gold standard of design to estimate therapist effects, but unfortunately it may not always be feasible in small trials.

Do the results presented here mean that in small trials, the number of therapists should always be minimized to increase precision? We do not think so. There are practical reasons to consider involving a greater number of therapists. For example, therapists may drop out during the course of a trial. It is useful to have a pool of trained backup therapists available if needed. If there are only, say, two therapists involved in a trial, and one therapist stops participating, the trial timeframe is likely to be extended (because all treatment falls to a single therapist) and the trial itself might be jeopardized. For this reason alone, it is probably useful to consider involving at least 3 or 4 therapists, even in small trials. Further, although one may not be able to formally generalize results to a larger population of therapists based on a small trial with a non-random sample of therapists and a fixed effects analysis, using a greater number of therapists may provide at least a qualitative sense of inter-therapist variation in ability to administer a treatment. This may be helpful for future research, including the design of larger trials aimed specifically at estimating therapist effects. In other words, inter-therapist variation in a small trial could be used as pilot data to support the design of a larger trial.

In the present trial, analysis of therapist effects was a secondary goal (in fact, it was an ad hoc goal that was not even considered at the time that the trial was designed). If estimation of therapist effects is a primary goal, it may be worthwhile to design a trial with a large enough sample of therapists to enable a convincing random effects analysis.

REFERENCES

Brown H, Prescott R. Applied mixed models in medicine. 1999. New York: Wiley.

Grissom RJ, Kim JJ. Effect sizes for research: A broad practical approach. 2005. Mahwah, NJ: Earlbaum.

Loeb KL, Wilson GT, Labouvie E, Pratt EM, Hayaki J, Walsh BT, Agras WS, Fairburn CG. Therapeutic alliance and treatment adherence in two interventions for bulimia nervosa: a study of process and outcome. *J Consult Clin Psychol*. 2005 Dec;73(6):1097-107.

Jannoun L, Munby M, Catalan J, Gelder M. A home-based treatment program for agoraphobia: replication and controlled evaluation. *Behavior Therapy*. 1980, 11, 294-305.

Lewis M, Morley S, van der Windt DAWM, Hay E, Jellema P, Dziedzic K, Main CJ. Measuring practitioner/therapist effects in randomised trials of low back pain and neck pain interventions in primary care settings. *Eur J Pain*, 2010; doi:10.1016/j.ejpain.2010.04.002.

Mathews AM, Johnston DW, Lancashire M, Munby M, Shaw PM, Gelder MG. Imaginal flooding and exposure to real phobic situations: treatment outcome with agoraphobic patients. *Br J Psychiatry*. 1976 Oct;129:361-71.

Muthén LK, Muthén BO. How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*. 2002 4, 599-620.

Walwyn R, Roberts C. Therapist variations within randomized trials of psychotherapy: Implications for precision, internal and external validity. *Statistical Methods in Medical Research*. 2010; 19: 291-315.

Walters SJ. Therapist effects in randomized controlled trials: What to do about them. *Journal of Clinical Nursing*. 2009; 19: 1102-1112.

Webb CA, DeRubeis RJ, Barber JP. Therapist adherence/competence and treatment outcome: A meta-analytic review. *J Clinical and Consulting Psych*. 2010; 78(2): 200-211.

Williams MS, Ebel ED, Wagner BA. Monte Carlo approaches for determining power and sample size in low-prevalence applications. *Prev Vet Med*. 2007 Nov 15;82(1-2):151-8. Epub 2007 Jun 25.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Doug Thompson
Assurant Health
500 West Michigan
Milwaukee, WI 53203

Work phone: (414) 299-7998

E-mail: Doug.Thompson@Assurant.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.