

Cost-Effective Design for Dental Randomized Clinical Trials with Longitudinal Observations

Arthur F. Hefti¹ and Mark C. Yang²

¹College of Dentistry, The Ohio State University, Columbus, OH 43210
and ²Department of Statistics, University of Florida, Gainesville, FL 32610

Short Title: Cost-effectiveness of randomized clinical trial

Abstract

In general, randomized clinical trials (RCT) in dentistry involve longitudinal observations. In such studies, the total cost is a function of the number of study subjects and visits, the study duration, and the type and number of examinations at each visit. In this paper, we derived the minimum cost design for longitudinal RCTs with two treatment arms and multiple visits. We optimized the number of subjects, visits, and repeated measurements under the constraints of the requirements for statistical significance, power, and minimum total study cost. A SAS macro was written and made available on the World Wide Web, so interested clinical investigators can easily find optimal designs. The application of the program is illustrated using an example.

Keywords: Randomized Clinical Trial - Cost - Cost-Effectiveness - Sample Size - Repeat Measurements - Dental

Corresponding author: A.F. Hefti, 4133 Postle Hall, College of Dentistry, The Ohio State University, Columbus, OH 43210, USA.

Introduction

Pocock (1) defined the randomized clinical trial (RCT) as “a form of planned experiment, involving patients and designed to elucidate the most appropriate treatment of future patients with a given medical condition.” The goal of a clinical trial can be described as the testing of a new therapy’s safety and efficacy in treating a health condition or fighting a disease. In the dental area, RCTs are frequently conducted to evaluate oral health care products, e.g., a new anti-tartar toothpaste, a mouth rinse with improved antimicrobial properties, or to assess the benefits of novel therapeutic procedures in direct comparison to the present standard of care. It is important to emphasize that RCTs by definition embrace all principles of the scientific method, i.e., the statement of a specific hypothesis, a written protocol that details the study design, execution and analysis, the data collection, management and analysis, and the reporting of results and their interpretation. Recently, fundamental aspects of the RCT have been expertly reviewed from an ethical point of view (2), including subject recruitment, study design and sample size determination. Furthermore, a comprehensive treatise on clinical trial design and analysis in periodontics has become available following a symposium sporting the same title (3).

Similar to the RCT in medicine, the RCT in dentistry (DRCT) can be a major undertaking with respect to planning, execution, monitoring, and also cost. In search of better therapy and ways to contain the cost of health care, RCTs have become the premier instrument to determine therapeutic benefits scientifically. The rapid development of what can be called the trial industry is reflected in the numbers presented in Table 1. Clearly, in the 1980s and 1990s there has been a constant increase in the number of RCTs published in peer reviewed journals. For obvious reasons, accurate cost projection is critical in the competitive world of clinical trials (4,

5). To our knowledge, however, information on clinical trial cost calculation and cost optimization is not available in the dental literature. By and large, the cost of a DRCT depends on the difficulty of subject recruitment, trial size, the type of interventional procedures, the clinical measurements, study duration, and the associated paper trail. Mathematically, trial size and accuracy of measurement are directly related to each other. Given the significance level and effect size, they determine the level of power that can be achieved in a particular trial. Generally spoken, at a fixed level of power better measurement accuracy will reduce the required sample size.

The assessment of clinical dental parameters, e.g., caries, attachment level, probing depth, gingival inflammation, or plaque, is difficult and subject to sizable measurement error (6-12), which in turn affects the ability of a statistical test to detect a true difference. In clinical research, the magnitude of measurement error can be reduced by 1) performing additional (repeat) measurements in the same subject, 2) improving measurement technology, and 3) examiner training. Of course, any one of the proposed approaches will affect study cost and it is of critical interest to study the tradeoff between error size and sample size in view of minimal study cost.

Table 1: Result of a Medline search using the terms "clinical trial" (I) "clinical trial and cost" (II), "clinical trial and dental" (III), and "clinical trial and dental and cost" (IV)

Publication Year	(I)	(II)	(III)	(IV)
1981 - 1985	472	8	16	< 1
1986 - 1990	709	17	28	< 1
1991 - 1995	1066	47	41	1
1996	1248	76	41	0

In this paper, we will present a costing model for DRCTs with fixed or flexible number of examination visits. We will optimize study cost by suitable choice of sam-

ple size and number of measurement repeats at any required trial power. We will show that the inclusion of such calculations in the planning phase of a DRCT can substantially lower the cost. In the first part of the paper we will develop the theoretical background. In the second part we will present examples.

Material and methods

DRCT costing model concept: There are many ways of budgeting the cost for a clinical trial. We approached DCTR cost estimation by identifying five different cost objects:

- (A) one-time expenses;
- (B) expenses that depend on study duration and accrue even if no subjects were enrolled in the study;
- (C) expenses due to subject recruitment;
- (D) expenses originating from baseline activities, subject up front compensation, and therapeutic intervention;
- (E) expenses for examination activities at visits subsequent to baseline (e_1), and due to repeat measurements (e_2).

Each cost object may include several cost items. Some of the cost items do not change with the enrollment size and study length. They are considered fixed cost. Other cost items vary directly with the size of the enrollments and/or study length and are referred to as variable cost. A list of possible cost items is presented in Table 2. Our costing model includes direct cost only. Usually, indirect cost is a fixed

percent of direct cost and subject to negotiation between the funding agency and the research organization. Therefore, indirect cost were not included in the present model.

Table 2: Expenses that may occur during a DRCT. Letters in subtitles refer to cost objects A to E as described in Material and Methods.

Cost Objects and Items	
<u>A: One-time expenses</u>	<u>C: Cont.</u>
Protocol development	Screening
CRF development and testing	
Equipment	<u>D: Baseline and intervention.</u>
Various reports	Baseline examination
IRB submission	Randomization
	Pre-treatment
	Administrative cost per patient
<u>B: Site administration and development</u>	Therapeutic Intervention
Utility	
Space lease	<u>E: Follow up</u>
Janitorial services	Follow-up examinations
Repairs	Repeat examinations
Administrative support	Subject compensation
	Compliance assessment
<u>C: Recruiting and screening</u>	Paper trail
Study advertisement	Data management, analysis
Pre-screening	
Subject information, informed consent	

Expenses included in cost object A are independent of study duration and sample size. Cost object B expenses are independent of the sample size, but related to total study length. For example, utility bills or space leasing fees will accrue independent of the number of subjects enrolled in the study. We assume such expenses will increase proportional to the number of subject visits. Cost object C reflects sample size and the difficulty of recruiting an adequate subject panel. For example, while the recruitment of several hundreds of 20 to 65-year-old subjects exhibiting mild to

moderate gingivitis is easy and can be accomplished in a short time, the recruitment of 50 subjects of similar age and exhibiting several sites of moderate dentinal hypersensitivity may turn out very difficult and time consuming. Cost object D depends on sample size, as well as procedure type and frequency. A pre-treatment of subjects before they receive therapeutic intervention is often required by the study protocol. For example, an anti-tartar study may require that subjects present free of any visible calculus before they start using the products. Also, the expenses associated with the therapeutic intervention will be posted under cost object D. Cost object E, which includes two components e_1 and e_2 , varies with type, extent and frequency of the clinical examinations, and with sample size. Component e_1 represents expenses associated with single follow-up examinations while component e_2 covers expenses generated by repeat measurements.

In summary, the total cost of a DRCT can be represented by:

$$TC = A + Bk + Cn/q + Dn + nke_1 + n(k + 1)me_2. \quad [1]$$

Equation [1] includes several design parameters. Parameter $k(t = 0, 1, 2, \dots, k)$ identifies the number of visits that follow the baseline visit. The success rate during the recruiting process is reflected in parameter q , and n is the number of study subjects enrolled at Baseline. The expected number of subjects who must be approached in order to enroll n subjects is the quotient n/q . The number of measurement repeats during a single visit is m . Parameters k , n and m will be determined by the algorithm presented below.

DRCT longitudinal modeling, testing and cost optimization. There are a large number of longitudinal models that can be applied to repeated measures (13-16). Let $\eta_i(t)$ and $\zeta_j(t)$ denote the true response of the i th subject in treatment group 1 and

the j th subject in treatment group 2, respectively. Let their observed values at visit t be

$$x_i(t) = \eta_i(t) + \epsilon_{it}; \quad \text{and} \quad y_j(t) = \zeta_j(t) + \epsilon_{jt}, \quad [2]$$

where ϵ_{it} and ϵ_{jt} are observation discrepancies. The indexes $i = 1, 2, \dots, n_1$, $j = 1, 2, \dots, n_2$, and $t = 0, 1, 2, \dots, k$, where n_1 and n_2 are the numbers of subjects in the two treatment groups. We also assume the time intervals between visits are equal and $t = 0$ means the baseline response. The two functions $\eta_i(t)$ and $\zeta_j(t)$ are subject and treatment dependent. There is no universal way to model them without some knowledge of the relation between the treatments and the disease. We limit our consideration to the situation that during the treatment period, $\eta_i(t)$ and $\zeta_j(t)$ are monotonic functions which include the following possibilities,

1. If the treatment is effective, the improvement gap between treatment and no treatment increases with time.
2. The treatment stabilized the condition, but when there is no treatment, the condition deteriorates with time. (Note that no treatment (control) is considered a “treatment”. Also, ‘treatment and control’ can be a new treatment and a standard treatment.)
3. If there is no effect by the treatment or the control, function $\eta_i(t)$ or $\zeta_j(t)$ is a constant, which is a special case of a monotonic function.

Obviously, improvement cannot be monotonic indefinitely. However, we may expect that a monotonic response subject to discrepancy and error is a good approximation during a limited period of time. Again, there are many monotonic functions. To choose one model for many subjects is at best an approximation. We feel that to use a linear model as the first degree approximation is reasonable if there is no evidence

to reject it (see for example the gradual model in (17) and (18)). Thus we let

$$\eta_i(t) = \alpha_i + \gamma_1(i)t, \quad \text{and} \quad \zeta_j(t) = \beta_j + \gamma_2(j)t. \quad [3]$$

This model covers subject variation, i.e., even in the same treatment group, we do not expect that slopes $\gamma_1(i)$ or $\gamma_2(j)$ are the same for all subjects (see Fig. 1(I)). We thus assume that $\gamma_\ell(i)$ is normally distributed with mean μ_ℓ and variance σ_s^2 for $\ell = 1, 2$. Moreover, we consider the ϵ 's in [2] are independent and identically distributed as normal random variables with mean 0 and variance $\sigma_r^2 + \sigma_m^2/m$ where σ_r^2 is the discrepancy between the true mean and the linear line and σ_m^2 is the measurement variance (see Fig. 1 (II)). Obviously, the discrepancy cannot be decreased by repeated measurements at the same visit, but the measurement variance is reduced to σ_m^2/m when the response is measured m times. If the linear model [3] is a perfect fit over time except measurement error, then $\sigma_r^2 = 0$.

Testing the treatment effect is equivalent to testing

$$H_o : \mu_1 = \mu_2 \quad \text{against} \quad H_1 : \mu_1 \neq \mu_2. \quad [4]$$

Since the intercepts α_i and β_j in equations [3] are unknown and cannot be assumed to be equal in most practical situations, using a maximum likelihood approach for hypothesis testing [4] would be very complicated. A very intuitive approach that produces a simple and exact statistical test is to estimate $\hat{\gamma}_1(i)$ and $\hat{\gamma}_2(j)$ using ordinary regression estimates,

$$\hat{\gamma}_1(i) = \sum_{t=0}^k t(x_i(t) - \bar{x}_i)/a, \quad i = 1, 2, \dots, n_1;$$

$$\hat{\gamma}_2(j) = \sum_{t=0}^k t(y_j(t) - \bar{y}_j)/a, \quad j = 1, 2, \dots, n_2, \quad [5]$$

where $a = k(k+1)(k+2)/12$, $\bar{x}_i = \sum_{t=0}^k x_i(t)/(k+1)$ and $\bar{y}_j = \sum_{t=0}^k y_j(t)/(k+1)$. Since $\hat{\gamma}_1(i)$ and $\hat{\gamma}_2(j)$ both have normal distributions with the same variance but different means μ_1 and μ_2 , the commonly used test is to reject H_0 at α level of significance if

$$T = \frac{|\bar{\gamma}_1 - \bar{\gamma}_2|}{S_p \sqrt{1/n_1 + 1/n_2}} > t_{n_1+n_2-2, \alpha/2}, \quad [6]$$

where

$$\begin{aligned} S_p^2 &= \left(\sum_{i=1}^{n_1} (\hat{\gamma}_1(i) - \bar{\gamma}_1)^2 + \sum_{j=1}^{n_2} (\hat{\gamma}_2(j) - \bar{\gamma}_2)^2 \right) / (n_1 + n_2 - 2), \\ \bar{\gamma}_1 &= \sum_{i=1}^{n_1} \hat{\gamma}_1(i) / n_1, \quad \bar{\gamma}_2 = \sum_{j=1}^{n_2} \hat{\gamma}_2(j) / n_2, \end{aligned}$$

and $t_{n_1+n_2-2, \alpha/2}$ represents the upper $\alpha/2$ quantile of a t -distribution with $n_1 + n_2 - 2$ degrees of freedom. We can also find the minimum cost for the execution of this clinical trial to satisfy a given significance level (α) and power requirement ($(1 - \beta)$) at $|\mu_1 - \mu_2| = \Delta\mu$. Because there are too many independent variables in the minimum cost equation, it is not feasible to provide a comprehensive designs table. Furthermore, although the derivation of the optimum design is not mathematically difficult, we could not find any published tables or descriptions (see e.g. (19)) that could serve the dental research community conveniently. Therefore, we wrote an SAS macro to do this. When the required parameters are entered into this program, the output will be the optimal design with number of visits k , number of subject, $n_1 = n_2 = n$, number of measurements m and the total cost for conducting this trial. Note that we let $n_1 = n_2 = n$ in the design stage, but use general n_1 and n_2 in the testing formula [6] because there are usually uneven dropouts during a clinical trial. The program

also provides an option to fix the number of patients, and/or the number of visits and/or the number of measurements. It happens in practice that the number of visits is fixed due to administrative necessity or the number of measurements is fixed at 1 due to clinical reasons. In particular, if k is fixed at 1, we expect $\Delta\mu$ difference between treatments at the final (only) visit after baseline. The use of the SAS macro and its World Wide Web site are provided in Appendix I. A derivation of the optimal design is given Appendix II.

Examples.

Suppose the costs A, B, C, D, e_1 and e_2 , the screening success rate q and the standard deviations $\sigma_s, \sigma_r, \sigma_m$ can be reasonably estimated. An example is listed in Table 3. The figures for cost objects A through E were based on a clinical trial conducted at the UF Periodontal Disease Research Center. The objective of the study was to compare the effect of a metalloproteinase inhibitor and a placebo in the presence of scaling and root planning on clinical attachment level change in subjects with clinically diagnosed periodontal disease. Full-mouth attachment levels were assessed using manual probing. Repeat measurements were planned only for the sites exhibiting clinically diagnosed periodontal disease. The double blind study was set up for a maximum of 9 months duration and conducted in a 3000 sqft clinical research facility. The between visit interval was set at three months. Six operatories were used full time during each examination phase. Two calibrated periodontists were responsible for the examinations and were assisted by two recorders. A 90-minute slot was scheduled per examination. Two experienced dental hygienists provided scaling and root planning therapy at baseline following the periodontal examinations. Scaling and root planning was restricted to periodontal pockets exhibiting inflammation and/or

clinically detectable calculus, and did not exceed 60 minutes per quadrant. Subjects were compensated \$75 per visit. A certified clinical research associate coordinated the study. Data entry, management and analysis were performed by the funding agency and not included in the budget.

Table 3: A list of costs and standard deviations for a dental clinical trial. With the exception of A and B all costs were calculated per subject.

A	B	C	D	e_1	e_2	q	σ_s	σ_r	σ_m
\$8000	\$3709	\$129	\$355	\$250	\$45	0.1	1.5	0.5	1.0

Moreover, suppose we require $\alpha = 0.05$ and power = 0.90 when the slope difference $|\Delta\mu| \geq 1.0mm/\text{visit}$. Then the output of the computer program shows that the optimal (minimum cost) design will include

$$\begin{aligned}
 n &= 57 && \text{subjects per treatment} \\
 m &= 2 && \text{repeated measurements per visit, and} \\
 k &= 2 && \text{visits.}
 \end{aligned}$$

The total cost will amount to \$153,073 (see Appendix I).

If the number of visits is fixed at, e.g., $k = 3$, then the minimum cost design includes $n = 54$ subjects per treatment and $m = 1$ measurement per visit. The total cost would amount to \$158,117. Apparently, the tradeoff is that more visits can result in a smaller number of subjects, but due to the large subject variation, the reduction in subject costs does not fully compensate the increase in cost due to more visits.

We should expect a smaller number of subjects and a greater number of visits if σ_s is smaller and σ_r larger. This is indeed the case. When σ_s in the table is decreased to 1.0, σ_r increased to 1.0, and all the other parameters remain unchanged then the optimal design would include

$n = 31$ subjects per treatment,
 $m = 1$ measurements per visit, and
 $k = 3$ visits.

The total cost of such a clinical trial would amount to \$98,952.

Fig. 2 shows how optimal designs are affected by subject variation (σ_s), measurement error (σ_m), and linear model discrepancy (σ_r). The values used for the computation are the same as in Table 3 except the variable displayed on the x-axis. We can see that σ_s affects the number of subjects n more than the number of visits k and the number of measurement replicates m . Basically, the measurement variation σ_m affects only the number of measurement repeats m . The linear discrepancy σ_r affects both k and m , and in this case, the numbers of repeats and visits compensate one another in reducing the model fitting error. In Figure 2, the centers of the x-axes correspond to the values used in Table 3. Thus, this type of graphical information can be used to assess the robustness of the design if the size of the various errors is not well known at the design stage.

Notice that the optimal figures of m , n , and k depend not only on the variances, but also on the size of the various cost objects. Thus, if the cost structure changes, the relation between m , n , k and the σ 's will also change. The computer program, of course, permits to explore the effects of such changes with no difficulty.

Discussion

Though we have laid down a reasonable model for longitudinal clinical trial, there are several practical situations that require additional consideration.

1. Usually, a two sided test [4] is used in clinical trial as required by FDA. The

reason is that we should not bias the test in favor of the assumption that the treatment is effective. If a one sided test is to be used, then one can still use our program by doubling the α level, e.g., if α is set at 0.05 for a one sided test, the α input for our program should be 0.10. The power in this case may not be exact, but the discrepancy is usually very small.

2. The effect of the recruitment success rate q to the total cost is reflected in the cost formula [1] with its expected value Cn/q . The real cost can be smaller or larger depending on the luck during the recruitment. For large n , the variation above or below the expected number should be negligible, but for small n , a negative binomial distribution should be used, i.e., the total number of subjects to be screened is a random variable with a negative binomial distribution with parameters n and q (20). The user has to use this distribution to compute the risk of under and over estimation of recruitment cost.
3. When the clinical trial is designed for multipurpose, i.e., there is more than one outcome to be measured, the sample size requirement for each measurement should be computed. The required sample sizes are usually different and one has either to use the most important outcome as the guideline, or to choose a sample size that satisfies all of them.

Conclusion

The minimum cost design for a general dental clinical trial of two treatments with longitudinal observations is derived. Tables are too voluminous to list because of the large number of cost objects but a computer program is made available for interested users.

References

1. Pocock SJ. *Clinical Trials. A Practical Approach.* John Wiley & Sons Ltd. Chichester, 13th reprint, 1996; 1.
2. Dennison DK. Components of a randomized clinical trial. *J Periodont Res* 1997; 32:430-438.
3. AAP. 1996 Joint Symposium on Clinical Trial Design and Analysis in Periodontics. *Ann of Periodont* 1997; 2(1):1-369.
4. Lesser ML. Guidelines for budgeting biostatistics involvement in research projects. *Statistics in Medicine* 1996; 15:2127-2133.
5. Piantadosi S, Patterson B. A method for predicting accrual, cost, and paper flow in clinical trials. *Controlled Clinical Trials* 1987; 8:202-215.
6. Pine CM, ten Bosch JJ. Dynamics of and diagnostic methods for detecting small carious lesions. *Caries Res* 1996; 30:381-388.
7. Badersten A, Nilveus R, Egelberg J. Reproducibility of probing attachment level measurements. *J Clin Periodont* 1984; 11:475-485.
8. Goodson M. Clinical measurements of periodontitis. *J Clin Periodont* 1986; 13:446-455.
9. Marks RG, Magnusson I, Taylor M, Clouser B, Maruniak J, Clark WB. Evaluation of reliability and reproducibility of dental indices. *J Clin Periodont* 1991; 20:54-58.

10. Clark WB, Magnusson I, Namgung YY, Yang MCK. The strategy and advantage in use of an electronic probe for attachment measurement. *Advances in Dental Res* 1993; 7(2):152-157.
11. Hefti AF. Periodontal probing. *Critical Reviews in Oral Biology and Medicine* 1997; 8(3):336-356.
12. Zappa U, Simona C, Graf H, Case D, Thomas J. Reliability of single and double probing attachment level measurements. *J Clin Periodont* 1995; 22:764-771.
13. David H, Martin C. *Practical longitudinal data analysis*, Chapman & Hall, London, 1996.
14. van Eye A. (Ed.) *Statistical methods in longitudinal research*, Vol. I, II, Academic Press, Inc., Boston, 1996.
15. Laird NM, Ware JH. Random-effect models for longitudinal data. *Biometrics* 1982; 38:963-974.
16. Yang MCK, Carter R. One way analysis of variance with time series data, *Biometrics*, 1983; 39:747-751.
17. Yang MCK, Marks RG, Clark WB. and Magnusson, I. Prediction power of various models for longitudinal attachment level change. *J Clin Periodont* 1992; 19:77-83.
18. Yang MCK, Namgung YY, Marks RG, Magnusson I, Clark WB. Change detection on longitudinal data in periodontal research. *J Periodont Res*, 1993; 28:152-160.
19. Desu MM, Raghavarao D. *Sample Size Methodology*, Academic Press, Inc., Boston, 1990.

20. Johnson NL, Kotz, S, *Discrete Distributions*, Wiley, New York, 1969.

Appendix I. Optimal Design by a SAS program

This program can be found in the clinical-trial directory of

www.stat.ufl.edu/~yang/

The program, data file, and output file are saved as “cdesign.sas”, “cdesign.dat” and “cdesign.out”, respectively. The data file is subject to change to fit a user’s need. One way to check whether your system is compatible to our is to run “cdesign.sas” and see whether your output is exactly the same as “cdesign.out”.

The input file for the three cases in the example is available as file cdesign.dat. It has the following structure:

```
----- File cdesign.dat -----
1----!!!!!!!!!!!!!!< Be careful >!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
2- First 23 lines in this input file are comment lines. Do not change them.
3- Input data starts on line 24.
4!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
5/****Definitions for input parameters *****/
6/* A   -one time cost           : */
7/* b   -duration cost/visit     : */
8/* c   -screening cost/patient  : */
9/* d   -cost/patient            : */
10/* e_1 -administration cost/vis/pat: */
11/* e_2 -exam cost/vis/pat/measure : */
12/* q   -rate of success        : */
13/* dr  -difference bt two slopes : */
14/* sig_s -subject s.d          : */
15/* sig_m -measurement s.d      : */
16/* sig_r -time variation wrt slope : */
17/* alpha -significance level    : */
18/* power - power for the test   : */
19/* n_min, n_max: minimum & max. numbers for subjects/treatment: */
20/* k_min, k_max: minimum & max. numbers for visits           : */
21/* m_min, m_max: minimum & max. numbers for measurements   : */
22/*****
```

```

23 [A]      [b]      [c]      [d]      [e_1]  [e_2]  [q]      [dr]      [sig_s]  [sig_m]  [sig_r]
[alpa] [power] [n_min] [n_max] [k_min] [k_max] [m_min] [m_max]
8000.0 3709.0 129    355    250    45.0  0.1    1.00    1.50    1.0    0.5
0.05  0.90  2      500    1      20    1      4
8000.0 3709.0 129    355    250    45.0  0.1    1.00    1.00    1.0    1.0
0.05  0.90  2      500    1      20    1      4
8000.0 3709.0 129    355    250    45.0  0.1    1.00    1.00    1.0    1.0
0.05  0.90  2      500    3      3      1      4

```

The first 23 lines are descriptions. Note that line 23 printed as two lines here is actually a single line from [A] to [m__max]. The same happens in printing for the next three data lines. This is clear when the data set “cdesign.dat” is used. The notation or the terms in the input follow the paper. Input rows 19, 20, and 21 permit to select a fix number design if that minimum allowed number is the same as the maximum number. Users can try simultaneously any number of cases. There is no restriction on the number of digits in the input figures, but they must be in the right order and the figures must be separated by at least a one-space gap as required in standard SAS input. The last three data lines are the three cases mentioned in the example section.

The output of the first case is shown in cdesign.out below. The result is self explanatory. The last row “pw-power at optimal” is the power which should be very close to the input requirement [power]. It may differ slightly due to the discreteness of n , m and k . These outputs have been checked by a simulation program.

```

----- File cdesign.out -----
*****
***** input parameters *****
*****
INP_CH                                INP_PA
**** A      -one time cost              : 8000.0000
**** b      -duration cost/visit         : 3709.0000
**** c      -screening cost/patient      : 129.0000
**** d      -cost/patient                : 355.0000
**** e_1    -administration cost/vis/pat : 250.0000

```

```

**** e_2   -exam cost/vis/pat/measure           :    45.0000
**** q     -rate of success                     :    0.1000
**** dr    -difference bt two slopes            :    1.0000
**** sig_s -subject s.d.                       :    1.5000
**** sig_m -measurement s.d.                  :    1.0000
**** sig_r -time variation wrt slope           :    0.5000
**** alpha -significance level for test        :    0.0500
**** power -required power for the test        :    0.9000
**** n_min -min. no. subj./treatment allowed   :    2.0000
**** n_max -max. no. subj./treatment allowed   :   500.0000
**** k_min -minimum number of visits allowed   :    1.0000
**** k_max -maximum number of visits allowed   :   20.0000
**** m_min -minimum no. of measurements allowed:    1.0000
**** m_max -minimum no. of measurements allowed:    4.0000

```

```

*****
***** optimal solution *****
*****

```

```

OUP_CH          OUP_PA
**** n  - number of individuals      :    57.0000
**** m  - number of measurements     :    2.0000
**** k  - number of visits           :    2.0000
**** TC - total cost                 :  153073.0000
**** pw - power at optimal           :    0.9043

```

Appendix II Derivation of Minimum Cost Design

By [2], [3] and $\gamma_1(i) \sim N(\mu_1, \sigma_s^2)$, we may write

$$x_i(t) = \alpha_i + \mu_1 t + \delta_i t + \epsilon_{it},$$

where $\delta_i \sim N(0, \sigma_s^2)$ and $\epsilon_{it} \sim N(0, \sigma_r^2 + \sigma_m^2/m)$.

It can be shown that $\hat{\gamma}_1(i)$ in the estimation formula [5] is normally distributed

with

$$\begin{aligned}
E\hat{\gamma}_1(i) &= \mu_1, \text{ and} \\
\sigma_0^2 &\equiv \text{Var}(\hat{\gamma}_1(i)) = \frac{1}{a^2} \mathbf{b}' \boldsymbol{\Sigma} \mathbf{b}, \\
\mathbf{b} &= [b_1, b_2, \dots, b_k], \quad b_i = i - k/2, \quad i = 1, 2, \dots, k, \\
\boldsymbol{\Sigma} &= (\sigma_{ij}) \text{ with} \\
\sigma_{ij} &= \begin{cases} ij\sigma_s^2 & \text{if } i \neq j \\ i^2\sigma_s^2 + \sigma_r^2 + \sigma_m^2/m & \text{if } i = j. \end{cases}
\end{aligned}$$

The distribution of $\hat{\gamma}_2(j)$ is the same except $E\hat{\gamma}_2(j) = \mu_2$. Consequently, under the alternative hypothesis, $\bar{\gamma}_1 - \bar{\gamma}_2 \sim N(\mu_1 - \mu_2, 2\sigma_0^2/n)$, or T in [6] has a noncentral t -distribution with $2n - 2$ degrees of freedom with noncentrality parameter

$$\lambda = \sqrt{n}(\mu_1 - \mu_2)/(\sqrt{2}\sigma_r).$$

Since n , k and m are discrete, our program does an exhaustive search on n , k , m . The output represents optimum n , k and m , satisfying the significance level and power requirements with the minimum total cost under [1].

When the fixed k -option is used, the search is limited to n and m .

Legends:

Fig. 1. (I) Possible *mean* responses of 4 subjects 1, 2, 3, and 4 in treatment group 1 when the linear response model by Eq. [3] is the correct model. The slopes are supposed to come from a normal population with mean μ_1 and variance σ_s^2 .

(II) Using subject 1 as an example, σ_r is the standard deviation of the time discrepancy from the ideal straight line, and σ_m is the standard deviation of the

measurement error.

(III) When the data are observed according to Eq. [2], the linear trends are blurred by the errors.

Fig. 2. Relation between optimum design numbers (n, k, m) and σ_s , σ_m and σ_r . Only the value in the x-axis is subject to change in any of the three figures. All the other values are the same as in Table 3. The points from top to bottom denote the changes of the required number of subjects n , the changes of the number of visits k and the changes of the number of measurements m . The numbers inside the border were those for the smallest and the highest value on the x -axes.



