

RegRand: Statistical software for the multiple-baseline design

MATTHEW J. KOEHLER and JOEL R. LEVIN
University of Wisconsin, Madison, Wisconsin

RegRand (Version 1.0) is Macintosh-based software that enables a researcher to conduct a non-parametric statistical analysis of the data from Koehler and Levin's (1998) recently described regulated randomization single-case multiple-baseline design. Regulated randomization design and analysis principles are reviewed in relation to an educational research application and a step-by-step illustration of them in relation to the RegRand program is presented.

The purpose of this article is to introduce some personal computer software associated with a recently described statistical procedure for analyzing the data from a *single-case*, multiple-baseline design. The multiple-baseline design is an often implemented research strategy of clinical and educational interventionists (e.g., Kazdin, 1992; Kratochwill & Levin, 1992). With this design, two or more experimental "units" typically consisting of either individuals (e.g., clients or students) or aggregates (e.g., small groups or classrooms) are (1) observed in a time-series fashion during a *control* or *baseline* phase, and (2) sequentially (i.e., one unit at a time) administered some type of treatment or intervention, with observations continuing throughout the *intervention* phase. Although the multiple-baseline design itself has been accepted as part of the single-case researcher's armament for some time now, appropriate statistical techniques for analyzing the data produced by such designs have not. Building from the nonparametric randomization rationales previously offered by Edgington (1975) and Marascuilo and Busk (1988), Koehler and Levin (1998) recently proposed a versatile and potentially powerful analytic strategy. That strategy, known as *regulated randomization*, is briefly described below, followed by a detailed account of the downloadable personal computer software that can be used to conduct the corresponding inferential statistical analysis.

THE MULTIPLE-BASELINE DESIGN

We describe the multiple-baseline design in the context of introducing a new classroom intervention, presumed to result in more efficacious student outcomes relative to the preintervention classroom conditions. In what is known as the *basic AB design*, measurements of

the desired outcome are taken at m fixed intervals during the preintervention baseline (A) phase of the study and similar measurements are taken at n fixed intervals during the intervention (B) phase. Thus, for example, suppose that the number of positive academic classroom behaviors exhibited by all students in the classroom are recorded for 5 consecutive school weeks prior to the introduction of an intervention designed to increase those behaviors (the baseline phase). The intervention is implemented in the classroom, and the number of positive behaviors continues to be observed and recorded over the next 5 school weeks (the intervention phase). If the intervention is effective, one would expect to see a greater number of positive academic behaviors, on the average, during the intervention phase (between Weeks 6 and 10) than during the baseline phase (between Weeks 1 and 5).

The multiple-baseline design augments the basic AB design by attempting to replicate the desired outcome through a staggered introduction of the intervention to additional experiment units. Suppose, therefore, that a total of three classrooms were candidates for the academic behaviors intervention over the same 10-week period. In multiple-baseline fashion, all three classrooms might be observed during a baseline phase during Week 1. Then, as the intervention is introduced in Classroom 1 starting on, say, Week 2, Classrooms 2 and 3 would continue in their baseline phases. During Week 5, the intervention is introduced in Classroom 2, while Classroom 3 remains in its baseline phase. Finally, Classroom 3 begins receiving the intervention during Week 8, with all three classrooms continuing to receive the intervention and associated measurements through Week 10. Evidence supporting the efficacy of the intervention would be obtained if the increase in the average amount of positive classroom behavior occurs (1) across classrooms, and (2) within each classroom only after the intervention has been introduced to that particular classroom, and not before. It has been argued that the multiple-baseline design, with its replication and stagger features, provides much needed research qualities of both intervention generalizability (replications across units) and intervention discrimination

RegRand 1.0 can be downloaded for use on a PowerPC Macintosh system from the following Web address: www.education.wisc.edu/edpsych/research/regrand/. Address correspondence to M. J. Koehler, CEPSE, College of Education, Michigan State University, East Lansing, MI 48824 (e-mail: mkoehler@msu.edu).

(selectively timed responsiveness to the intervention), thereby enhancing the credibility of attributing the outcomes to the intervention (see, e.g., Levin, 1992, p. 217).

The Koehler-Levin Regulated Randomization Procedure

With the standard multiple-baseline design, the staggered intervention "start points" for each unit are typically determined on a fixed, a priori basis (e.g., Classroom 1's intervention phase begins during Week 2, Classroom 2's during Week 5, and Classroom 3's during Week 8). The randomization model proposed by Koehler and Levin (1998) requires that the researcher initially specify a potential start-point *interval* for each unit (say, during *either* Week 2 or Week 3 for one of the classrooms, during *either* Week 5 or Week 6 for another classroom, and during Week 8 for the last classroom). Following both random assignment of units to start-point intervals and random selection of an actual start point within the specified interval for each unit, the regulated randomization analysis proceeds to calculate the probability of obtaining a composite effect (a summary measure reflecting the difference between intervention and baseline outcomes) as extreme as, or more extreme than, the effect actually observed. If that probability is less than or equal to an a priori Type I error probability (e.g., $\alpha = .05$), one can conclude that there is a statistical difference between the intervention and baseline outcomes.

A personal computer program, RegRand, has been developed to compute all possible randomizations of the data on the basis of user-specified input, along with the associated probabilities.

THE REGRAND MICROCOMPUTER PROGRAM

For a detailed discussion of the multiple-baseline regulated randomization procedure, including its underlying logic, statistical basis, and examples of its application, readers should refer to the Koehler and Levin (1998) article. For a clinical research implementation of the regulated randomization design and analysis, see McKie (1998).

System Requirements

RegRand (Version 1.0) requires a PowerPC (PPC) Macintosh computer with 10 MB of free RAM. In theory, RegRand should work with Macintosh OS 7.5.1 or higher. However, this version of the program has only been tested under MacOS 8.0 and 8.1. At this time, there is no IBM PC-compatible version of the program.

Notation

In the microcomputer program, the following notation is used:

N = the number of independent units (e.g., classrooms, clients) in the study;

U_1, U_2, \dots, U_N refer to the individual units; and

T = the total number of repeated measurements on the outcome variable;

N and T are determined before data collection, and T is the same for all units (i.e., each U_i is measured T times). The number of baseline and intervention assessments can and will vary between units, however.

A Sample RegRand Design Specification and Analysis

Design specification. RegRand 1.0 features design windows, one of which is displayed in Figure 1. In Figure 1, there are $N = 3$ independent classrooms to which the intervention will be sequentially introduced (represented by the three 1s in the "Design" column). One randomly chosen classroom receives the intervention beginning at either the second or third outcome-assessment occasion (during either Week 2 or Week 3). A different randomly chosen classroom receives the intervention beginning at either the fifth or sixth measurement occasion (either Week 5 or Week 6). The last classroom receives the intervention beginning at the eighth measurement occasion (Week 8).

In this example, there are 24 possible assignments of classrooms to start points, as represented by Koehler and Levin's (1998) Formula 1, namely:

$$N! \prod_{i=1}^N k_i,$$

where N is the total number of units (i.e., classrooms) and k_i is the number of potential start points designated for a particular unit (i.e., $k_1 = k_2 = 2$ and $k_3 = 1$, in the present example). Thus, there are $3! \times 2(2)(1) = 6 \times 4 = 24$ possible outcomes in the resulting randomization distribution, consisting of the following sets of ordered potential start points:

- | | | | |
|------------|-------------|-------------|-------------|
| 1. [2,5,8] | 7. [3,8,5] | 13. [6,2,8] | 19. [8,3,5] |
| 2. [2,6,8] | 8. [3,8,6] | 14. [6,3,8] | 20. [8,3,6] |
| 3. [2,8,5] | 9. [5,2,8] | 15. [6,8,2] | 21. [8,5,2] |
| 4. [2,8,6] | 10. [5,3,8] | 16. [6,8,3] | 22. [8,5,3] |
| 5. [3,5,8] | 11. [5,8,2] | 17. [8,2,5] | 23. [8,6,2] |
| 6. [3,6,8] | 12. [5,8,3] | 18. [8,2,6] | 24. [8,6,3] |

The three numbers in each bracket indicate which of the T measurement occasions correspond, respectively, to an intervention start point that is randomly assigned to the first, second, and third randomly sampled classrooms. Thus, the first ordered triplet [2,5,8] represents an intervention start point just prior to the 2nd week's assessments for the first classroom, just prior to the 5th week's assessments for the second classroom, and just prior to the 8th week's assessments for the third classroom.

Analysis. Once the researcher has specified the number of measurement occasions and potential start points for each classroom (as above), one of the potential start

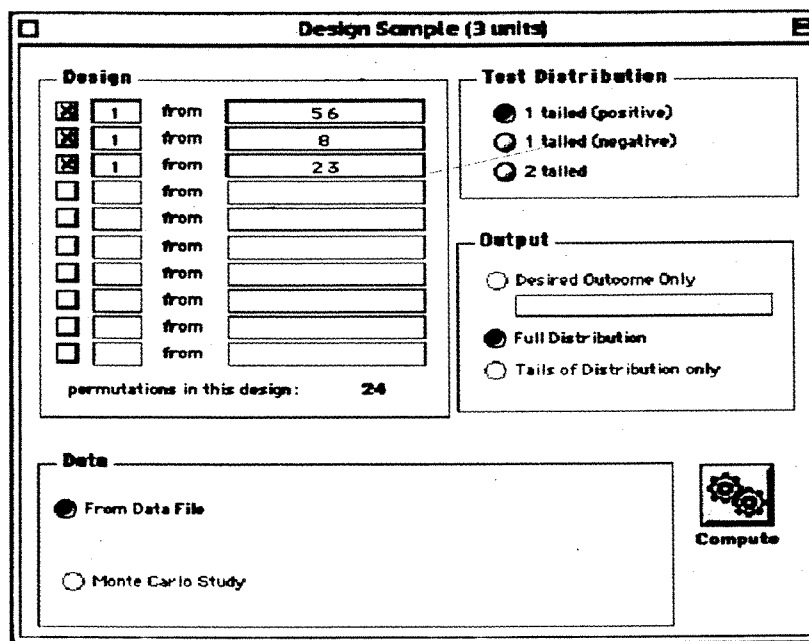


Figure 1. Sample design window from RegRand 1.0.

point designations should be randomly selected for the actual conduct of the study. Suppose, for example, that the ordered triplet [6,8,2] was randomly selected and that there are to be $T = 10$ outcome assessments for each classroom. The researcher would then collect the data, making sure that the intervention for U_1 (which in the present example represents C_1 , Classroom 1) commences just prior to the 6th week's measurements (T_6), thereby yielding five pre-intervention and five intervention periods. Similarly, C_2 would receive seven pre-intervention and three intervention periods, and C_3 would receive one preintervention and nine intervention periods.

In this example, suppose that the actual data collected were as indicated in Table 1. Koehler and Levin's (1998) regulated test statistic is computed as follows: First, for each classroom, the mean of the pre-intervention measures (M_A) is subtracted from the mean of the intervention measures (M_B). For Classroom 1 of the present example, $M_{1A} = (2.5 + 3 + 4.5 + 2 + 3)/5 = 3.0$, $M_{1B} = (3 + 4 + 3 + 5 + 3)/5 = 3.6$, and $M_{1B} - M_{1A} = 0.6$. Similarly, for Classroom 2, $M_{2A} = (2.5 + 1 + 3 + 2 + 4 + 1 + 2.5)/7 = 2.2857$, $M_{2B} = (5.5 + 4 + 4)/3 = 4.5$, and $M_{2B} - M_{2A} = 2.2143$; and for Classroom 3, $M_{3A} = 1$, $M_{3B} = (2 +$

$4 + 2.5 + 5 + 3 + 3.5 + 2.5 + 5 + 4)/9 = 3.5$, and $M_{3B} - M_{3A} = 2.5$. Then, the mean of these three mean differences is calculated as $(0.6 + 2.2143 + 2.5)/3 = 1.7714$.

The statistical significance of the test statistic is determined by comparing it to the complete randomization distribution of the statistic, given the design specifications and the observed data. In other words, the actual data are used to calculate the test statistic that corresponds to each of the other 23 potential start point assignments that could have been selected. Once this information is gathered, the outcomes are sorted and significance probabilities (p values) are assigned in such a way that the most "extreme" outcome is assigned a p value of $1/(\text{number of possible assignments})$, the second most "extreme" outcome is assigned a p value of $2/(\text{number of possible assignments})$, and so on. The definition of "extreme" depends on whether the researcher has specified on an a priori basis either a directional (one-tailed) test (in which case "extreme" is the largest number in the specified direction) or a nondirectional (two-tailed) test (in which case "extreme" is the largest number in absolute value). For the present example, assume that the researcher had specified $\alpha = .05$ and a one-tailed alternative in the pos-

Table 1
Hypothetical Data For Three Classrooms, 10 Weeks of
Outcome Assessment Per Classroom, and Intervention Start Points
Corresponding to Weeks 6, 8, and 2, for Classrooms 1, 2, and 3, Respectively

Classroom	Week									
	1	2	3	4	5	6	7	8	9	10
1	2.5	3.0	4.5	2.0	3.0	3.0	4.0	3.0	5.0	3.0
2	2.5	1.0	3.0	2.0	4.0	1.0	2.5	5.5	4.0	4.0
3	1.0	2.0	4.0	2.5	5.0	3.0	3.5	2.5	5.0	4.0

Note—| corresponds to the actual intervention start point for each classroom.

itive direction (i.e., that the intervention would *increase* performance), so that the $M_B - M_A$ test statistic would be positive. The resulting randomization distribution, presented in Table 2, reveals that the observed test statistic of 1.7714 is the most extreme of all 24 possible statistics and therefore is associated with a cumulative probability (p value) of .0417. Because .0417 is less than $\alpha = .05$, the hypothesis of no intervention effect can be rejected in favor of the alternative hypothesis stating that there is a positive effect of the intervention.

USING THE REGRAND PROGRAM

A regulated randomization analysis with RegRand 1.0 is conducted in the following three steps: (1) A regulated randomization design must be specified; (2) a text file containing the data for all units needs to be created; and (3) RegRand generates the distribution of the test statistic, given the design (from Step 1) and the data (from Step 2).

Specifying the Regulated Randomization Design

When RegRand 1.0 is first started, a blank (empty) untitled design window appears. Designs can also be saved and opened within the RegRand program by using the "Save" and "Open" commands under the "File" menu. Files can also be opened by double clicking on any previously saved design. We illustrate the design window operations again in Figure 1.

Partitions. The key to understanding a RegRand design is understanding partitions. Only "enabled" partitions

contribute to a design specification; a partition is enabled when the corresponding box is checked. In the three-unit Figure 1 window, three partitions are enabled and seven are "disabled."

Partitions specify how the units are assigned to the intervention start points. For our three-classroom example, suppose that it is decided that the intervention will begin just prior to either the second or third observation period for one of the classrooms, just prior to either the fifth or sixth observation period for another classroom, and just prior to the eighth observation period for the last classroom. In addition, suppose that Classroom 1 is randomly assigned to the 5/6 partition, Classroom 2 to the 8 partition, and Classroom 3 to the 2/3 partition. These specifications can be seen in Figure 1.

The sum of enabled partition numbers in the left column of Figure 1 indicates the total number of units in the design (N). Note that it is possible to have more than one unit assigned to a given partition, although with the regulated randomization design and corresponding calculations, different units assigned to the same partition cannot be assigned to the same intervention start point. Thus, in a design where, say, a 3/4/5 partition is designated for two of the units, the intervention for one of the units would begin at one of the three randomly selected start points (3, 4, or 5) and the intervention for the other unit would begin at a start point randomly selected from the two remaining ones.

If there are errors in the partition specifications, the "permutations in this design" area will show "- - -" and the "Compute" button will be disabled. Some other common errors are the following: (1) The list of potential start-point numbers cannot include a "1" because every participant must have at least one preintervention session; (2) the list of potential start-point numbers must contain numeric characters separated by spaces; and (3) the list of potential start-point numbers must have at least as many entries as there are units in that partition.

Number of randomization outcomes. The number of randomization outcomes associated with the specified design is based on the information provided in the enabled partitions. As indicated in Koehler and Levin's (1998) more general Formula 2, that number is equal to

$$N! \prod_{i=1}^P \binom{k_i}{n_i}$$

where N is the total number of units, P is the number of enabled partitions, k_i is the number of potential start points associated with the i th partition, and n_i is the number of (nonoverlapping) units associated with the i th partition

$$\sum_{i=1}^P n_i = N.$$

Creating Data Files

Before RegRand 1.0 can generate the appropriate randomization distribution, a data file must be created. This

Table 2
Complete Randomization Distribution and
Associated One-Tailed (Positive) Significance Probabilities
for the Hypothetical Data Randomization

Randomization Outcome	$M_B - M_A$	p value
[8,2,6]	0.5746	1.0000
[5,2,8]	0.6111	.9583
[6,2,8]	0.6444	.9167
[3,6,8]	0.8069	.8750
[8,2,5]	0.8274	.8333
[2,6,8]	0.8741	.7917
[8,3,6]	0.9079	.7500
[5,3,8]	0.9444	.7083
[3,5,8]	0.9653	.6667
[6,3,8]	0.9778	.6250
[2,5,8]	1.0324	.5833
[8,3,5]	1.1607	.5417
[3,8,6]	1.2006	.5000
[8,6,3]	1.2038	.4583
[2,8,6]	1.2677	.4167
[8,6,2]	1.3079	.3750
[8,5,3]	1.3621	.3333
[3,8,5]	1.4534	.2917
[8,5,2]	1.4663	.2500
[2,8,5]	1.5205	.2083
[5,8,3]	1.6339	.1667
[6,8,3]	1.6673	.1250
[5,8,2]	1.7381	.0833
[6,8,2]	1.7714	.0417*

Note.—*One-tailed significance probability corresponding to the outcome actually obtained.

file can be created either in word processors that permit saving as "text only" (ASCII files) or in spreadsheet or statistical programs that support saving as "tab-delimited data." Data files must be saved in this way for RegRand to recognize them. The following format is required for data files: There must be one line for each unit in the study; the *T*-measured outcomes for each unit must be on the same line, separated by either spaces or tabs; there must be the same number of measured outcomes for all units; and each line of the file (including the last) must end with a single return character.

Generating the Distribution

Once the design has been specified in the design window and there is a saved data file accessible from a directory or from a diskette, RegRand will generate the randomization distribution. Before clicking on "Compute," the user must be sure that the "Test Distribution" section of the window ("1 tailed positive," "1 tailed negative," or "2 tailed") is set appropriately. After clicking on "Compute," the user is asked to specify: (1) the data file to be used, and (2) a name and location for the to-be-created output file that will contain the randomization distribution. Although the resulting text-formatted output file is saved as a Microsoft Word document, it can be accessed by any application that is capable of opening text files. (The output file cannot be viewed within Version 1.0 of RegRand.)

Possible causes of errors. An error will occur while RegRand attempts to generate the randomization distribution if (1) the number of units specified in the design does not match the number of lines in the data file; (2) the number of measured outcomes is not the same in each line of the data file; (3) the number of observations per line in the data file is smaller than the largest number listed in the enabled partitions; (4) there are other errors associated with the data file (e.g., the file is open in another application, nonnumeric characters or an extra return character are present); or (5) the design results in more

than 100,000 randomization outcomes or if RegRand does not have enough memory allocated to it for a given design; in the latter situation, it is possible to increase the memory allocation using the "Get Info" menu option in the Finder.

Known Limitations

Version 1.0 of RegRand has a few known limitations: (1) The program is limited to 10 units and no more than 20 measured outcomes per unit. (2) Checking to make sure that files are available and that there is sufficient disk space for an output file does not always occur. In that case, a generic error message is provided. (3) As may be seen in Figure 1, there is a place holder for Monte Carlo sampling investigations of the Koehler-Levin (1998) regulated randomization procedure, but at present that particular feature is not operational.

REFERENCES

- EDGINGTON, E. S. (1975). Randomization tests for one-subject operant experiments. *Journal of Psychology*, *90*, 57-68.
- KAZDIN, A. E. (1992). *Research design in clinical psychology* (2nd ed.). Boston: Allyn & Bacon.
- KOEHLER, M. J., & LEVIN, J. R. (1998). Regulated randomization: A potentially sharper analytical tool for the multiple-baseline design. *Psychological Methods*, *3*, 206-217.
- KRATOCHWILL, T. R., & LEVIN, J. R. (Eds.) (1992). *Single-case research design and analysis: New developments for psychology and education*. Hillsdale, NJ: Erlbaum.
- LEVIN, J. R. (1992). Single-case research design and analysis: Comments and concerns. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New developments for psychology and education* (pp. 213-224). Hillsdale, NJ: Erlbaum.
- MARASCUILLO, L. A., & BUSK, P. L. (1988). Combining statistics for multiple-baseline AB and replicated ABAB designs across subjects. *Behavioral Assessment*, *10*, 1-28.
- McKIE, A. (1998). *Effectiveness of a neoprene hand splint on grasp in young children with cerebral palsy*. Unpublished master's thesis, University of Wisconsin, Madison.

(Manuscript received February 8, 1999;
revision accepted for publication December 8, 1999.)