DOCUMENT RESUME

ED 331 843 TM 016 271

AUTHOR	Janosky, Janine E.
TITLE	An Overview of the Analysis of a Single-Subject Design with Recommendations.
PUB DATE	Apr 91
NOTE	29p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 3-7, 1991).
PUB TYPE	Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Data Interpretation; Graphs; Literature Reviews; Observation; *Research Design; Research Reports; *Statistical Analysis
IDENTIFIERS	*Single Subject Designs; *Smoothing Methods

ABSTRACT

The statistical analysis of data from a single-subject design is somewhat controversial. The procedure most often chosen to examine data from a single-subject design involves the visual inspection of the graphed outcome variable over time. Problems associated with this procedure are discussed, and questions about the use of traditional statistical tests are reviewed. To supplement visual inspection of the data, the non-parametric smoother proposed by J. W. Tukey (1977) is presented as an appropriate and useful technique for interpreting such data. Two working examples are presented with single-subject data from: (1) baseline observations of the level of a particular hormone; and (2) observations of the behavior of a subject. Four graphs illustrate these working examples. A 46-item list of references is included. (SLD)

An Overview of the Analysis of a Single-Subject Design with Recommendations

Janine E. Janosky, Ph.D.
University of Pittsburgh
Department of Clinical Epidemiology and Preventive Medicine
M200 Scaife Hall
Pittsburgh, PA 15261

Running Head: Single-subject

AERA/NCME PRESENTATION, CHICAGO, APRIL 1991

Abstract

The statistical analysis of data from a single-subject design is somewhat controversial (Gay, 1987; Weiner & Eisen, 1985). The procedure most often chosen for examining data from a single-subject design involves the visual inspection of the graphed outcome variable over time (Gay, 1987). Difficulties of this procedure have been suggested by Ottenbacher (1986). The use of traditional statistical tests has also been questioned (Blumberg, 1984; Kazdin, 1984; Jones, Vaught, & Weinrott, 1977; Michael, 1974). As a complement to the visual inspection of the observed data, the nonparametric smoother proposed by Tukey (1977) is presented as a possibly appropriate and useful technique for the examination of data from a single-subject research paradigm. Two working examples are presented.

An Overview of the Analysis of a Single-Subject Design with Recommendations

A frequently used quasi-experimental design involves multiple measurements over time on a single subject (n=1). This design has been labeled intrasubject replication, within-subject design, repeated-measures design, time-series design, individual organism design, A-B, A-B-A design, or a single-subject design (Gay, 1987; McReynolds & Thompson, 1986; McLaughlin, 1983; Kazdin, 1982; Cook & Campbell, 1979; Hersen & Barlow, 1976; Kratochwill, et al., 1974; and Sidman, 1960). Most often, this design is used to study changes over time. The changes over time, among others, can be investigated as behavioral, developmental, seasonal, or circadian changes. Two examples of possible research paradigms are: (1) the study of a token reinforcement system in controlling head-banging behavior among severely retarded children and (2) the investigation of seasonal or circadian psychological changes over a specified period of time (Moore-Ede, Sulzman, & Fuller, 1982).

Some applications of the design involve a series of measurements (baseline), intervention measurements (treatment), and a series of measurements (baseline). This series can also repeat, involving additional interventions. This application of the design is most often applied in psychological, educational, or medical research where the series of measurements is referred tc as A, and the intervention is termed B (Kearns, 1986). The A-B, A-B-A, A-B-A-B or other versions of the design have been used to investigate topics such as behavioral modification, stuttering, psychotherapy, and drug research (Connell & Thompson, 1986; Borg & Gall, 1983).

> The data resulting from the best of experimental designs is of little value unless subsequent statistical analyses permit the investigator to test the extent to which obtained differences exceed chance fluctuations (Gottman, McFall, Barnett, 1969, p. 301).

Visual Inspection

The procedure most often suggested for the analysis of the data from a single-subject design involves a visual inspection of a graphic presentation of the results (Gay, 1987; Borg & Gall, 1983). This procedure involves plotting the data on a bivariate plot, time on the horizontal axis and outcome scores on the vertical axis. The assessment of the change or impact of the intervention is made by visually examining the plot. Parsonson and Baer (1978), Campbell (1988), and Kazdin (1982) state that visual inspection is a useful and appropriate procedure for determining the effects of an intervention. Dattilo and Nelson (1986) state that visual inspection permits the magnitude of change to be assessed by changes in mean performance over repeated measurements or by examining the performance level during shifts from baseline to intervention to baseline. Jones, Vaught, & Weinrott (1977) provide specific guidelines for visually analyzing the data from a single-subject design.

However, given the potential for the introduction of biases and inaccuracies in plotting, care should be taken when attempting to evaluate change in this manner (Tufte, 1983; Campbell, 1974). Distorted plots, broadened or limited axes, inappropriate scales, etc. can all potentially lead to misinterpretation. Ottenbacher (1986) investigated agreement among raters in evaluating visual plots. It was suggested that considerable disagreement occurred between therapists when changes were rated as significantly improved versus nonimproved by the use of visual inspection alone; the agreement was at the level of chance. The level of agreement did not improve when the comparison was judgment of improvement based upon visual analysis versus quantitative analysis (Ottenbacher, 1986). As cited in Kazdin (1984), additional studies have suggested that in a variety of circumstances, researchers have disagreement in their interpretation of data via visual inspection (DeProspero and Cohen, 1979; Gottman and Glass, 1978; Jones et al., 1978).

In summary, as a descriptive technique, visual inspection of the raw data may be useful; yet, care should be taken when visual inspection of the raw data is used as the sole means of interpretations.

Statistical Hypothesis Testing

Kazdin (1984) has suggested the use of statistical analyses in order to provide more informative procedures than visual inspection. More specifically, it is suggested that traditional t and F tests be employed. Kazdin (1984) suggests the use of the t-test for the analysis of A-B designs, and the use of the analysis of variance for the analysis of A-B-A-B designs (i.e., designs with more than two phases). It has been suggested that the use of these tests would enable the researcher to evaluate whether the differences between the phase means are statistically different.

Both the t-test and the analysis of variance are parametric statistical tests, and both of these tests have statistical assumptions. Kazdin (1984) addresses one of these assumptions and cautions the use of these tests when the data are serially dependent (e.g., adjacent observations are highly correlated). However, there are additional statistical assumptions to these tests that might not be met for analyzing the data from a single-subject design. As an example, for the t-test these assumptions include: (1) normality of observations within phase, (2) equality of variances between phases, and (3) independence between observations. In order to investigate whether the first two assumptions hold, a test for normality and a test for equality of variances can be used. The implemented research design determines whether the third assumption is met. In a single-subject design, all across time reported observations are for one subject. Therefore, the observations are most likely dependent. Given the probable violation of the independence of observations assumption, reservations should be observed when using these statistical tests. These arguments, cautioning the use of these tests, are also presented in Bock (1975) and Weiner and Eisen (1985).

The use of curve fitting has been recommended for the analysis of data from a single-subject design. For the A-B design one curve would be fit for the baseline (A) and one for the treatment (B). A test of statistical significance can then be used to test for slope and/or intercept differences (Mood, 1950). For the single-subject design, this approach has two potential difficulties: (1) the violation of the assumption of linearity or nonrandomly distributed residuals and (2) violation of the assumption that the repeated observations are independent samples of a random variable (Gottman, McFall, & Barnett, 1969; Holtzman, 1967). Given the possible violation of the assumptions, the use of curve fitting should be cautioned.

The use of traditional statistical tests in single-subject research is a somewhat controversial issue (Gay, 1987; Weiner & Eisen, 1985). Since the assumptions of the traditional statistical tests are probably violated when used with a single-subject design, the use of these tests might not be entirely appropriate.

Another proposed technique for the analysis of single-subject designs is time-series analysis (Dattilo & Nelson, 1986; Tryon, 1984; Chatfield, 1982; Tryon, 1982; Jones, Vaught, and Weinrott, 1977; Glass, Willson, Gottman, 1975). The purpose of time-series analysis is to detect reliable changes in slope and level. Especially proposed as useful, for the single-subject design, is the autoregressive integrated average (ARIMA) technique. Unlike the t-test and curve-fitting, with the time-series analysis there is an explicit allowance for statistical dependence among the observations at different points in time. However, a potential limitation of the analysis of a single-subject design using time-series involves the number of recommended data points. The number of recommended data points per phase is 50 to 100 (Borg and Gall, 1983; Hartmann et al., 1980); thus practical use of time-series for the single-subject design is most likely limited.

As a simplified time-series analysis technique, the C statistic has been proposed as an alternative to the ARIMA technique (Tryon, 1984; Tryon, 1982). However, as discussed by Blumberg (1984), there are many difficulties with the C statistic. Among other criticisms, the critical value does not change appreciably, irrespective of the number of observations.

Thus the statistical significance is directly related to the number of observations (Tryon, 1982).

Other descriptive and inferential, parametric and nonparametric, techniques have been proposed. Among other techniques, Edgington has proposed the use of randomization tests (Edgington, 1980), and the split-middle technique has been proposed as a descriptive and an inferential technique (White, 1971, 1972, 1974). General applicability of the randomization technique has been limited by the difficulty of required computations (Kazdin, 1984; Conover, 1971). The use of the split-middle technique, unlike time-series, seems to have been limited by the required number of data points per phase (White, 1974).

Nonparametric Smoothing

Exploratory data analysis (EDA) is a set of techniques proposed by Tukey (1977). One of the EDA proposed techniques is sequential smoothers. It has been suggested that nonparametric smoothing can be used to more fully understand a process over time (Tryon, 1983; Velleman and Hoaglin, 1981) such as data from a single-subject research design.

The smoothing technique involves specifying a relatively smooth curve from a series of time-bound points. For each step a comparison is made with some of the adjacent points. The fitting process is analogous to a regression equation consisting of two additive components, smooth plus rough (Data=smooth plus rough) (Tukey, 1977, p. 208). In terms of regression analysis, the smooth can be conceptualized as the regression term or fit of the model, and the rough term as the residual term. Smoothing the observations results in reducing the random variation in the measurements. This difference is most likely due to the excess "noise" or "roughness" in these observations. One conceptualization of the noise in the data is as measurement error due to less than perfect reliability in the measurement instrument. Since the smoothing procedure is nonparametric, assumptions which restrict the use of the traditional t and F tests for the application in the single-subject design are not necessary. The recommended number of data points per phase is 6 (Velleman and Hoaglin, 1981), thus allowing a much wider practical applicability than time-series.

Nonparametric Smoothing: Working Examples

Figure 1 contains the observed data from a single-subject design. The purpose of the study was to categorize the baseline (basal) observations of the level of a particular hormone (i.e. Did the hormone level exhibit a particular pattern?). Measurements were taken over 120 time points, that is, each measurement was taken every hour for a total of 5 days. By examining the observed data, a prominent pattern is not immediately evident; there appears to be a large amount of scatter.

These observed data were analyzed using a nonparametric smoother. Figure 2 contains the observed and smoothed data from this study. By examining the smoothed data, a cyclical pattern or a circadian rhythm seems to be evident. Approximately every 15 to 20 hours a low (around 5.0) is reached. This pattern was not as strongly evident in the observed data.

Insert Figures 1 and 2 About Here

The smoothed data, presented in Figure 2, were smoothed using the RSMOOTH command in Minitab (Ryan, Joiner, and Ryan, 1985). RSMOOTH uses a moving median smoother 4253HR to smooth the data (Velleman & Hoaglin, 1981). The 4253HR used in Minitab is a compound smoother. Compound smoothers combine several elementary smoothers by resmoothing and reroughing (Velleman & Hoaglin, 1981). The 4253HR smoother starts with a running median of 4 adjacent observations, recentered by 2. It then resmooths by 5 observations, by 3, and then applying the hanning smoother. The hanning smoother (H) consists of multiplying by weights in each averaging operation. These resultant smoothed data are reroughed (e.g., residual are calculated), and then the entire process is repeated (R) (Velleman & Hoaglin, 1981). According to Velleman & Hoaglin (1981) and Velleman (1980), the 4253HR compound smoother seems to perform quite well in general application. Additional discussions of nonparametric smoothers can be found in Velleman & Hoaglin (1981), Velleman (1980), and Tukey (1977). In addition to Minitab, nonparametric smoothers are also available in other statistical analysis packages (SPSS, 1980).

Figure 3 contains the observed data from a single-subject A-B-A study. The number of occurrences of a specified behavior were recorded. Twenty measurements were made within each phase. The means and standard deviations within each phase were: (A) mean=6.92, SD=1.90; (B) mean 6.95, SD=1.11; and (A) mean=7.01, SD=2.00. Based upon these calculated summary values it appears as if no differences were observed for either the means or the standard deviations, across phases. In addition, by visual inspection of the data presented in Figure 3, it seems as if no differences were observed between the phases.

Figure 4 contains the observed and smoothed data from this study. The data were smoothed within phase. The aforementioned smoothing technique was employed. A somewhat different conclusion is reaching by examining the smoothed data when compared with the conclusions reached from the observed data. The smoothing of the data has reduced the variability. Initially during the B phase the number of occurrences drops from the previous phase. This is followed by a steady increase to the previous phase A level, and followed by a steady increase. Then this level is initially maintained during the final A phase, with a steady stair-like decrease. By examining the smoothed data, it appears as if the number of occurrences drops initially during the B phase with a broadened climb to the previous level, followed by a plateau with a decrease. This interpretation does not seem as evident by examining either a plot of the observed data or the summarized descriptive statistics.

Insert Figures 3 and 4 About Here

Conclusion

In addition to the visual display of the observed data, the nonparametric smoothing technique may be useful in aiding in the interpretation of the data from a single-subject design. Smoothers are available in most statistical analysis packages. By eliminating the rough or residual from the observed data, while at the same time not inappropriately applying a statistical test, the process or observations can be more easily interpreted. One often overlooked advantage of a single-subject design is the continuous record of fluctuations over time; however, the smoother allows the separation of systematic fluctuations from the random fluctuations (Gottman, McFall, & Barnett, 1969). Smoothing might provide an informative understanding of the process or behavior outcomes over time while not violating any applicable statistical assumptions.

References

Blumberg, C.J. (1984). Comments on "A simplified time-series analysis for evaluating treatment interventions". *Journal of Applied Behavior Analysis*, *17(4)*, 539-542.

Bock, R.D. (1975). *Multivariate statistical methods in behavioral research*. New York: McGraw-Hill.

Borg, W.R. & Gall, M.D. (1983). *Educational research: An introduction*. New York: Longman.

Campbell, P.H. (1988). Using a single-subject research design to evaluate the effectiveness of treatment. *The American Journal of Occupational Therapy*, *42(11)*, 732-736.

Campbell, S.K. (1974). *Flaws and fallacies in statistical thinking*. New Jersey: Prentice-Hall, Inc.

Chatfield, C. (1982). *The analysis of time series: An introduction*. New York: Chapman and Hall.

Connell, P.J. & Thompson, C.K. (1986). Flexibility of single-subject experimental designs. Part III: Using flexibility to design or modify experiments. *Journal of Speech and Hearing Disorders*, *51*, 194-203.

Conover, W.J. (1971). *Practical nonparametric statistics*. New York: Wiley.

Cook, T.D., & Campbell, D.T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Boston: Houghton Mifflin Company.

Dattilo, J. & Nelson, G.D. (1986). Single-subject evaluation in health education. Health Education Quarterly, 13(3), 249-259.

DeProspero, A. & Cohen, S. (1979). Inconsistent visual analysis of intrasubject data. Journal of Applied Behavior Analysis, 12, 573-579.

Edgington, E.S. (1980). Validity of randomization tests for one-subject experiments. Journal of Educational Statistics, 5, 235-251.

Gay, L.R. (1987). Educational research: Competencies for analysis and application. Columbus: Merrill Publishing Company.

Glass, G., Willson, V., & Gottman, J. (1975). Designs and analysis of time-series experiments. Boulder, CO: Associated University Press.

Gottman, J.M. & Glass, G.V. (1978). Analysis of interrupted time-series experiments. In T.R. Kratochwill (Ed.), Single-subject research: Strategies for evaluating change (pp. 197-237) New York: Academic Press.

Gottman, J.M., McFall, R.M., & Barnett, J.T. (1969). Design and analysis of research using time series. Psychological Bulletin, 72(4), 299-306.

Hartmann, D.P., Gottman, J.M., Jones, R.R., Gardner, W., Kazdin, A.E., and Vaught, R. (1980). Interrupted time-series analysis and its application to behavioral data. Journal of Applied Behavior Analysis, 13, 543-559.

Hersen, M. & Barlow, D.H. (1976). Single case experimental designs: Strategies for studying behavioral change. New York: Pergamon.

Holtzman, W.H. (1967). Statistical models for the study of change in the single case. In C. Harris (Ed.) Problems in measuring change. Madison: University of Wisconsin Press.

Jones, R.R., Vaught, R.S. & Weinrott, M. (1977). Time-series analysis in operant research. Journal of Applied Behavior Analysis, 10, 151-166.

Jones, R.R., Weinrott, M.R., & Vaught, R.S. (1978). Effects of serial-dependency on the agreement between visual and statistical inference. Journal of Applied Behavior Analysis, 11, 277-283.

Kazdin, A.E. (1984). Statistical analyses for single-case experimental designs. In D.H. Barlow & M. Hersen. Single-case experimental designs: Strategies for studying behavior change (2nd ed., pp. 285-324). New York: Pergamon Press.

Kazdin, A.E. (1982). Single case research designs: Methods for clinical and applied settings. New York: Oxford.

Kearns, K.P. (1986). Flexibility of single-subject experimental designs. Part II: Design selection and arrangement of experimental phases. Journal of Speech and Hearing Disorders, 51, 204-214.

Kratochwill, T., Alden, K., Demuth, D., Dawson, D., Panicucci, C., Arnston, P., McMurry, N. Hempstead, J., & Levin, J. (1974). A further consideration in the application of an analysis-of-variance model for the intrasubject design. Journal of Applied Behavior Analysis, 7, 629-634.

McLaughlin, T.F. (1983). An examination and evaluation of single subject designs used in behavior analysis research in school settings. Educational Research Quarterly, 7(4), 35-42.

McReynolds. L.V. & Thompson, C.K. (1986). Flexibility of single-subject experimental designs. Part 1: Review of the basics of single-subject designs. Journal of Speech and Hearing Disorders, 51, 194-203.

Michael, J. (1974). Statistical inference for individual organism research: Mixed blessing or curse? Journal of Applied Behavior Analysis, 7, 647-653.

Mood, A.F. (1950). Introduction to the theory of statistics. New York: McGraw-Hill.

Moore-Ede, M.C., Sulzman, F.M., & Fuller, C.A. (1982). The clocks that time us. MA: Harvard University Press.

Ottenbacher, K.J. (1986). Reliability and accuracy of visually analyzing graphed data from single-subject designs. The American Journal of Occupational Therapy, 40(7), 464-469.

Parsonson, B.S. & Baer, D.M. (1978). The analysis and presentation of graphic data. TR Kratochwill (Ed.) Single subject research: Strategies for evaluating change. New York: Academic.

Ryan, B.F., Joiner, B.L., & Ryan, Jr., T.A. (1985). Minitab Boston: PWS-KENT Publishing Company.

Sidman, M. (1960). Tactics of scientific research. New York: Basic Books.

Statistical Packages for the Social Sciences (1988). Chicago, Illinois.

Tufte, E.R. (1983). The visual display of quantitative information. CT: Graphics Press.

Tukey, J.W. (1977). Exploratory Data Analysis. Reading, MA: Addison-Wesley Publishing Company.

Tryon, W.W. (1984). A simplified time-series analysis for evaluating treatment interventions: A rejoiner to Blumberg. Journal of Applied Behavior Analysis, 17, 543-544.

Tryon, W.W. (1983). Digital filters in behavioral research. Journal of the Experimental Analysis of Behavior, 39, 185-190.

Tryon, W.W. (1982). A simplified time-series for evaluating treatment interventions. Journal of Applied Behavior Analysis, 15, 423-429.

Velleman, R.F. (1980). Definition and comparison of robust nonlinear data smoothing algorithms. Journal of the American Statistical Association, 75, 609-615.

Velleman, R.F. & Hoaglin, D.C. (1981). Applications, basics, and computing of exploratory data analysis. Duxbury Press.

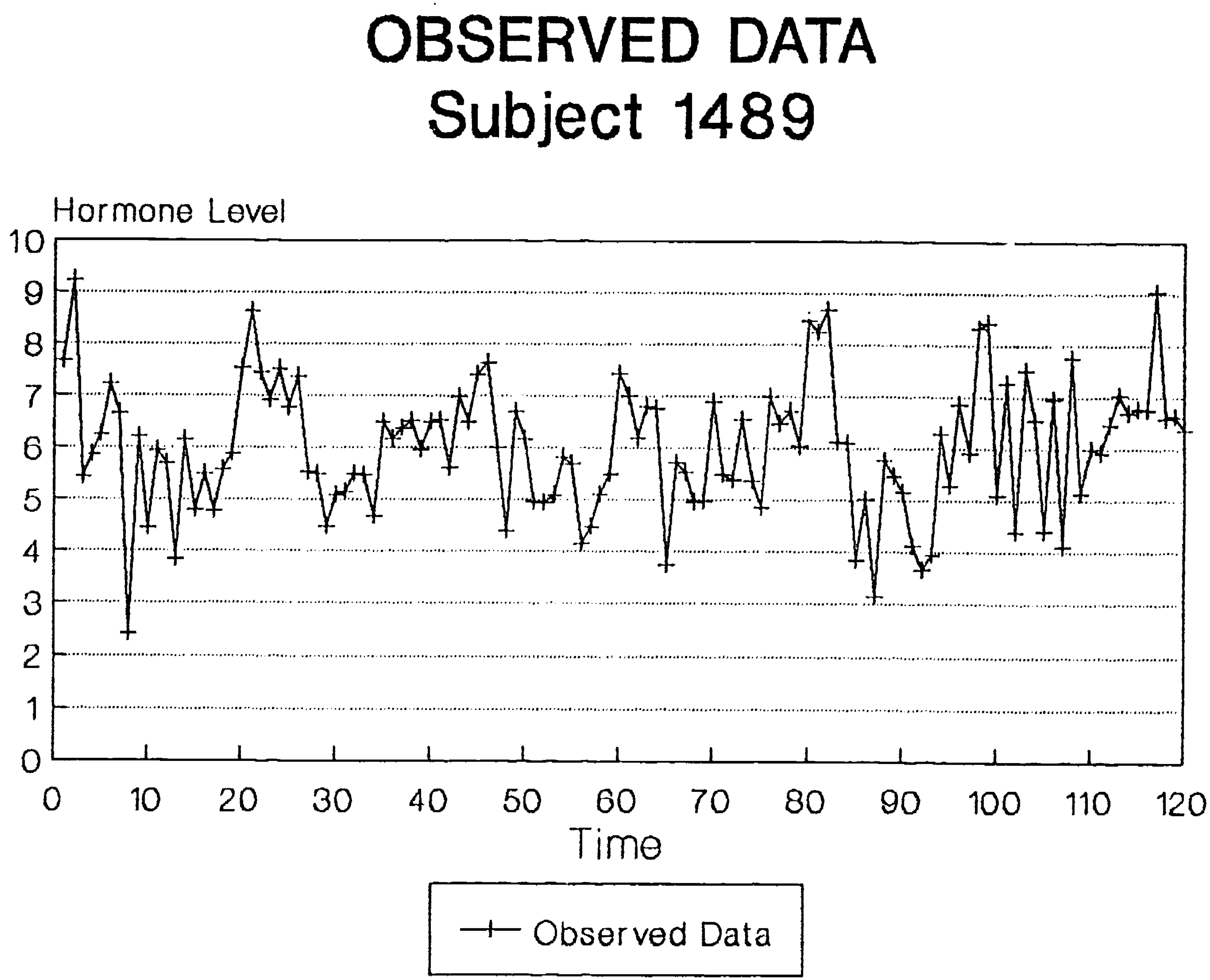
Weiner, I.S. & Eisen, R.G. (1985). Clinical research: The case study and single-subject designs. Journal of Allied Health, 191-201.

White, O.R. (1971). A glossary of behavioral terminology. Champaigm, IL: Research Press.

White, O.R. (1972). A manual for the calculation and use of the median slope: A technique of progress estimation and prediction in the single case. Eugene, OR: University of Oregon, Regional Resource Center for Handicapped Children.

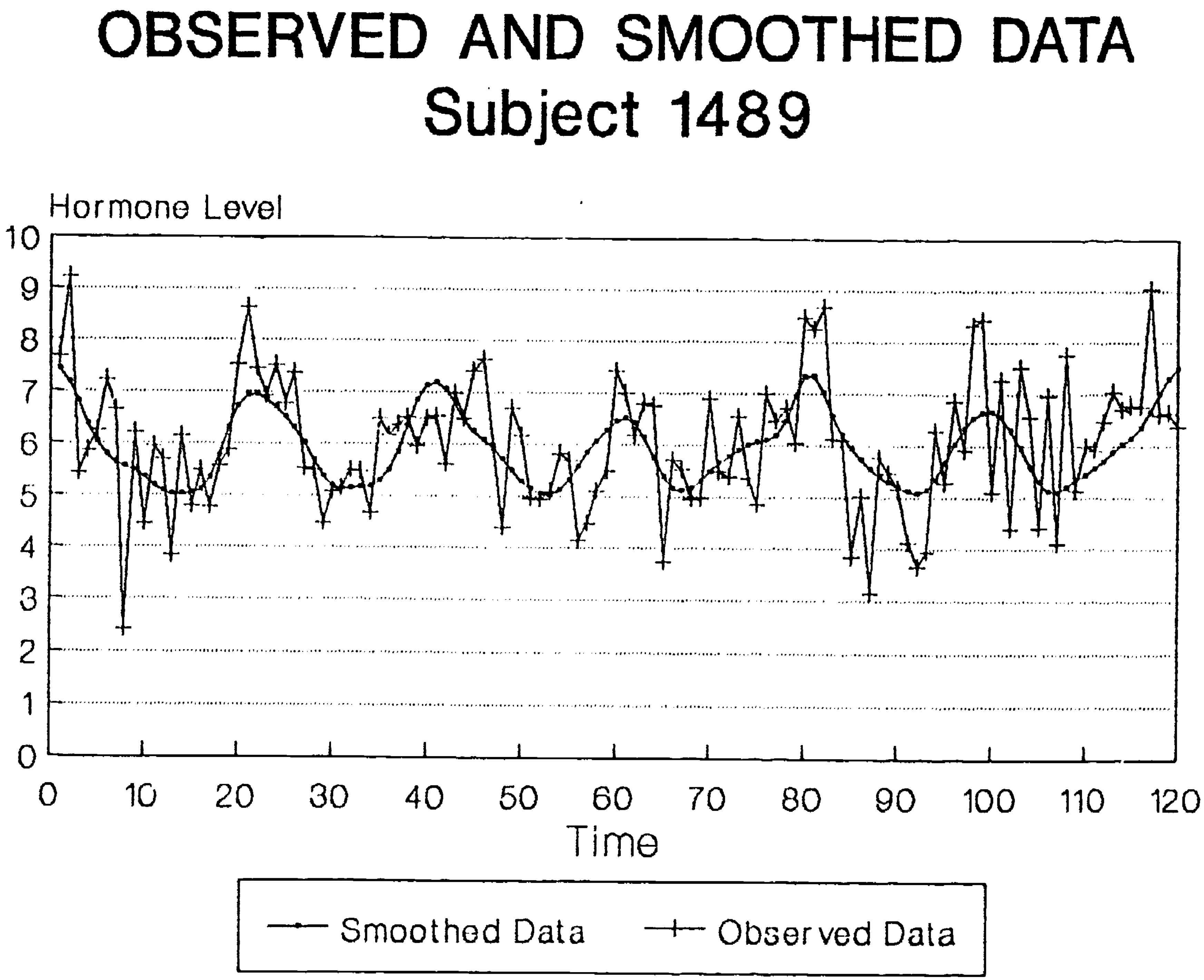
White, O.R. (1974). The "split-middle". A "quickie" method of trend estimation. Seattle, WA: University of Washington, Experimental Education Unit, Child Development and Mental Retardation Center.

Figure 1



OBSERVED DATA
Subject 1489
Hormone Level
10
9
8
7
6
5
4
3
2
1
0
0
10
20
30
40
50
60
70
80
90
100
110
120
Time
Observed Data

Figure 2



OBSERVED AND SMOOTHED DATA
Subject 1489
Hormone Level
10
9
8
7
6
5
4
3
2
1
0
0
10
20
30
40
50
60
70
80
90
100
110
120
Time
Smoothed Data
Observed Data

Figure 3

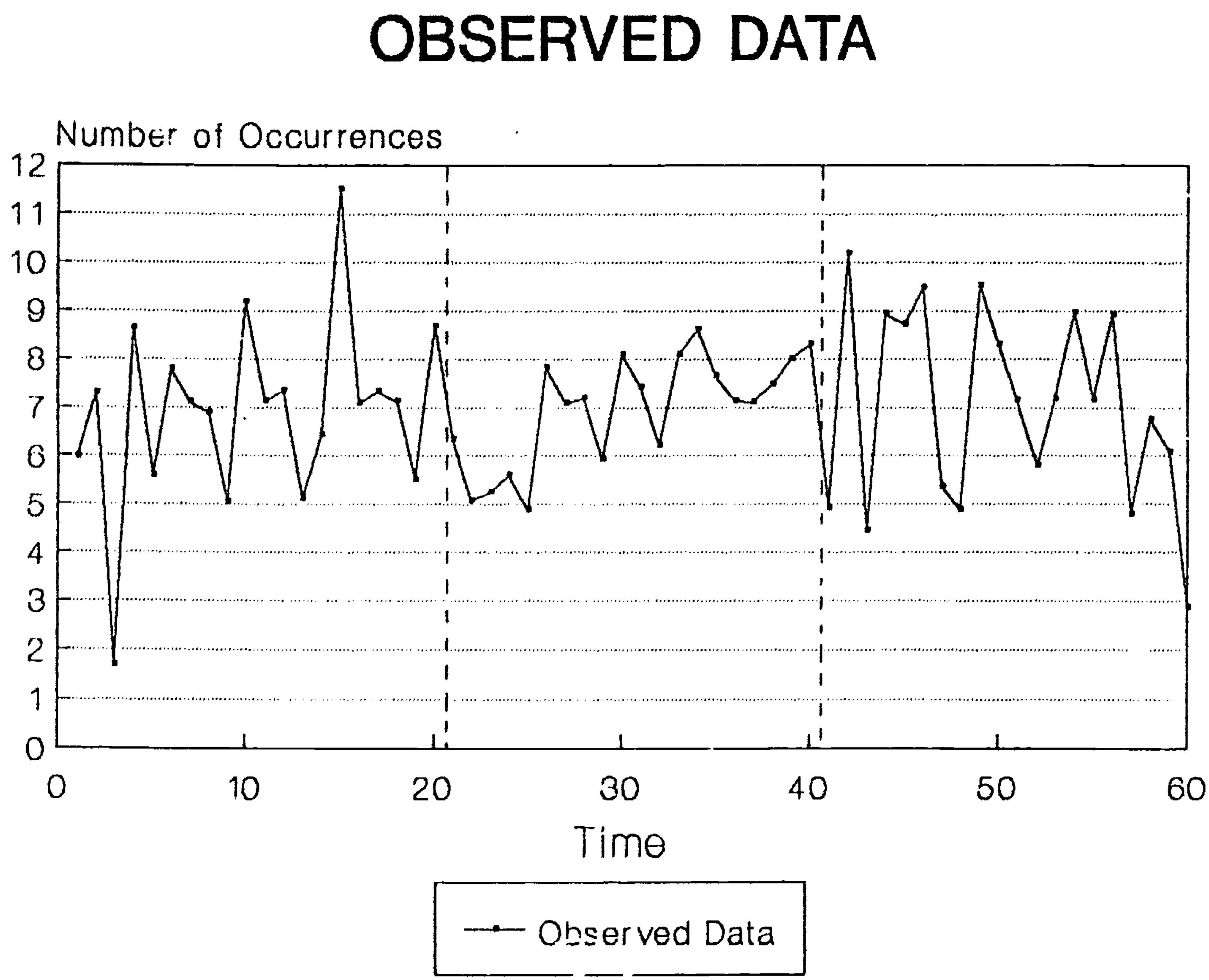
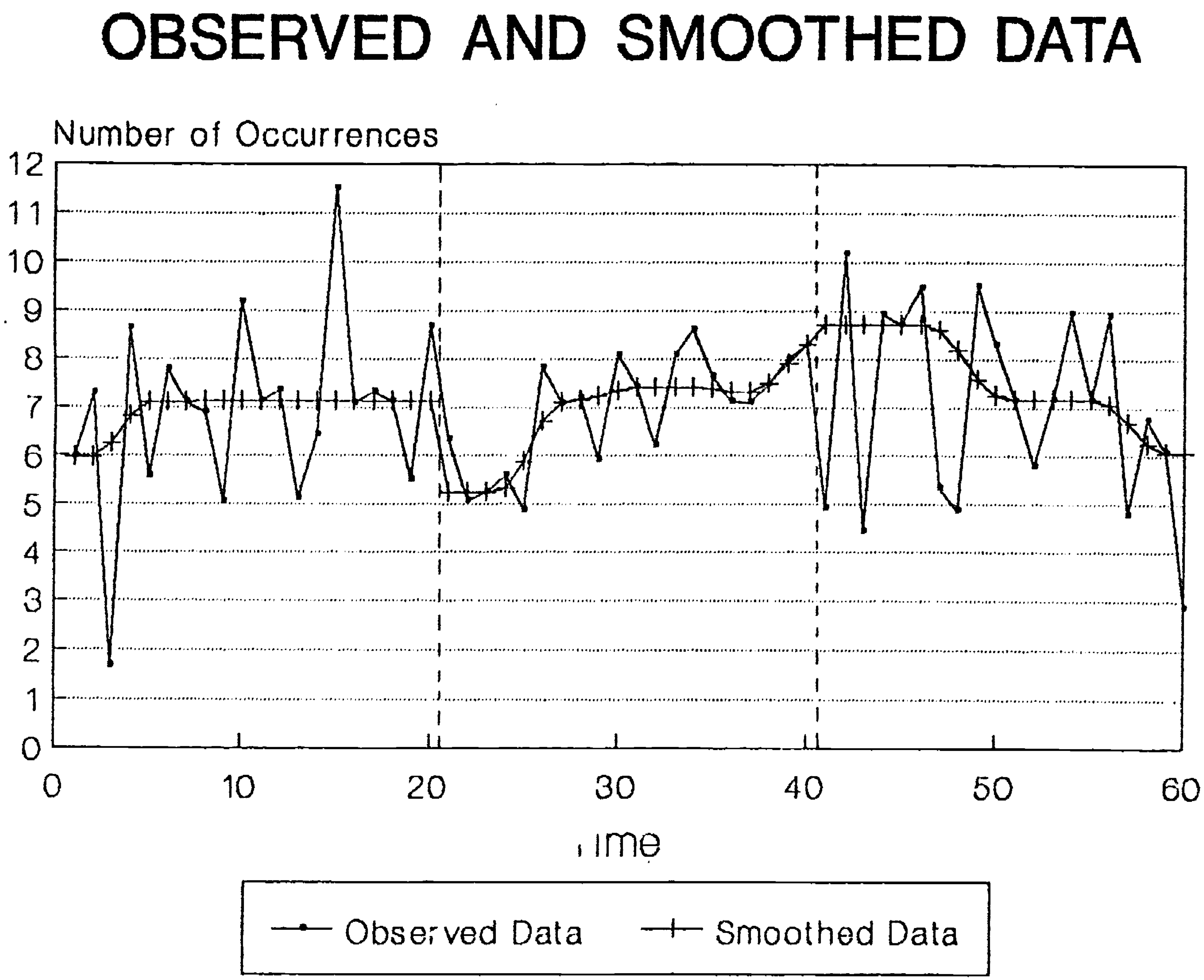

OBSERVED DATA
Number of Occurrences
12
11
10
9
8
7
6
5
4
3
2
1
0
0
10
20
30
40
50
60
Time
Observed Data

Figure 4


OBSERVED AND SMOOTHED DATA
Number of Occurrences
12
11
10
9
8
7
6
5
4
3
2
1
0
0
10
20
30
40
50
60
ime
Observed Data
Smoothed Data