TEMPORAL AND CLIMATIC VARIABLES IN NATURALISTIC OBSERVATION¹

MARK B. RUSSELL AND MARTHA E. BERNAL

UNIVERSITY OF DENVER

Home-observation data on 5- to 7-yr-old boys collected over 2 yr were examined for systematic variations in rates of desirable and undesirable behaviors associated with several temporal and climatic variables. Significant effects associated with time of day, day of the week, precipitation, and temperature were found. No significant effects on the naturalistic observation data were found for environmental factors associated with lunar phase. It was noted that the correlational nature of the findings did not obviate the necessity for control of the influence of temporal and climatic variables. Several methodological strategies for such control were discussed.

DESCRIPTORS: environmental influences, control, observational data, methodology, systematic error, recording and measurement, children

Variance in observed behavior, except that which results from the source manipulated by the investigator, comprises the noise or error term. Thus, error includes the accumulated effect of all nontargeted influences that serve randomly or systematically to contribute to measurement error in observed scores. Despite attempts to minimize or eliminate possible error sources via sophisticated designs and statistical procedures, signal and noise are necessarily mingled in every set of observations (Diamond, 1959). As the magnitude or directionality of the error increases, results become more obscured. The present report examined potential error sources, namely day of the week, time of day, precipitation, temperature, and lunar phase, that are accessible to manipulation by the care-

ful investigator. The hypothesis here was that the specified temporal and climatic variables would be associated with systematic fluctuations in the magnitude of behavioral scores. This information should be of particular interest to clinical intervention researchers who rely on naturalistic observation measures to evaluate treatment outcome.

METHOD

From December 7, 1971 to May 30, 1972, and from December 10, 1972 to April 16, 1973, naturalistic observation data were collected in the homes of 5- to 7-yr-old boys. The parents of 38 boys reported discipline problems at home; parents of 38 boys described them as no more noncompliant, disruptive, or aggressive than the average boy. These 76 subjects were recruited by mail and telephone through a large urban public-school district. The deviant subjects' families were to be offered a behavioral parent-training intervention and the normal subjects' families were asked to volunteer so as to permit collection of normative data. All families were paid from \$10 to \$12 for taking part in the observations. From four to six 30-min observations were conducted in their homes with all family

¹This work was supported by NIMH Grant Number MH20922-04 awarded to the second author. Requests for scoring manuals describing the behavior code in greater detail or for reprints should be addressed to Martha E. Bernal, Department of Psychology, University of Denver, 2327 East Evans, Denver, Colorado 80210. Thanks are due to Paul M. Rosen for his suggestions regarding the analyses, to Cheryl L. Russell for collecting the weather information, to Terrie Wasson for typing the manuscript, and to Juel Ann North, Susan Kreutzer, and Leola Schultz for training the observers and collecting and processing the data.

members present. These observations occurred before treatment for the deviant boys. Families were asked to remain in the observation space, refrain from turning on the television, stereo, or radio and from making telephone calls, and to allow only the household members to be present. Observations were scheduled twice a week at the same time of day at the family's convenience. Data from both the deviant and normal subjects were cast into a common pool of 300 observations. Separate analyses for the deviant and normal children were not performed, as the purpose of the present report was to demonstrate systematic sources of uncontrolled variability, not to describe subject type differences.

The Bernal and North (1972) coding system was used; details of observer training, coding, *etc.*, may be found in Bernal, Delfini, North, and Kreutzer (1976). Well-trained professional observers coded three summary categories for the targetted child: Annoying, Deviant, and Desirable behaviors. The Annoying category consisted of verbal abuse, noisemaking, whining, and motor behavior. The Deviant category included name calling, destructiveness, and physical negative. The Desirable category contained on task, physical positive, appropriate verbal interaction, appropriate nonverbal interaction, and altruism. Labels for the three summary categories were used for convenience and did not necessarily reflect the connotations of their generic derivatives; rather, they reflected behavior categories differentiated on an *a priori* basis in terms of the presumed degree of acceptability or severity of the behaviors subsumed under them. Only the first occurrence of a behavior in each of the categories within each 15-sec time sample was coded. Therefore, the 12 behaviors subsumed under the three categories could not be grouped together. Interobserver agreement was calculated for each category by means of the formula, $[(\text{Agreement on occurrences} + \text{Agreement on nonoccurrences}) / \text{Number of time samples per observation}] \times 100$. Occurrence was defined as scoring of any behavior within a given category, *i.e.*, agreement on the specific behavior

scored within the category was not required. Category agreement scores averaged 0.94 for Desirable, 0.93 for Annoying, and 0.99 for Deviant. These agreement figures were obtained once weekly and averaged over 42 checks.

There were five variables. *Lunar phase* was defined as full, new, or interphase. A full moon and a new moon were defined as the date of occurrence $\pm$ one day. All other days were interphase. *Day of the week* was the second variable. *Time of day* was defined as the time at which the observation began, divided by hour from morning to night. The restricted range of times of day here examined, 3:30 to 8:00 p.m., resulted from the practical issue that home observations had to be scheduled during the time between the finish of school and bedtime. *Precipitation* was defined as water equivalent of ≤ 0.01 in. occurring between 7:00 a.m. and 8:00 p.m. on the day of the observation. *Temperature* was recorded as the highest temperature reached for the day, divided into four ranges: ≤ 35 degrees, 36 degrees to 50 degrees, 51 to 65 degrees, and ≥ 66 degrees.

All 300 observations were cast into a matrix with observations as rows and variables as columns. That is, each observation was identified in terms of its relevance to each variable, *e.g.*, observation 1 was at interphase, on Monday, at 4:30, there was precipitation, and the temperature was in the ≤ 35 -degree range. Then, the observations were distributed into the appropriate lists for the various analyses, *e.g.*, lists of observations that had been conducted on Mondays, on Tuesdays, *etc.*, were compiled for the analyses for the day of the week and lists of observations that had occurred on days when the temperature was less than 35 degrees, between 36 degrees and 50 degrees, *etc.*, were made for the analyses for temperature. From these lists, samples of equal size were randomly drawn using the random-number table and procedures outlined by Wallis and Roberts (1958). Chi-square tests of independence revealed that the distribution of subject types, *i.e.*, deviant and normal, across variable

Table 1
Time Sample Rate per Minute for Behavior Categories in the Various Analyses

Variable	Annoying		Deviant		Desirable	
	M	SD	M	SD	M	SD
LUNAR PHASE						
Full (N = 20)	0.39	0.41	0.01	0.02	3.72	0.31
New (N = 20)	0.68	0.65	0.03	0.05	3.64	0.36
Interphase (N = 20)	0.47	0.39	0.03	0.06	3.71	0.24
DAY OF THE WEEK						
Monday (N = 40)	0.42	0.41	0.01	0.03	3.76	0.22
Tuesday (N = 40)	0.58	0.53	0.02	0.03	3.58	0.45
Wednesday (N = 40)	0.71	0.66	0.09	0.25	3.44	0.60
Thursday (N = 40)	0.65	0.70	0.03	0.06	3.61	0.47
Friday (N = 40)	0.75	0.64	0.08	0.33	3.55	0.57
TIME OF DAY						
3:30-5:00 (N = 60)	0.70	0.73	0.09	0.31	3.46	0.71
5:01-6:30 (N = 60)	0.64	0.66	0.03	0.09	3.58	0.42
6:31-8:00 (N = 60)	0.47	0.48	0.02	0.03	3.71	0.33
PRECIPITATION						
Precipitation (N = 60)	0.50	0.55	0.02	0.04	3.64	0.43
No Precipitation (N = 60)	0.64	0.57	0.08	0.29	3.57	0.50
TEMPERATURE						
≤35° (N = 30)	0.62	0.62	0.06	0.15	3.47	0.63
36° to 50° (N = 30)	0.68	0.70	0.04	0.05	3.59	0.50
51° to 65° (N = 30)	0.55	0.61	0.05	0.22	3.62	0.50
≥66° (N = 30)	0.53	0.41	0.02	0.04	3.66	0.30

categories, *e.g.*, precipitation *versus* no precipitation, did not deviate significantly from chance for any of the five variables. Thus, any obtained behavior-rate differences across variable categories were not artifactual to differing proportions of subject types in these categories. As the number of observations did not distribute evenly across lists within each analysis, *e.g.*, there was a preponderance of interphase days in the lunar phase analysis and relatively few observations on days with a temperature less than 35 degrees in the temperature analysis, different sample sizes resulted for the various analyses.

Time sample rate-per-minute data are presented in Table 1; tests of the distributions of the means are presented in Table 2. In analyses involving comparison of two events, *e.g.*, precipitation *versus* no precipitation, *t* tests were performed for each of the behavioral categories. In analyses involving comparison of more than two events, *e.g.*, days of week, simple

one-way ANOVA's were calculated for each category. Additionally, tests of linear trend were made for the temporal variables, time of day and day of the week. These variables were examined to see if behavioral category scores increased or decreased in a significantly linear fashion as the week progressed or as the day grew later.

Lunar Phase

Inspection of Table 2 reveals no significant effects associated with lunar phase for any of the child-behavior categories.

Day of the Week

The difference in mean Deviant behaviors across days of the week was statistically significant (Table 2). As demonstrated in Figure 1, subjects presented more Annoying and Deviant behaviors as the week progressed. Analysis of linear regression for Annoying means revealed a significant linear trend ($p < 0.05$)

Fig. 1. Time sample rate per minute as a function of day of the week.

with a nonsignificant residual: $F_{reg} + 5.91$ ($df = 1,195$), $F_{res} = 0.61$. Also, the linear regression ratio and its residual were significant for Deviant means: $F_{reg} = 8.42$ ($p < 0.01$), $F_{res} + 4.11$ ($p < 0.05$).

Time of Day

The distribution of the Desirable category means was significant (Table 2). Results are presented graphically in Figure 2. Analyses of linear regression for the Annoying, Deviant, and Desirable categories yielded F ratios of 3.87, 4.90, 7.21 ($df = 1.177$), with residuals of 0.12, 1.00, and 0.30 respectively. Regression ratios for Deviant and Desirable were significant ($p < 0.05$ and $p < 0.01$); the F_{reg} for Annoying just missed significance; all of the residuals were nonsignificant.

Precipitation

The difference between means for days with precipitation and days without precipitation for the Deviant category was significant (Table 2), with more Deviant behaviors present on days without precipitation.

Temperature

The F ratio for the Deviant category again reached significance in the temperature analyses (Table 2), with more Deviant behaviors associated with colder days (Table 1).

DISCUSSION

Naturalistic observation data collected over 2 yr were analyzed for systematic effects associated with various temporal and climatic conditions. Several significant effects were found. The children of this sample became increas-

Table 2
Tests^a of Distributions of Means

Variable	df	Annoying	Deviant	Desirable
Lunar Phase	2,57	$F = 1.76$	$F = 0.50$	$F = 0.33$
Day of the Week	4,195	$F = 1.91$	$F = 5.33^{**}$	$F = 2.26$
Time of Day	2,177	$F = 1.97$	$F = 2.66$	$F = 3.73^*$
Precipitation	118	$t = 1.00$	$t = 2.00^*$	$t = 0.10$
Temperature	3,116	$F = 0.41$	$F = 3.50^*$	$F = 0.80$

^a t tests used in two-group comparisons; simple one-way ANOVA used in comparisons of three or more groups.
* $p < 0.05$
** $p < 0.01$

Fig. 2. Time sample rate per minute as a function of time of day.

ingly annoying and deviant as the week progressed, exhibited more desirable and fewer deviant behaviors as bedtime approached, presented fewer deviant behaviors on rainy or snowy days, and were more deviant on cold days. Only the analyses for lunar phase failed to demonstrate a systematic source of error for any of the three summary child-behavior categories. Of the 15 analyses for main effect, four reached significance. Of the 11 nonsignificant analyses for main effect, two were demonstrated to be significantly linear functions. The num-

ber of significant effects was well above the experiment-wise error prediction rate, and was seen as strong support for the hypothesis of systematic variability in the data associated with temporal and climatic variables.

A lack of effect associated with lunar phase was also reported by Chapman (1960) in his sample of functional psychotics, schizophrenics, and paretics. A temperature effect was demonstrated in the self-reports of Faust and Ladewig's (1972) drug-addicted and depressive patients. Still, generalization of the present findings to populations with different subject characteristics, *e.g.*, sex or age, may not be fully justified. Replication with other samples is encouraged, as it is requisite to the question of generalizability.

Given the correlational nature of the present findings, caution should be exercised in constructing causal statements. It was not the purpose of the present study to demonstrate that time of day or day of the week *per se* caused differential behavior rates. A more plausible interpretation of the reported findings might be that some third variable or variables associated with the temporal and climatic conditions was responsible for the observed systematic behavioral change. These causal variables might include, either singly or in combination, the approach of bedtime, anticipation of the weekend, the lack of temptation to play outside on snowy or rainy days, favorite television programs, and so on. The list of possible speculations appears endless. Despite the lack of causal evidence, it should be noted that whatever variables were responsible for the reported results, these variables were associated with temporal and climatic factors. Thus, experimental control of the identified temporal and climatic factors would serve to reduce the confounding influence of the true causal events. Although the nature of this control is necessarily indirect, it may be that the identified temporal and climatic factors are more readily observable and measurable than their as-yet-unidentified correlate components. Control of

these variables may, therefore, be a more pragmatic and expedient approach to minimizing systematic uncontrolled variability in observational data.

Of perhaps greater practical utility than the direction of the reported effects to investigators of child behaviors was the presentation of several components of the error term that do not fulfill the assumption of random distribution and, as such, can contribute to systematic uncontrolled variability in the data. In essence, the results reaffirmed the necessity to institute procedures to control as many influences as possible other than the treatment of interest, and provided empirical evidence for possible confounds associated with certain environmental events. To exemplify an extreme hypothetical situation, the investigator using the present subject pool who collected pretreatment measures on Fridays at 3:30 p.m. on clear, cold days and collected followup data on Mondays at 6:30 p.m. on warm, rainy days might see a significant reduction in problem behaviors regardless of intervention technique. The more careful researcher should be aware of the systematic effects associated with temporal and climatic variables, and should institute procedures for randomization or counterbalancing of their influence across experimental conditions. Minimization of systematic sources of uncontrolled variability via randomly determined observation periods should be of particular relevance to clinical investigators using single-subject designs, since the variability of the data is itself used as a dependent measure in single-subject analyses.

Of the two classes of environmental variables examined here, control of temporal factors appears to be more salient, because it is obviously difficult to predict the weather on an *a priori* basis and because most clinical intervention research that uses observational data does not involve months or years of data collection. The careful methodologist might eliminate the systematic influence of climatic variables, or for that matter the temporal variables, via *post hoc*

analyses in a covariance design in which the environmental influences are treated as covariates, and the behavioral observation scores are adjusted accordingly. It would seem sufficient, however, when the data do not range over a period of seasonal change, to recommend that control and treatment groups data be collected in the same time structure. That is, systematic change associated with climatic variables would have the most deleterious effect when, for example, the treatment-group data are collected in the spring and the control-group data in the winter.

For the temporal factors, it is recommended that time of day and day of the week should be the same for all observation periods. When this is not possible, *e.g.*, when several observations are made per week, care should be taken to sample equally the various days of the week or times of day. Furthermore, the days and times sampled should not be allowed to vary systematically across experimental groups. When these recommendations are followed, the identified variables will still contribute to uncontrolled variability in the obtained data. Their contribution, however, will be transformed from a systematic effect to random noise.

The argument here is that unless clinical researchers recognize the potential influence of temporal and climatic variables on their observational measures, they may incorrectly attribute change in behavioral scores to a treatment intervention, while the change may well be associated with systematic environmental influences, or, conversely, real treatment effects may be masked.

REFERENCES

- Bernal, M. E., Delfini, L. F., North, J. A., and Kreutzer, S. L. Comparison of boys' behaviors in homes and classrooms. In E. J. Mash, L. A. Hamerlynck, and L. C. Handy (Eds), *Behavior modification and families*. New York: Brunner/Mazel, 1976. Pp. 204-227.
- Chapman, L. J. A search for lunacy. *Journal of*

- Nervous and Mental Disease*, 1961, **132**, 171-174.
- Diamond, S. *Information and error*. New York: Basic Books, 1959.
- Faust, V. and Ladewig, D. Wettereinfluss auf drogenabhängige und depressive patienten. *Confinia psychiatrica*, 1972, **15**, 149-161. (English abstract)
- Wallis, W. A. and Roberts, H. V. *Statistics: a new approach*. Glencoe, Ill.: The Free Press, 1958.
- Received 9 October 1975.
(Final acceptance 6 December 1976.)