BASELINE MEASUREMENT OF RUNNING AWAY AMONG YOUTH IN FOSTER CARE

By

LUANNE R. WITHERUP

A THESIS PRESENTED TO THE GRADUATE SCHOOL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE DEGREE OF MASTER OF SCIENCE

UNIVERSITY OF FLORIDA

2005
# TABLE OF CONTENTS

LIST OF FIGURES........................................................................................................... iii
ABSTRACT....................................................................................................................... iv
INTRODUCTION .............................................................................................................. 1
Incidence of Running Away ....................................................................................... 1
Prior Research on Running Away .......................................................................... 3
Purpose of the Current Investigation .................................................................... 6
METHOD ........................................................................................................................... 8
Inclusion Criteria and Demographics .................................................................. 8
Data Collection ........................................................................................................ 8
Data Analysis .......................................................................................................... 9
Observer-Panel Evaluation ................................................................................... 12
Inter-Observer Agreement .................................................................................. 14
RESULTS ......................................................................................................................... 16
Single-Subject Interval-Based Measures .......................................................... 16
Single-Subject Episode-Based Measures .......................................................... 18
Observer-Panel Single-Subject Evaluation ......................................................... 25
Group-Size Analysis ............................................................................................ 27
DISCUSSION................................................................................................................... 31
REFERENCES ................................................................................................................. 37
APPENDIX A  SINGLE-SUBJECT INTERVAL-BASED MEASURES .......... 41
B  SINGLE-SUBJECT EPISODE-BASED MEASURES .................................. 58
C  GROUP INTERVAL-BASED MEASURES ....................................................... 75
BIOGRAPHICAL SKETCH ............................................................................................ 82
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Example single-subject interval-based measures (Finding 1)</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>Example single-subject interval-based measures (Finding 2)</td>
<td>21</td>
</tr>
<tr>
<td>3</td>
<td>Example single-subject interval-based measures (Finding 3)</td>
<td>22</td>
</tr>
<tr>
<td>4</td>
<td>Example single-subject interval-based measures (Finding 4)</td>
<td>23</td>
</tr>
<tr>
<td>5</td>
<td>Example single-subject episode-based measures</td>
<td>24</td>
</tr>
<tr>
<td>6</td>
<td>Observer evaluation- baseline acceptance</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>Observer evaluation- measure type</td>
<td>26</td>
</tr>
<tr>
<td>8</td>
<td>Example group interval-based measures</td>
<td>28</td>
</tr>
<tr>
<td>9</td>
<td>Observer evaluation- group-size analysis</td>
<td>30</td>
</tr>
<tr>
<td>10</td>
<td>Multiple-baseline designs</td>
<td>35</td>
</tr>
<tr>
<td>11</td>
<td>Single-subject interval-based measures</td>
<td>41</td>
</tr>
<tr>
<td>12</td>
<td>Single-subject episode-based measures</td>
<td>58</td>
</tr>
<tr>
<td>13</td>
<td>Group interval-based measures</td>
<td>75</td>
</tr>
</tbody>
</table>
Abstract of Thesis Presented to the Graduate School of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Master of Science

BASELINE MEASUREMENT OF RUNNING AWAY AMONG YOUTH IN FOSTER CARE

By

Luanne R. Witherup

December 2005

Chair: Timothy R. Vollmer
Major Department: Psychology

The primary purpose of this study was to demonstrate the usefulness of several different behavioral measures of running away by evaluating various measurable dimensions of running away for a sample of children in foster care. Participants included 84 runaways residing in one service district of the Florida Department of Children and Families (FDCF). All data were obtained from existing databases managed by FDCF. Seven baseline measures were calculated for each runner including (a) the number of run initiations, (b) the proportion of opportunity days in which the child initiated a run, (c) the number of days the child spent on the run, (d) the proportion of opportunity days that the child spent on the run, (e) the duration of successive run episodes, (f) successive episode inter-response times and (g) successive initiation inter-response times. In addition, an observer-panel evaluation was conducted to identify baselines that would be suitable for treatment evaluations. A group-size analysis was conducted by aggregating data from the original sample. Results demonstrate the usefulness of evaluating several
measurable dimensions of running away, and highlight the benefits of including
duration-based measures and correcting for opportunity. However, results suggest that
treatment evaluations may only be possible via an analysis of groups of runners rather
than individual subjects. Several conceptual implications related to the assessment and
treatment of runaway youth are discussed.
INTRODUCTION

Running away is possibly one of the most severe forms of problem behavior exhibited by adolescents (Biehal & Wade, 1999). Results of various studies indicate that children who run away from home are more likely than other children to abuse drugs (e.g., de Man, 2000; Edelbrock, 1980; Kennedy, 1991; Koopman, Rosario, & Rotheram-Borus, 1994; Yates, MacKenzie, Pennbridge, & Cohen, 1988), commit crimes (e.g., Abbey, Nicholas, & Bieber, 1997; Powers, Eckenrode, & Jaklitsch, 1990), engage in prostitution (e.g., Cohen, MacKenzie, & Yates, 1991; Yates, MacKenzie, Pennbridge, & Swofford, 1991), contract sexually transmitted diseases (e.g., Cohen et al., 1991; Yates et al., 1991), attempt suicide (e.g., Kennedy, 1991; Powers et al., 1990), join street gangs (e.g., Yoder, Whitbeck, & Hoyt, 2003), skip school (e.g., de Man, 2000; Sullivan & Knutson, 2000), and drop out of school (e.g., Yates et al, 1988). Research also indicates that runaways are likely to be physically and sexually victimized while on the run (e.g., Abbey et al, 1997; Hoyt, Ryan, & Cauce, 1999; Yates et al., 1991).

Incidence of Running Away

Given the serious risks listed above, several government reports and research studies have attempted to estimate the incidence of running away among youth in our society. A study by Hammer, Finkelhor, and Sedlak (2002) estimated that in 1999 approximately 1,682,900 children, representing approximately 2.6% of all youth in the United States, either ran away from home or were forced out by their caretakers (U. S. Bureau of the Census, 2000). Although this figure is most likely an overestimate of
running away due to the inclusion of children forced out by caregivers, difficulties in discriminating between such children have been noted by the researchers attempting to produce such estimates (e.g., Hammer et al., 2002).

Running away among foster children has received considerable attention in recent years due to the heightened publicity surrounding children missing from substitute care. Agencies responsible for the well being of these children must abide by policies and procedures designed to accurately track the whereabouts of missing children (Florida Statutes, 937.022, 2004). For this reason, state agencies often require foster parents to abide by strict reporting procedures for missing children. For example, Florida foster parents are required by state law to immediately report to the Florida Department of Children and Families (FDCF) when a child has run away from their home (Florida Administrative Code, 65C-13.010(4)(k)3, 2004). Due to these legal mandates, estimates of running away among foster children are potentially more accurate than estimates of running in the general population. Even so, considerable variability exists among estimates of foster children who run from care (Kaplan, 2004). According to the Department of Health and Human Services (2001), 9,112 foster children were on the run as of September 30, 2001. This value represents approximately 2% of the children in care on that date. The National Survey of Child and Adolescent Well-Being reported that 11% of 700 children in care for one year had attempted to run away (Department of Health and Human Services, 2003). Even higher estimates were obtained by Fasulo, Cross, Mosley, and Leavey (2002) during a study of 147 adolescents residing in specialized foster care. Results indicated that 44% of the children ran away at least once during their stay in care and 22% of the children ran permanently. However, the high
incidence of running found in that particular study may be due to the specific characteristics of children in specialized foster care. Several studies have also examined the extent to which children exit the child welfare system via a run episode. Such estimates range from as low as 2% to as high as 21% (Courtney & Barth, 1996; Department of Health and Human Services, 2001).

Prior Research on Running Away

A considerable number of research studies have focused on identifying risk factors associated with an increased likelihood of running away. Unfortunately, a cohesive literature base on risk factors for running away does not exist due to considerable variability in definitions and measurement procedures. Although mixed results are a frequent occurrence, some factors found to be associated with an increased risk of running away include a history of maltreatment (e.g., Famularo, Kinscherff, Fenton, & Bolduc, 1990; Yoder, Whitbeck, & Hoyt, 2001), increased age (e.g., Courtney & Wong, 1996; Yoder et al., 2001), being female (e.g., Courtney & Wong, 1996; Fasulo et al., 2002), placement instability (e.g., Abbey et al, 1997; Kashubeck, Pottebaum, & Read, 1994), placement in a group home (e.g., Courtney & Wong, 1996), a prior history of running (e.g., Kashubeck et al., 1994; Siegel & Callesen, 1993), drug and alcohol use (e.g., de Man, 2000; Edelbrock, 1980), emotional problems (e.g., Cohen et al., 1991; de Man, 2000), attempted suicide (e.g., Cohen et al., 1991; Edelbrock, 1980), academic problems (e.g., Rogers, Segal, & Graham, 1994; Sullivan & Knutson, 2000), and truancy (e.g., Sullivan & Knutson, 2000; Yates et al, 1988).

It is surprising to note that no investigations to date have attempted to obtain repeated measures of running away for individual children. Rather, researchers generally
categorize children as either runaways or non-runaways but make no attempt to track the occurrence of run episodes on an individual or even a group basis. This is unfortunate given that an effective behavioral analysis of running away necessitates the use of repeated measurement in order to properly evaluate behavior change. For example, if one were to evaluate the efficacy of an individual or large-scale behavioral intervention for running away, some type of baseline measure would be required.

The lack of research on repeated measurement of running away might be explained by considering the methodology commonly used by behavior analysts attempting to assess and treat problem behavior. Specifically, direct observation and repeated measurement of behavior-environment relations are standard components of traditional behavioral assessment procedures. Typical child problem behavior such as verbal and physical aggression, noncompliance, and tantruming are readily suited for such assessment procedures due to their relatively high rate (e.g., multiple times per day or hour) and overt nature. Conversely, running away is very low rate (e.g., once in a lifetime or multiple times per year) and is often initiated covertly, which makes direct observation unlikely and reduces the opportunity to obtain repeated measures.

Additionally, difficulties often arise when attempting to evaluate treatment effects for low-rate problem behavior. Specifically, the ability to produce a convincing behavior change upon the introduction of a treatment program is difficult when the baseline rate of problem behavior is very low. Modification of traditional assessment procedures and treatment evaluation methods may be required if behavior analysts plan to address this socially significant problem behavior.
Researchers attempting to study low-rate and covert problem behavior have employed several useful strategies to date. Application of these strategies to the study of running away may be a first step toward filling the existing research void. For example, researchers attempting to study covert problem behavior often make use of permanent behavior products (e.g., Grace, Thompson, & Fisher, 1996) or other naturally occurring records of behavior (e.g., McSweeny, 1978). A similar strategy may prove useful when attempting to study running away. More specifically, it may be possible to obtain missing child reports that are filed to law enforcement when a child runs away from home. Although such a report is not a direct product of the run episode, it is a direct product of the caregivers’ response and is presumably at least correlated with an actual run episode.

In cases where low-rate baselines make treatment evaluation difficult, one method that researchers have found to be effective is to alter the unit of interest by evaluating groups of individuals rather than individual subjects. For example, using both reversal (ABA) and multiple-baseline designs, several researchers have demonstrated behavior change across entire groups rather than for individual subjects. A study by Agras, Jacob, and Lebedeck (1980) demonstrated the effectiveness of a community-wide water conservation intervention by using a multiple-baseline-across-cities design. Similarly, a study by Kirchner (1980) evaluated the effectiveness of a helicopter patrol program at reducing residential burglaries using a multiple-baseline-across-neighborhoods design. Multiple-baseline designs have also been used across classrooms (e.g., Switzer, Deal, & Bailey, 1977), coal mines (Fox, Hopkins, & Anger, 1987), intersections (Van Houten & Retting, 2001), and highways (Van Houten & Nau, 1981). Reversal designs have also
been used to demonstrate behavior change for groups of individuals. For example, a study by Cope and Allred (1991) used a reversal design to demonstrate the effectiveness of a community wide intervention to deter illegal parking.

When attempting to study low-rate problem behavior, the use of duration measures is another possible strategy for evaluating potential treatment effects. At times, a change in the duration of problem behavior may be clinically significant even when the rate of behavior remains unchanged. In such cases, the use of duration measures may be more clinically relevant than rate measures by providing a more descriptive account of the problem at hand. This is clearly the case with the problem of running away given that children may be exposed to serious risks with each day spent on the run (Biehal & Wade, 1999). Surprisingly, government issued reports and research studies often focus only on the number of children on the run at a given time (e.g., Department of Health and Human Services, 2001) or the number of run episodes occurring over a specified period of time (e.g., Hammer et al., 2002). Little attention has been given to producing reliable estimates of the total amount of time children spend on the run.

**Purpose of the Current Investigation**

The primary purpose of the study was to demonstrate the usefulness of several different behavioral measures of running away by evaluating several measurable dimensions of running away for a sample of children in foster care. To accomplish this, naturally occurring behavior records in the form of missing child reports were obtained. Although any child or adolescent who runs away should be eligible for assessment and treatment at the clinical level, it is possible that many would have baselines insufficient to demonstrate experimental control of treatment effects due to the relatively low rate of
behavior. Therefore, a secondary purpose of this study was to examine (a) the likelihood of obtaining single-subject baselines suitable for treatment evaluations, and (b) the role of group size on the suitability of baselines for treatment evaluation. To quantify baseline suitability, a panel of observers was polled to evaluate the acceptability of individual and group baselines on the assumption that the baselines would eventually be used to evaluate the efficacy of a treatment. However, note that no treatment was actually evaluated as part of this investigation.
METHOD

Inclusion Criteria and Demographics

Data for all runaway foster children residing in one FDCF service district as of 10/12/04 were considered for inclusion in this study. A runaway was defined as a child who engaged in one or more run episodes between 9/1/01 and 10/12/04. This time interval was deemed by FDCF personnel to represent the most reliable period of data collection with respect to the accurate documentation of run episodes by children in foster care due to the implementation of a data system in September of 2001. Based on these criteria, 86 children were identified for inclusion. Two children were excluded from the analysis due to missing or insufficient information. Of the 84 runaways included in the analysis, 42 were female and 42 were male. The median age was 16 years (range 10-17 years), the median number of run episodes was 2 (range 1-19 episodes), the median number of days spent on the run was 10 (range 1-441 days), and the median number of years spent in foster care was 2 (range 0.12–15.6 years).

Data Collection

Data were obtained from two databases managed by FDCF. Data on run episodes were obtained from the Missing Child Tracking System (MCTS). This database records the initiation and recovery dates of all run episodes based on missing child reports that are filed to the Florida Department of Law Enforcement (FDLE). Data were obtained for all episodes initiated between 9/1/01 and 10/12/04. Demographic information including gender, age, and time spent in foster care was also obtained from the MCTS.
The MCTS database contained only some of the information needed to carry out this investigation. Therefore a second database called HomeSafenet (HSn) was used to obtain additional information. Specifically, placement history reports were obtained for each child from HSn. Such histories contain the placement and removal dates for each placement episode a child experiences while in foster care. These histories also include information about placements at lockdown facilities such as juvenile detention.

**Data Analysis**

**Interval-Based Measures**

The four interval-based baseline measures, described below, were calculated for each child across 30-day intervals beginning with the child’s first day in care or 9/1/01, whichever was later, and ending with the last completed interval expiring on or before 10/12/04. The average number of intervals evaluated for each child was 25 (range 1-37 intervals). At times, researchers or clinicians may wish to evaluate the effect of a particular intervention on groups rather than individuals (e.g., multiple-baseline-across-groups design). Therefore, 80 of the 84 runners were randomly selected for inclusion in a group-size analysis. A parametric group size analysis was accomplished by constructing 31 groups as follows: 16 groups of 5, 8 groups of 10, 4 groups of 20, 2 groups of 40, and 1 group of 80. For each group, the average value of each measure below was calculated across successive 30-day intervals. All measures were put into graphical form to allow for easy visual inspection.

**Number of run initiations**

The number of run initiations the child engaged in during each successive 30-day interval was calculated.
Proportion of opportunity days initiating a run

Due to the fact that run initiations cannot occur when a child is already on the run or when they are placed in a lockdown facility, the preceding measure may not provide an accurate account if the behavior is frequently prevented from occurring. This concern is warranted given that all of the children included in this analysis spent at least one day on the run and 40% of the children spent at least one day in a lockdown facility (median 0 days, range 0-876 days). Therefore, the number of opportunity days was calculated for each 30-day interval. For this measure, an opportunity day was defined as any day not spent entirely on the run or in a lockdown facility. Days in which a child spent only a portion of the day on the run or in a lockdown facility were considered to be opportunity days. Given that the number of opportunity days in each interval can fluctuate, the proportion (rather than the number) of opportunity days in which the child initiated a run episode was calculated for each successive 30-day interval.

Number of days spent on the run

Measures related to the occurrence of run initiations provide little information about the total amount of time a child spends on the run. Therefore, the total number of days spent on the run was calculated for each 30-day interval. Days in which a child spent at least some portion of the day on the run were considered to be a day spent on the run.

Proportion of opportunity days spent on the run

Due to the fact that there is no opportunity to be on the run while a child is placed in a lockdown facility, the preceding measure may not provide an accurate account for children who spend time in lockdown facilities. Therefore, the number of opportunity
days was calculated for each 30-day interval. For this measure, an opportunity day was defined as any day not entirely spent in a lockdown facility. Days in which a child spent only a portion of the day in a lockdown facility were considered to be opportunity days. Given that the number of opportunity days in each interval can fluctuate, the proportion (rather than the number) of opportunity days that the child spent on the run was calculated for each successive 30-day interval.

**Episode-Based Measures**

The following baseline measures were calculated based on an analysis of each child’s run episodes. These measures were not subject to the group size analysis. All measures were put into graphical form to allow for easy visual inspection.

**Run durations**

The duration of each run episode was calculated in days. Run episodes that were in progress on the date of data collection were indicated as such when displayed graphically. Therefore, minimum durations are depicted for such episodes rather than actual durations (i.e., final durations are unknown).

**Episode inter-response times**

The time elapsing between the end of each run episode and the beginning of the next episode was calculated in days. This measure was omitted for 29 children with only one run episode.

**Initiation inter-response times**

The time elapsing between successive run initiations was calculated in days. This measure was omitted for 29 children with only one run episode.
Observer-Panel Evaluation

In order to evaluate the acceptability of each baseline measure for use during treatment evaluations, a panel of five observers was constructed to evaluate the data. Observer panels of this sort have been used by prior researchers to assist with data interpretation for various purposes (e.g., Hagopian, Fisher, Thompson, & Owen-DeSchryver, 1997; Kahng et al., 1998). Both the single-subject and group data sets were subjected to observer evaluation. Initiation inter-response time was the only measure not subjected to observer evaluation given that the direction of behavior change associated with improvement is ambiguous. For example, an improvement in the rate of running would produce an increase in this measure, but an improvement in the duration of run episodes would produce no change in this measure. So although this measure may provide useful information about temporal patterning of run initiations, such a measure would not be appropriate for experimental evaluations of behavior change.

Observer-Panel Selection

Five individuals were selected for the observer panel based on their expertise in the field of Applied Behavior Analysis and their experience working with runaway foster children. Specifically, all observers possessed a doctorate degree, were Board Certified Behavior Analysts (BCBA), had at least one first-author publication in the Journal of Applied Behavior Analysis (JABA), and had work experience involving runaway foster children.

Observer-Panel Materials

Participation was requested from each observer selected and evaluation materials were provided upon acceptance. Observers were allowed to complete the evaluation
independently and were asked return materials upon completion. In addition to a basic
description of each measure as described above in the Data Analysis section, observers
were provided with the following written instructions:

The field of Applied Behavior Analysis has traditionally focused on the
assessment and treatment of high rate problem behavior such as self-injury
or aggression, which can occur multiple times per minute. In contrast,
behavior such as rape, murder, suicide and running away from home can
be extremely low rate. For this reason, research-based treatment
evaluations of such behavior may prove difficult if adequate baseline
measures cannot be obtained. In order to assess this difficulty with respect
to running away from home, we have compiled several relevant baseline
measures using a sample of 84 foster children who have run away at least
once. The data have been presented in both single-subject and group
formats. You have been selected to evaluate these data based on both your
expertise in the field of Applied Behavior Analysis and your experience
working with children who run away. Please evaluate the following data
sets under the assumption that you are a behavior analyst planning to
evaluate an intervention designed to address running away among
children. The assumption is that all children who run away could receive
interventions, but only a portion of those would be eligible for a proper
experimental evaluation (e.g., a multiple-baseline evaluation). After
reviewing the descriptions for each section of graphs (below), circle all
graphs that you feel represent adequate baselines by which to evaluate the
intervention.

Graphs packets were organized by type of measure rather than by individual
runner or group of runners. For instance, each page included baselines of a single
type (e.g., number of run initiations) for several of the individual runners or
groups of runners. A total of 599 graphs were presented. Interval-based single-
subject measures were presented first (336 graphs), episode-based single-subject
measures second (139 graphs), and interval-based group measures third (124
graphs). Participant and group numbers were not included, however group sizes
were denoted next to each group baseline.
Observer-Panel Data Analysis

To evaluate the likelihood of baseline acceptance, the total number of baselines designated as acceptable by a majority of the observer panel (i.e., at least 3 out of 5 observers) was calculated for each runner individually. Note that runners could attain a maximum of six acceptable baselines (i.e., all measures except for initiation inter-response times).

To evaluate possible differences in the likelihood of baseline acceptance based on type of measure and group size, the average proportion of observer acceptance was calculated for each interval-based and episode-based measure individually (excluding initiation inter-response times). Episode inter-response times that were omitted for runners with only one run episode were automatically designated as inadequate (i.e., proportion observer acceptance = 0). The proportion of observers designating the baseline as acceptable was first calculated for each of the baseline graphs individually. The average of these values was then calculated according to group size (interval-based measures only) and type of measure (e.g., number run initiations, days spent on the run).

Inter-Observer Agreement

Inter-observer agreement with respect to the calculation of all baseline measures was evaluated for 27 of the 84 children in this analysis (32%). A second observer calculated both the interval and episode-based measures for each child. An exact agreement measure was employed across all types of measures. For example, a disagreement was scored if Observer A counted two run episodes in a given interval and Observer B counted only one run episode. Average agreement was then calculated for each type of measure by dividing the total number of agreements by the total number of
agreements plus disagreements, then dividing by 100. Average agreement across all seven baseline measures was 99% (range 98% -100%).

Inter-observer agreement was also calculated with respect to ratings of acceptability among the observer panel. A pairwise exact agreement comparison was used to identify the extent to which each observer concurred with every other observer. A total of 599 acceptability ratings were obtained for each observer. Individual pairwise agreement scores were obtained by comparing each observer’s ratings (i.e., acceptable or not acceptable) with the ratings of each other observer (5 observers = 10 pairings). The average pairwise agreement score across all 10 pairings was 81% (range 68% -90%). An average pairwise agreement score was also obtained for each observer. For example, if Observer 1 agreed with Observer 2 on 100% of the graphs rated, but agreed with the remaining observers (i.e., 3, 4, and 5) on only 70% of graphs, the resulting average pairwise agreement score for Observer 1 would be 77.5%. Average pairwise agreement scores for the 5 observers were 72%, 82%, 82%, 84%, and 85%.
RESULTS

Single-Subject Interval-Based Measures

The utility of the four interval-based measures varied across children due to differences in run and lockdown histories. However, four potentially useful findings emerged when evaluating the measures obtained for individual children. Example datasets highlighting each of these four findings are presented in Figures 1-4. Interval-based measures for all 84 runners can be found in Appendix A. For Figures 1-4, each row depicts all four interval-based measures for a given runner. All measures are displayed across successive 30-day intervals along the x-axis. The number of data points displayed for each child will vary based on the amount of time spent in foster care. The first column depicts the number of run initiations, the second column depicts the proportion of opportunity days in which the child initiated a run, the third column depicts the number of days the child spent on the run (maximum 30 days), and the fourth column depicts the proportion of opportunity days that the child spent on the run. Missing data points (i.e., no data point between successive x-axis tick marks) will result for any intervals containing no opportunity days. Y-axis scales were adjusted on an individual basis to allow for proper analysis of trends. Baselines judged as acceptable for treatment evaluation by a majority of the observer panel (i.e., at least 3 out of 5 observers) are designated by light gray shading.

The first general finding that emerged from this analysis was that a majority of the children engaged in very few run episodes of minimal duration (51% of children ran less
than 3 times and spent less than 16 days on the run). Figure 1 depicts the interval-based measures for three such runners. With respect to behavioral trends, similar information is provided across all four measures for these children. In addition, these baselines were typically judged as unacceptable for treatment evaluation by the observer panel.

The second general finding was that baseline data for children with frequent run initiations was often times variable or on a downward trend across all interval-based measures. Figure 2 depicts the interval-based measures for three such runners. Data such as these were also viewed as unacceptable for treatment evaluation.

The third finding that emerged concerns the utility of duration measures and correcting for initiation opportunity. Given that children cannot initiate run episodes while on the run, improvements in the rate of run initiations were often accompanied by an increase in the amount of time spent on the run. In other words, spending time on the run artificially suppressed the rate of run initiations for several runners. In general, the utility of duration measures and correcting for initiation opportunity increased as time spent on the run increased for a given runner. Figure 3 depicts the interval-based measures for three runners who spent a substantial amount of time on the run (range 253-437 days). The advantages of using duration measures and correcting for initiation opportunity were apparent for these children. Note that although all runners showed a recent decline in number of run initiations (column 1), it became clear that these declines did not represent desirable outcomes once we corrected for initiation opportunity (column 2). Additionally, number of days spent on the run (column 3) and proportion of opportunity days spent on the run (column 4) were both high and stable for these
children. In these cases, the duration measures were most informative given the substantial amount of time these runners spent on the run, yielding high, stable, baselines.

The fourth general finding that emerged was the need to correct for opportunity to run for children who spent time in lockdown facilities. More specifically, we noted that when evaluating time spent on the run, the utility of correcting for opportunity increased as time spent in lockdown increased. Figure 4 depicts the interval-based measures for three runners who spent a substantial amount of time in lockdown facilities (range 229-402 days). Note that although all runners showed a recent decline in the number of days spent on the run (column 3), correcting for opportunity (column 4) indicates that these were forced improvements due to time spent in lockdown.

**Single-Subject Episode-Based Measures**

The episode-based measures provided a different way to evaluate behavior patterns by allowing for an explicit analysis of response duration and inter-response times that was not possible using interval-based measures. However, we found that the usefulness of such measures varied among children based on the total number of run episodes. More specifically, an analysis of trend in run duration was only possible for children engaging in two or more run episodes. Similarly, inter-response time trend analyses were only possible for children engaging in three or more run episodes.

Figure 5 contains the episode-based measures for 5 of the 84 runners in this analysis. Episode-based measures for all 84 runners can be found in Appendix B. Each row depicts all three episode-based measures for a given runner. All measures are displayed across successive run episodes along the x-axis. The number of data points displayed for each child will vary based on the total number of run episodes. The first
column depicts the duration of each run episode in days, the second column depicts successive episode inter-response times in days, and the third column depicts successive initiation inter-response times in days. Y-axis scales were adjusted on an individual basis to allow for proper analysis of trends. If a child was on the run as of the date of data collection, the associated run duration is designated as being in progress (IP) and represents only the minimum duration of the episode (i.e., final duration is unknown). Inter-response time measures were omitted for children who engaged in only one run episode. Baselines judged as acceptable for treatment evaluation by a majority of the observer panel are designated by light gray shading.

Data for runners R11, R43, and R83 in the top three rows are typical for children with few run episodes. Although data such as these provided limited information, it is important to note that with respect to run duration limited information may still prove useful. For example, the fact the runner R11 only remained on the run for 2 days suggests the possibility that she may be incapable of obtaining the basic needs required to maintain long absences from care (i.e., food, shelter). Such information could have important implications for treatment. Episode-based measures for children who engaged in many run episodes were inherently more informative. For example, data for runners R70 and R56 are much more descriptive due to the high number of run episodes.

In general, differences between episode inter-response time (column 2) and initiation inter-response time (column 3) were observed for runners with relatively long run episodes. For example, note the similarity in these two measures for runner R83 who had a maximum run duration of 9 days. In contrast, these measures differed substantially (note y-axes) for runner R56, who had a larger maximum run duration of 139 days.
Figure 1. Example single-subject interval-based measures (Finding 1). Each row represents data for one runner. Data for all 84 runners can be found in Appendix A.
Figure 2. Example single-subject interval-based measures (Finding 2). Each row represents data for one runner. Data for all 84 runners can be found in Appendix A.
Figure 3. Example single-subject interval-based measures (Finding 3). Each row represents data for one runner. Data for all 84 runners can be found in Appendix A.
Figure 4. Example single-subject interval-based measures (Finding 4). Each row represents data for one runner. Data for all 84 runners can be found in Appendix A.
Figure 5. Example single-subject episode-based measures. Each row depicts data for one runner. Data for all 84 runners can be found in Appendix B.
Observer-Panel Single-Subject Evaluation

The shaded baselines contained in Figures 1-5 provide specific examples of individual graphs judged as acceptable for treatment evaluations by a majority of the observer panel. Additional data analyses were also conducted to answer specific questions concerning the acceptability of baseline measures. One question of interest is the likelihood that a given runner would have one or more baselines judged as acceptable. Figure 6 depicts the percentage of runners attaining various degrees of baseline acceptance. The number of baselines judged as acceptable by a majority of the observer panel is displayed along the x-axis, with the percentage of runners along the y-axis. Given that initiation inter-response time was not subjected to observer evaluation, a maximum of six acceptable baselines was attainable. Results indicate that a large percentage of runners (62%) had no baselines judged as acceptable according to the majority criteria. The remaining 38% of the runners had at least one acceptable baseline and none of the runners had all six baseline measures judged as acceptable.

A second question of interest is was whether the likelihood of baseline acceptance would vary according to the type of baseline measure selected. Figure 7 depicts the average proportion of observer acceptance for all six baseline measures. A majority criterion was not employed, but rather the actual proportion of observers accepting each baseline graph was determined and then the average of these values was calculated for each type of measure. Recall that episode inter-response time baselines that were omitted for children with only one run episode were automatically considered unacceptable. Results for the interval-based measures (left side of graph) indicate that number of run initiations was the least accepted type of baseline measure (0.17), followed by the
Figure 6. Observer evaluation- baseline acceptance. Baselines approved by a majority of the observer panel were considered acceptable.

Figure 7. Observer evaluation- measure type. Average proportion observer acceptance for all single-subject baseline measures.
proportion of opportunity days initiating a run (0.20), the number of days spent on the run (0.23), and the proportion of opportunity days spent on the run (0.25). Therefore, initiation measures (i.e., number of run initiations and proportion of opportunity days initiating a run) were less accepted than duration measures (i.e., number of days spent on the run and proportion of opportunity days spent on the run), and correcting for opportunity increased average acceptance for both types of measures. Episode-based measures are depicted on the right side of the graph. Episode inter-response times attained an average acceptance similar to that of the interval-based measures (0.20), and run durations attained the highest acceptance overall (0.30).

**Group-Size Analysis**

In general, the same considerations discussed previously regarding the utility of correcting for opportunity and using duration measures apply to the group data sets. In addition, the relative variability observed in the data for all measures was negatively correlated with group size, in that smaller groups generally demonstrated more variability than larger groups. This result was predictable given that the aggregation of single-subject data will necessarily capture more behavior and reduce variability. Nonetheless, this result is potentially informative because it suggests that studying runaway behavior may be best accomplished using groups of runners.

Figure 8 contains the interval-based measures for 5 of the 31 groups constructed for the group analysis (one group of each size is displayed). Data for all 31 groups can be found in Appendix C. Each row depicts all four interval-based measures for a given group of runners. The group average of each measure across successive 30-day intervals is displayed on the y-axis. Although the number of intervals completed for each child
Figure 8. Example group interval-based measures. Each row represents data for one group. Group sizes are in bold next to each group number. Data for all 31 groups can be found in Appendix C.
varied, the last interval for each group represents the last completed interval for all children in the group. Baselines judged as acceptable for treatment evaluation by a majority of the observer panel are designated by light gray shading.

Although the shaded graphs in Figure 8 provide examples of acceptable group baselines, a more detailed parametric analysis of degree to which group size would increase baseline acceptance was also conducted. Figure 9 depicts the average proportion of observer acceptance according to both type of measure (legend) and group size (x-axis). Single-subject results were included for comparison (interval-based measures only).

Not surprisingly, results indicated that average acceptance increased as the size of the group increased. In addition, consistent with the single-subject analysis, duration measures fared better than or equal to initiation measures across all group sizes. Although maximum acceptance was reached by group size 20 for duration measures, initiation measures did not reach maximum acceptance until group size 80. One unexpected finding is worth noting as well. Correcting for initiation opportunity (i.e., proportion of opportunity days initiating a run) did not produce any increases in acceptance as was observed in the single-subject analysis. In fact, correcting for initiation opportunity actually decreased acceptance for group sizes 5, 10, and 20.
Figure 9. Observer evaluation- group-size analysis. Average proportion observer acceptance is depicted according to both by group size (x-axis) and type of measure (legend).
DISCUSSION

Given the lack of behavioral research targeted at the problem of running away, even the most basic issue of measurement has yet to be thoroughly addressed. Difficulties surrounding how and what to measure with respect to running away must be resolved before more complex issues such as the identification of behavioral function can be addressed. The current investigation demonstrated the usefulness of several different behavioral measures of running away and examined their suitability for use during treatment evaluation.

In general, results indicate that runaway behavior can be quantified along several measurable dimensions, but that the usefulness of a particular type of measure will vary across children. More specifically, the utility of using duration measures (in addition to rate measures) and of correcting for opportunity was most apparent for children with extensive run and lockdown histories. Similarly, episode-based measures including run durations and inter-response times were more descriptive for children with a higher number of run episodes. Results of the observer-panel evaluation suggest that single-subject baselines may often be unacceptable for treatment evaluations, with 62% of runners in this sample having no baseline measures viewed as acceptable. However, results also indicate that using duration-based measures, correcting for opportunity, and evaluating groups of runners may allow for successful treatment evaluations by capturing the occurrence of more behavior and reducing variability in the data.
For the clinical assessment and treatment of running away, virtually all types of baseline measures have the potential to provide useful information. Even baselines similar to those in Figure 1, which were judged as unacceptable for use during treatment evaluations, have the potential to provide extremely useful information when used in conjunction with other assessment procedures. For example, assume that subsequent assessment by a behavior analyst revealed that runner R28 (see Figure 1) was separated from her siblings during the same interval containing her only run episode. This information may lead to an effective preventive intervention based on the possibility that separation from siblings serves as the primary establishing operation.

Although all baseline measures may prove informative in some respect, results of this study highlight the need for both clinicians and researchers to carefully consider the possible implications of the type of baseline measure they choose to use (or not use). Arbitrary selection of a baseline measure could obscure pertinent information and ultimately hinder treatment effectiveness or undermine the detection of important treatment outcomes. Although the use of various types of baseline measures would ensure the most thorough analysis possible in all cases, results of this study suggest that certain types of measures may be especially useful for children with high run durations and substantial lockdown histories. For example, the utility of duration measures was apparent for children with lengthy run durations given that associated changes could not be observed through the use of rate measures alone. This point is especially important for agencies emphasizing recovery efforts for children already on the run, given that the effect of such efforts may be reflected largely through the use of duration-based measures.
Another consideration suggested by the results of this study is the need to correct for a lack of opportunity when measuring running away. For example, certain environmental circumstances may prevent the occurrence of run episodes (e.g., time spent in lockdown facilities), and failure to correct for this lack of opportunity may distort baseline data and alter interpretations regarding behavior change. The utility of correcting for opportunity was most apparent for children with high run durations and substantial lockdown histories. However, it is important to correct for opportunity in all cases in order to ensure the most accurate account of each child’s run history. Childcare agencies should carefully consider the above implications when establishing data collection requirements and performance standards related to runaways.

Based on the results of the observer-panel evaluation, treatment evaluation is clearly an area in which both clinicians and researchers will face the most difficulties. Specifically, the evaluation of treatment effects necessitates the use of individual baselines that are capable of demonstrating convincing behavior change. This applies to both clinical evaluations of single cases and research-based evaluations using multiple-baseline designs. However, as results of this study suggest, such baseline measures may be difficult or impossible to obtain for a majority of individual runaways. Therefore, clinicians and researchers attempting to conduct treatment evaluations will face a difficult challenge. One strategy suggested by the results of this study is the use of duration-based measures (e.g., number of days spent on the run) rather than initiation-based measures (e.g., number of run initiations). In the current study, duration-based measures were more likely than initiation-based measures to be judged as acceptable for use during a treatment evaluation. Results of the observer-panel evaluation also suggest that grouping
runaways in the context of single-subject methodology logic (i.e., multiple-baseline-across-groups) may prove to be an effective strategy. Baseline acceptability in the present study increased over that of single subjects for all group sizes, including as few as five runaways per group. Using groups of 20 or greater may almost guarantee acceptable baselines depending on the type of measure selected. These strategies will allow behavioral researchers to conduct treatment evaluations for running away without abandoning single-subject research design logic or being forced to rely on anecdotal report of treatment effectiveness. Obviously, the use of reversal designs or intentional baseline extensions during a multiple-baseline design would be too dangerous. Thus, the use of naturally occurring baselines appears to be the most promising approach to treatment evaluation. Figure 10 demonstrates how a multiple-baseline evaluation might be conducted using either a single-subject or group format. Individual runners (left column) or groups of runners (right column) with acceptable baselines were arbitrarily selected for the purpose of this demonstration.

One limitation of the current study is that the reliability of the data contained in the FDCF databases was not explicitly examined. Although reporting and data-entry errors are almost inevitable, the rate and magnitude of such errors has not yet been determined. Evaluations aimed at the identification and correction of such errors will be necessary to ensure that FDCF databases are a reliable source of information. For example, it may be possible to identify errors by cross checking multiple databases for conflicting information or by comparing database records to a secondary source of information such as foster parent report. The mechanisms needed to effectively carry out
such evaluations are not yet in place, therefore the retrospective nature of this analysis made it difficult to assess database reliability.

The current investigation was also limited in scope, in that it focused solely on the measurement of running away and did not directly evaluate any specific assessment or treatment procedures. However, given the critical role of measurement in behavioral methodology, this study was designed to serve as a catalyst for such endeavors. Future research in this area should seek to develop and test behaviorally-based assessment and treatment methods. Some investigations currently being conducted by our own research team include (a) analysis of various foster child characteristics associated with running away, (b) assessment of maintaining variables for running based on child and caregiver
verbal report and (c) assessment of run probability by placement type (e.g., group homes) and individual caregiver (e.g., a particular foster caregiver).

Behavioral researchers should also begin to consider the issue of prevention when attempting to address severe problem behavior such as running away. Given that even a single episode of running away can pose serious risks (Biehal & Wade, 1999), the prevention of running among children who do not yet exhibit such behavior is a top priority (Kaplan, 2004). Unfortunately, traditional single-subject research methods are not readily suited for an analysis of preventive interventions. Therefore behavioral researchers will ultimately face yet another methodological challenge in their effort to address this socially significant problem behavior.
REFERENCES


Figure 11. Single-subject interval-based measures. Each row represents data for one runner.
Figure 11. Continued
Successive 30-day Intervals

Figure 11. Continued
Figure 11. Continued
Figure 11. Continued
Figure 11. Continued
Figure 11. Continued
# Run Initiations

Proportion Opp Days Initiating Run

# Days on Run

Proportion Opp Days Spent on Run

Successive 30-day Intervals

Figure 11. Continued
Figure 11. Continued
Figure 11. Continued
Figure 11. Continued
Figure 11. Continued
Figure 11. Continued
Figure 11. Continued
Successive 30-day Intervals

Figure 11. Continued
Figure 11. Continued
Figure 11. Continued
Figure 12. Single-subject episode-based measures. Each row represents data for one runner.
Figure 12. Continued
Run Durations  | Episode IRTs  | Initiation IRTs

Successive Run Episodes

Figure 12. Continued
Successive Run Episodes

Figure 12. Continued
Successive Run Episodes

Figure 12. Continued
Figure 12. Continued
Successive Run Episodes

Figure 12. Continued
Figure 12. Continued
Run Durations | Episode IRTs | Initiation IRTs
---|---|---
R41 | R42 | R43
R44 | R45

Successive Run Episodes

Figure 12. Continued
Run Durations | Episode IRTs | Initiation IRTs
--- | --- | ---

Successive Run Episodes

Figure 12. Continued
Successive Run Episodes

Figure 12. Continued
Figure 12. Continued
Figure 12. Continued
Run Durations  

Episode IRTs  

Initiation IRTs  

Successive Run Episodes

Figure 12. Continued
Run Durations | Episode IRTs | Initiation IRTs

Successive Run Episodes

Figure 12. Continued
Run Durations

Episode IRTs

Initiation IRTs

Successive Run Episodes

Figure 12. Continued
Figure 12. Continued
Figure 13. Group interval-based measures. Each row represents data for one group. Group sizes are in bold next to each group number.
Figure 13. Continued
Figure 13. Continued
30-day Intervals

Figure 13. Continued
30-day Intervals

Figure 13.  Continued
Figure 13. Continued
Figure 13. Continued
BIOGRAPHICAL SKETCH

Luanne Witherup graduated from the University of Florida in 2001 with a B. S. in psychology and a minor in business administration. Following graduation, she became employed as a Behavior Analyst/Research Assistant for the University of Florida Behavior Analysis Services Program (BASP). BASP provides services to foster children and foster families throughout the state of Florida and is sponsored by the Florida Department of Children and Families. Ms. Witherup began her graduate studies at the University of Florida in 2002 and is currently pursuing a doctorate degree in psychology with a specialization in applied behavior analysis. Her primary research interests involve the development of assessment and treatment methodologies for runaway foster children.