

Single-case experimental designs in veterinary research

Richard B. Evans, PhD; Michael G. Conzemius, DVM, PhD; Duane A. Robinson, DVM; Scott R. McClure, DVM, PhD; Jessica A. Dahlberg, BS; Thomas D. Brown, PhD

Inferential studies to investigate the health of rare, exotic, or companion animals are often challenging because sample groups of sufficiently large sizes are difficult to obtain. This may be a result of limited availability of exotic or endangered animals or the ethical desire to limit the number of animals involved in painful or terminal studies. In standard study designs with adequate statistical power, the number of animals required for inclusion is often greater than the number that can be obtained or feasibly managed. Also, the cost of providing long-term care for research animals is often greater than the available funding for research in animal health. The purpose of this article is to provide an overview of single-case study designs and analyses and outline how these types of analyses may be used for animal studies when conditions warrant the use of a single subject.

Specific statistical methods and study designs have been developed for use when very few subjects are available; the latter are known as small N study designs (where N represents the number of study subjects) or single-case study designs when 1 study subject is used. These study designs were developed primarily in the field of human behavioral science^{1,2} and were used to monitor and evaluate patients and clinical practice.³ Sometimes, the characteristics of an individual patient (eg, age, sex, or disease status) may not match the inclusion criteria of published research projects that evaluate treatment options. In such situations, the proponents of evidence-based medicine have suggested using single-case designs to assess treatment effects for a particular patient.⁴ Single-case study designs are also called single-subject or single-system designs. The latter term may be useful when applying the method to an agglomeration of units that are treated as an entity, such as a litter, herd, farm, or production facility. For convenience in this article, we use the term subject or single case to refer to an individual animal and an aggregation of animals that is treated as an entity.

Single-case designs are used to assess a subject over time (ie, before, during, or after 1 or more interventions such as surgical, medical, and behavioral treatments or biosecurity measures). Usually, the subject is also assessed during a postintervention follow-up period.

The idea behind the design is to collect enough data to allow statistical comparisons among the different phases of treatment; the intent is to determine whether a change in the course of a medical problem was a consequence of the intervention rather than chance. The data analysis must take into consideration the repeated-measures (ie, autocorrelated) aspect of the data, which may induce correlation among the observations.

There are arguments for and against the use of single-case designs in research. Group-based methods provide inference for a supposedly typical subject, but single-case designs infer the effect of treatment on a single subject. Therefore, with regard to provision of study results that can be generalized to a population, single-case designs cannot replace classical designs. Nevertheless, single-case designs are longitudinal and provide insight into the disease process and the effect of medical intervention over time.

Single-case study designs are members of a large class of statistical designs known as quasi-experiments. Quasi-experiments typically lack 1 or more features (eg, control groups, randomization, or causal hypotheses) of a complete experiment. However, quasi-experiments have been widely used to identify trends or develop hypotheses in scientific fields such as criminal justice and behavioral sciences, for which subject characteristics or ethical issues make it difficult to perform a traditional experiment. Single-case designs and analyses differ from case study reports in that they provide quantitative, inferential information about disease processes and interventions, which is important for the medical community and not available in a case study report.

In some research (eg, kinetic studies), the temporal process of the disease process is known or expected. The objective of such a study is to assess features of the data attributable to the physiologic process that are affected by interventions.

In this review, we begin with a description of designs for single-case experiments, followed by a discussion of the baseline and intervention phases. Data analyses appropriate for single-case designs are outlined, and a description of a single-case experiment to assess variation in the gait of an emu is provided; the latter is further discussed, as are single-case study designs in general.

Single-case Study Designs

In a single-case study, a disease and the effect of treatment over time are monitored in 1 subject; outcomes are measured serially at predetermined inter-

Received February 7, 2005.

Accepted May 5, 2005.

From the Orthopaedic Research Laboratory, College of Veterinary Medicine, Iowa State University, Ames, IA 50011 (Evans, Conzemius, Robinson, McClure, Dahlberg); and the Orthopaedics Biomechanics Laboratory, University of Iowa, Iowa City, IA 52242 (Brown).

Supported by the National Institutes of Health (No. 5R01AR049919-02).

Address correspondence to Dr. Evans.

PVF Peak vertical force

GRF Ground reaction force

GEE Generalized estimating equation

vals. These studies have 2 or more phases; the first is the baseline phase, followed by phases in which measurements are obtained after the implementation or cessation of interventions. Each phase consists of a set of serially measured outcomes. A common design is denoted AB, where A represents the baseline phase and B represents the intervention phase. Another common design is denoted ABA, in which there is a baseline phase and an intervention phase and then the intervention is withdrawn for the third (follow-up) phase. Note that the follow-up phase is designated A because it is identical to the baseline phase with respect to the intervention. In general, single-case designs are denoted with capital letters, with each letter representing a unique phase. There are many possible designs and common classical designs (eg, multi-factorial designs) that can be accommodated in the single-case framework.⁵

Data collected as part of an experiment conducted by one of the authors (SRM) to assess the analgesic effect of shockwave treatment on a horse with caudal heel pain provide an example of data summarized from a single-case study of the AB design (Figure 1). In this experiment, PVF (the maximum amount of downward force during stance phase, which is decreased in association with lameness) was assessed at intervals after the initial observation; mean PVF (adjusted for body weight [N/kg]) of the left forelimb of the horse, determined while the horse was trotting over a force platform during a single kinetic analysis session, was plotted against time (days) from the initial observation. There were 5 datum points in the baseline phase (A) and 7 datum points in the follow-up phase (B); in this experiment, there were no data from the intervention phase, which involved a single shockwave treatment that lasted only a few minutes and was administered on day 5, 8 hours before the next observation. The informational value of the single-case design is apparent graphically even before formal statistical inferences are made among phases; the summarized data indicate a

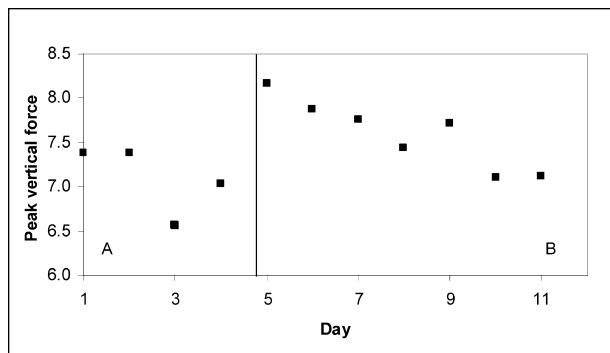


Figure 1—Plot of PVF (adjusted for body weight; N/kg) at intervals after the initial assessment in 1 horse with caudal heel pain to illustrate summarized data from a single-case study comprised of a baseline phase (A) and follow-up phase (B) after shockwave treatment on day 5. Each datum point is the mean PVF of the left forelimb of the horse while trotting over a force platform during a single kinetic analysis session. The solid vertical line demarks the baseline phase (labeled A) from the follow-up phase (labeled B). Datum points indicate a stable baseline followed by a large (relative to baseline) PVF at day 6, suggestive of a strong initial analgesic effect, which is reduced over subsequent days.

stable baseline followed by a large (relative to baseline) PVF at day 6, suggestive of a strong initial analgesic effect that is decreased over subsequent days. The next step is to use statistical inference to attribute the apparent difference in phases to the effect of treatment, rather than to chance. The statistical techniques for this example are described in a subsequent section.

Variables that can be measured reliably over time should be used for single-case studies. Binary variables, scores, and continuous measures can be used, although the method of statistical analysis may differ depending on data type. Endpoint variables, such as death, are not suitable because the observation of a positive result ends the ability to collect data serially and terminates the study. Typically, objective measures of outcome have fewer sources of variability than subjective measures and may provide more stability in the outcome assessment. For example, it is preferable to assess gait by use of kinetic or kinematic analyses, rather than by use of subjective lameness scores.

Baseline Phase in a Single-case Study Design

The purpose of the baseline phase is to obtain a reference for comparison with the effect of intervention and subsequent phases. A sufficient number of baseline measurements must be obtained over time to adequately characterize the state of the subject before intervention. The data may have large or small variation and may be increasing, decreasing, or oscillating over time. Unstable data during the baseline phase may make it difficult to compare with other phases because the unstable data fail to characterize the subject in a manner that can be easily summarized with a single statistic (eg, a mean value). For example, a lame animal may become increasingly lame during the baseline phase as a consequence of disease progression; a force-time plot of PVF would reveal a decrease in values for the lame limb over time. Consequently, the mean of the baseline data would include early values that no longer represent the subject's kinetic parameters and inflate the mean GRF value.

If the data values are increasing or decreasing (ie, there is a trend of sorts), some other phase summary may be a better single-point representation of the phase. Common summaries include the slope (from simple linear regression analysis), median, and SD value. Alternatively, data transformations (eg, logarithms) may adjust the data to a scale that permits easy summarization and analysis. Transformations are most useful when the scientific questions can be answered on the transformed scale. However, it may be the effect of treatment on the baseline trend that is of interest. For example, it is possible to conceive of circumstances in which a morning hormonal rise that leads to thyroid gland activity might be the baseline situation and the need is for a drug to suppress or enhance the rate of rise. In that instance, removing the trend would fail to capture point of the study.

Some characteristics of phase plots can be determined graphically (Figure 2). With data that are associated with trend (ie, data that are increasing or decreasing with time), observations obtained at the beginning of the phase are predictably different from

those at the end of the phase. With relatively unstable data, the datum points may appear to be centered around a horizontal line, but the SD value is large and there is large uncertainty about the phase mean. In contrast, ideal baseline phase data have low variation around an overall mean that represents the phase.

Intervention Phases in a Single-case Study Design

The phases are demarked by the introduction or cessation of 1 or more interventions. It is important to clearly define the intervention in the context of time. Some interventions fit into the phase framework, such as treatment of wounds or application of therapeutic horseshoes for treatment of fractures of the distal phalanx. These interventions are implemented over a period that is long enough to obtain serial measurements for comparison with baseline measurements.

However, the characteristics of some clinical procedures prohibit data collection, and these interventions cannot be used directly as a phase. These interventions are implemented over a period that is too short to obtain measurements or indispose the subjects to measurement; for example, the duration (eg, several hours) of an orthopedic surgical procedure may be too short to allow collection of data, and concomitant anesthesia of the patient prevents gait analysis. In that situation, data may be collected from the baseline and follow-up (postoperative) phases. Alternatively, the intervention may be defined to include the surgery and postoperative rehabilitation.

Selection of Measurement Intervals

It is important to determine the number of observations in a phase before the study begins. It has been shown that modifying phase lengths on the basis of

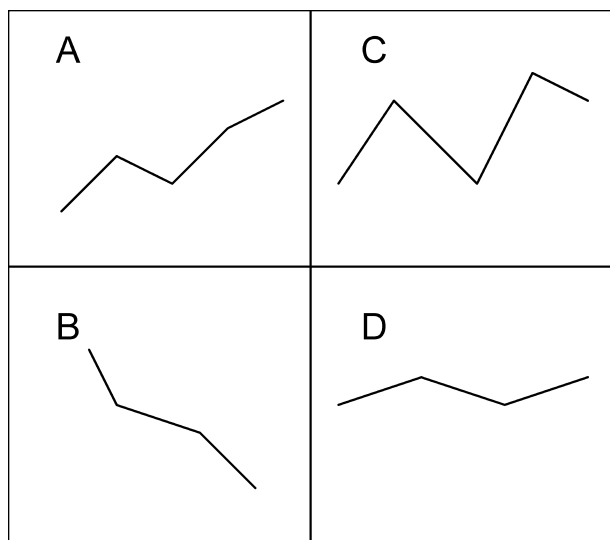


Figure 2—Set of hypothetical plots to demonstrate some characteristics of phase plots. A and B—Data associated with a trend (increasing or decreasing). Observations obtained at the beginning of the phase are predictably different than those at the end of the phase. C—Example of relatively unstable data. Although the data appear to be centered around a horizontal line, the SD value is large and there is large uncertainty about the phase mean. D—Ideal baseline phase data. The variation is low around an overall mean that represents the phase.

inspection of the collected data can bias the results.⁵ Equal intervals between observations are not required, but unequal intervals may induce inconsistent correlation structures among the observations. Adjacent observations with short intervals between them may be more correlated than adjacent observations made at longer intervals. These correlation structures must be accommodated by data analyses. Finally, if > 1 intervention is used, the sequence and timing of implementation and cessation of interventions should be arranged so that there are no carryover effects (ie, the physiologic function after the intervention) among interventions. However, in some types of research (eg, kinetic studies), it is the carryover effect that is of primary interest.

Display of the Data

Graphical representation of data is a fundamental part of the analysis and interpretation of serial data because graphs visually display characteristics of the disease process and effects of interventions over time and allow identification of serial patterns. Single-case data are typically displayed with time on the x-axis, the outcome variable on the y-axis, and vertical lines demarking the phases (Figure 1). When the intervals between observations are not uniform, marks on the x-axis of the graph should be consistent with the data and have spacing proportional to the time between measurements.

Some additional visual, noninferential data analysis may also be included on graphs to emphasize features of the data or to simplify so-called noisy data. Typically, these are lines or joined line segments overlaid on the graph for each phase. Regression lines can be used to emphasize trends within a phase, or mean lines can be used to contrast phase means. Noisy data can be smoothed with running mean values and line segments joining smoothed data. Splines can also be fit to noisy data or nonlinear data to reveal an overall impression of the data pattern over time.

For example, the data described previously (collected as part of a single-case study [AB design] to assess the analgesic effect of shockwave treatment on a horse with caudal heel pain) can be plotted as PVF versus time with the superimposition of 2 regression lines (1 each for phases A and B; Figure 3). The lines summarize the effect of the shockwave intervention, which involves an initial increase in PVF relative to baseline (indicating a decrease in caudal heel pain in the affected hoof), followed by a decrease in the analgesic effect of the treatment. It is possible to statistically compare slopes of the regression lines.

Analysis of Single-case Data

For the analyses of data collected via single-case study designs, phases are treated as groups in a traditional study design, and mean values or other parameters of interest are compared among phases (ie, groups) to assess the effect of implementation or cessation of a treatment. The data are often serially correlated so that many standard statistical methods for comparing groups (eg, *t* tests, ANOVAs, or linear regression analy-

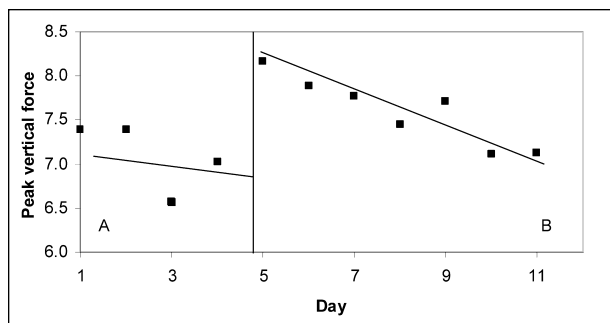


Figure 3—Same plot of PVF (adjusted for body weight; N/kg) at intervals after the initial assessment in 1 horse with caudal heel pain as in Figure 1, with a regression line superimposed on the plot for each phase. The lines summarize the effect of the shock-wave intervention: an initial increase in PVF relative to baseline (indicating a decrease in caudal heel pain in the affected hoof) followed by a decrease in the analgesic effect of the treatment. The slope of the first line is not significantly ($P > 0.05$) different from zero, but the second line has a significantly negative slope.

ses) are not valid; those methods assume independent observations. Improper assumption of independence will not bias estimates (eg, regression slopes or mean values) but may overstate significant findings.

There are 2 ways to assess the amount of correlation in the data. Intraclass correlation⁶ is a single-number summary of the correlation in the data. It is the ratio of between-phase variation to the overall variation in the data. If the ratio is near 1, then the variation between phases (eg, interventions) accounts for most of the variability and there is substantial correlation (also known as cluster effect) within the phases. Autocorrelation can be used to measure correlation among neighboring time points within each phase. It is a variation of the Pearson correlation (r^2), where within each phase, r^2 is calculated between the phase data and similar data that are shifted by 1 or more time points. In other words, the pairs of points (used to calculate r^2) are neighboring datum points. For example, if data were {1, 2, 3, 4, 5}, then the lag-1 autocorrelation for that set would be r^2 calculated on the set of pairs {(1, 2), (2, 3), (3, 4), (4, 5)}.

Randomization tests, repeated-measures ANOVAs, GEEs, and time-series models are the most common statistical methods used to compare groups and accommodate data dependence. It is important to consider which variables or parameters of the phase best represent the scientific question and should be compared among groups by use of a statistical test. Randomization methods can test the equivalence of virtually any parameter among groups by comparing estimates of the parameters. However, some statistics may not appropriately summarize a phase. For example, if the data in the baseline phase are stable but the data in the intervention phase have a trend, then comparison of group means may be misleading. In that instance, comparison of group slopes may be more appropriate. One advantage of randomization tests is flexibility; almost any statistical value can be used as a test statistic. Randomization tests are robust and easy to use and lack the need for model or distributional assumptions; however, for valid inference about group parameters, the underlying distributions must other-

wise be the same across groups.⁷ In other words, a randomization test may indicate significant differences among groups but it may not be the intended parameters (often median values) that are different, which may lead to misleading interpretation of the results.

Repeated-measures ANOVA is also easy to use but has strict assumptions about the correlation structure (called sphericity) in the data. Many statistical software programs automatically check the data for the sphericity assumption and provide adjustments to P values when the assumption is not met. Also, repeated-measures ANOVA compares group means, and in the case of data that have a trend or are unstable, the mean may not be an appropriate group summary.

A flexible statistical analysis method for correlated data is the use of GEEs.⁷ Generalized estimating equations are most often used in an ANOVA or regression setting, when some aspects of the data (eg, clustering or repeated measures) induce a correlation structure on the residuals. Generalized estimating equations can model several different correlation structures and are more flexible than repeated-measures ANOVAs. One important feature of the GEE method is that it is robust to departures in some model assumptions.

Time-series models are used to model trends and correlations in data and can be more powerful than randomization tests but rely on distributional assumptions and models that may be difficult to verify from small data sets (sets with relatively few serial observations/phase). Most time-series models are based on regression models and have a single dependent outcome that is modeled with several time points as independent variables. Time-series analyses often use the general concept of stationary data sequences. Phase data are considered stationary when there is no trend over time and the variability is consistent over time. If the data are nonstationary (eg, there is an increasing trend) and the investigator prefers a stationary sequence, then there are standard methods of transforming the data into a stationary sequence. One of the methods is that of differencing, in which adjacent points are subtracted.⁶

If there are more than 2 phases, then some subsets of the phases may be compared in different ways. For example, the means of the baseline and follow-up phases may be compared for equivalence, but the intervention phase may have a trend, so the slopes of the intervention and follow-up phases may be compared.

Confounding Over Time

Confounding variables in a single-case experiment are similar to confounders in traditional studies. These are variables that affect the outcomes but are disproportionately represented in phases. However, unlike traditional studies, the control (baseline) and postintervention data are not collected simultaneously in single-case experiments and a confounding variable may be introduced during the study. For example, key personnel may be replaced during the experiment, perhaps after the completion of a phase; a new investigator may bring subtle differences to the study that affect the outcomes either by altering the intervention or the assessments of outcomes.

If the single case under investigation is a system (eg, herd or food animal confinement), confounders also consist of changes in the population of the system between phases. The introduction of new livestock or the sale of some animals could imbalance variables such as breeds, sex, and age among phases and affect outcomes in combination with the intervention.

Example of a Single-case Experiment

Femoral head osteonecrosis is a debilitating disease that affects 20,000 to 25,000 Americans each year. Despite considerable improvements in early detection and staging of the disease, treatment options are limited. In an attempt to better understand this disease in humans, an animal model involving emus has been developed.⁸ Femoral head osteonecrosis can be induced in emus by use of a liquid nitrogen probe that is surgically inserted into the femoral head. To develop an accurate model and biomechanically evaluate the hip joint during locomotion, values of GRF are obtained via a force platform. To establish an emu gait analysis protocol that controls for sources of variation, it is important to know whether there is a day-to-day change in GRF values in emus (ie, a day effect). Rumph et al⁹ determined that 20% of variation in GRF values in dogs was attributable to a day effect.

Adult emus can be difficult to handle. To improve their behavior with regard to the force platform, 5 emu chicks were obtained when they were 5 days old. These birds were hand-raised and trained to walk across a force platform.^a During the training process and subsequent research studies, 4 emus failed to meet the study inclusion criteria or died of reasons unrelated to the assessment of GRF. Therefore, only one 1-year-old emu that had no history of lameness or orthopedic disease was available for assessment. The emu was trained by multiple handlers, one of which handled the emu during the gait analysis procedure.

It is difficult, time consuming, and costly to train emus, and procurement of additional trained emus for this study was not possible. A single-case study design was deemed an appropriate method with which to obtain GRF data, and the subsequent analysis was used to provide inference about the day-to-day variation in the GRF of the 1 remaining emu.

Procedures—The single-case study design for the emu study was ABCD, where A represents the baseline

phase and B, C, and D represent force platform data collections on different days. In effect, the interventions are days in this study. Each phase consisted of at least 5 valid serial measurements of GRF. Details of the force platform system and general kinetic analysis protocol have been reported.¹⁰

Trial velocity, acceleration, PVF, and vertical impulse (the area under the z-force force-time curve corresponding to the lame limb) were recorded and used in the analysis to assess gait. Summary statistics (including lag-1 autocorrelation) for each phase were calculated for all variables. Autocorrelation is the correlation among observations of a single variable (eg, PVF). Lag-1 autocorrelation is the correlation between adjacent observations. By use of statistical software,^b a randomization test (involving an F statistic) was used to test the hypothesis that phase means were equal among the group. An ANOVA table was used to calculate the percentage of the total error attributable to phase. However, an ANOVA test was not appropriate for these data, so the percentage was tested for significance against 0 by use of a randomization test.

Emu study results—Mean velocity, but not acceleration, differed among the phases and confounded the effect of velocity with GRF for day. Linear regression fit the data and was used to adjust the GRFs by controlling the effect of velocity. The velocity-adjusted data (the residuals from the regression) were used for the remainder of the analysis.

Summary statistics for unadjusted PVF and vertical impulse for each phase were calculated (Table 1). The GRF values for the emu during walking were consistent with bipedal gait, with PVF greater than the emu's body weight for each step. Inspection of the means and SD values suggested that there may be differences in the group means. However, on examination of the adjusted data by use of the randomization test, there was no significant ($P > 0.05$) difference in mean PVF or vertical impulse among phases.

The adjusted PVF data were plotted by trial number in serial order (Figure 4). The data did not appear visually to have trends that were a function of time or phase. The variation among phases was different, but equal variation among groups is not an assumption of a randomization test. The data for vertical impulse had no additional visual characteristics of interest (data not shown).

Table 1—Mean values \pm SD for PVF, vertical impulse, velocity, and acceleration during each phase of a single-case gait analysis study involving 1 emu.

Variable	Phase (No. of trials/phase)			
	A (n = 16)	B (5)	C (8)	D (7)
PVF	118.5 \pm 11.3	127.9 \pm 22.1	164.6 \pm 36.3	130.7 \pm 20.3
Vertical impulse	49.3 \pm 6.8	46.6 \pm 14.3	50.8 \pm 24.6	39.2 \pm 19.5
Velocity	1.0 \pm 0.3	1.1 \pm 0.5	1.4 \pm 0.6	1.4 \pm 0.5
Acceleration	-0.1 \pm 0.1	-0.2 \pm 0.3	1.4 \pm 0.6	-0.7 \pm 0.5

The single-case study design for the emu study was ABCD, where A represents the baseline phase and B, C, and D represent force platform data collections on different days. The data are unadjusted for velocity. The GRF during walking was consistent with a bipedal gait (PVF greater than the emu's body weight for each step). Inspection of the means and SD values suggests that there may be differences among the group means; however, evaluation of velocity-adjusted data by use of a randomization test revealed no significant ($P > 0.05$) difference in the means for PVF of vertical impulse.

The sample lag-1 autocorrelations for all the phases were small and not significantly different from 0. To obtain an indication of the amount of variation attributable to day, we used an ANOVA table to calculate the percentage of the total error attributable to the phase for PVF values. The phase effect accounted for 16.5% of the total variation, or in other words, the corresponding Pearson correlation was 0.165, which was not significantly ($P > 0.05$) different from zero.

Discussion of the single-case emu study—The GRF values determined for the healthy emu used in this single-case study did not have variation attributable to the day effect that was significantly different from zero. Although a generalization of findings in emus is not possible from single-case data, the single-case analysis provided inferential information showing that the changes in GRF among days were both inconsequential and consistent with chance.

The lag-1 autocorrelations were small. This does not imply that the trials were serially independent and that classical statistical tests are acceptable, but does obviate the need to model data dependence. However, a lame emu may have more serial dependence as its gait deteriorates or improves over the course of a force platform session.

The baseline phase seemed to cover a wide range of variation and adequately described the baseline status of the emu. Collecting more data in the baseline phase would not further describe the GRF in this emu. It is possible that, following a surgical procedure, an emu would have a within-phase trend in GRF. In that situation, the phase mean would not be representative of the phase and use of another gait parameter (eg, slope over time) may be better suited to describe the data. The randomization test, however, would still be appropriate.

Inspection of the plot of emu PVF by trial suggested that variances were different among the phases and that the underlying data distributions may have been different over days. For this study, there was no scientific reason for that conjecture and there was no pattern to the variances (eg, an increase in variance over time). The study addressed variation of the phase means, but interpretation of the phase means themselves when the variances are different must be guarded. The different variances, combined with the different number of trials in each phase, again illustrate the importance of using a randomization test rather than an ANOVA. The assumptions made to utilize an ANOVA are violated under these conditions.¹¹

Also, there was no apparent time trend in the phase means that was indicative of the emu learning to use the force platform. This emu was trained prior to the start of the study, and it was expected to be comfortable with the experimental protocol. Untrained emus would probably require a training phase.

The number of observations varied among the phases, and ideally, the sampling interval and number of samples would have been predetermined. However, the generally erratic behavior of an emu is not conducive to an organized experiment; sometimes phases were terminated because of emu or handler fatigue, and sometimes conditions favored collection of additional data. However, phases were never terminated

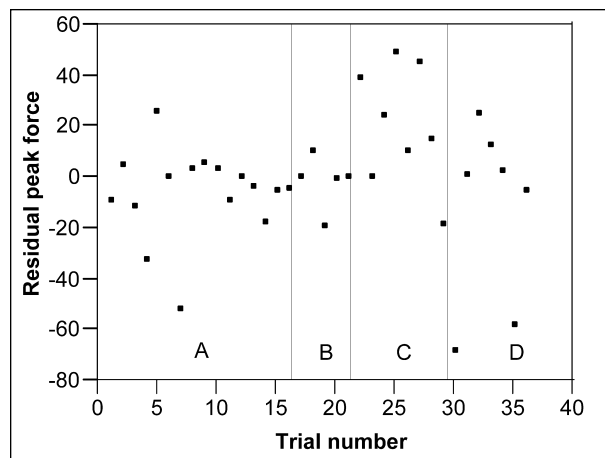


Figure 4—Plot of PVF (N/kg as a percentage of body weight, adjusted for velocity) by trial number in serial order during each phase of a single-case gait analysis study involving 1 emu. The vertical lines demarcate the phases A, B, C, and D, where A represents the baseline phase and B, C, and D represent force platform data collections on different days. The data do not appear visually to have trends that are a function of time or phase. The variation among phases is different, but equal variation among groups is not an assumption of a randomization test. However, by use of adjusted data, there was no significant ($P > 0.05$) difference among phase group means.

after examination of data and identification of trends, and no biases were introduced.

General Discussion of Single-case Study Designs

Single-case study designs have an important place in veterinary medical research because they are a formal mechanism by which inferences about treatments can be obtained when large sample sizes are difficult to achieve. They have advantages, compared with case studies. Single-case designs are quantitative and reveal patterns of outcomes associated with implementation and cessation of interventions. Evaluation of these patterns can reveal features of the disease process and, more importantly, provide inferences about the effect of interventions on that disease process. Via statistical analyses, P values can be calculated and used to ascertain whether changes among the phases (baseline, intervention, and follow-up) are due to chance or the disease process. However, results of single-case analysis cannot be generalized to a larger population of animals and should not be used as a surrogate for a classical study. It is an individual veterinarian's decision, based on experience, to use the results of a single-case study to direct treatment of another animal in a clinical setting.

There are several generalizations applicable to the single-case experiment concept. If several similar single-case experiments are available, then results can be combined into a single study by use of meta-analysis (in which summary statistics are combined, rather than the data being analyzed as a collection). That method is also useful when an investigator has more than 1 subject but too few for an experiment of classical statistical design. A single-case experiment can be conducted for each subject separately and then the results combined via meta-analysis to increase the

power of the study. That approach could have been used in the emu study described previously if all 5 emus had been available for data collection.

Alternatively, several subjects can be assessed individually in studies with single-case designs and the collection of data analyzed together. When the course of the disease is of interest, this approach permits univariate analyses of time or subject effects as well as the effect of treatment and the time-by-treatment interaction. These data are often analyzed with linear models, but there are other model-free analytical techniques for use in that situation.

There are other statistical designs that can be used when the sample size is small. A classic design is the balanced incomplete block design. Blocking is a technique that “provides local control of the environment to reduce experimental error.”¹² Common blocking criteria are proximity, physical characteristics, time, and management of tasks.¹² The result is that subjects are more or less uniform within a block. The term “incomplete” is derived from the fact that there are more treatments than subjects in the study so that each block is missing at least 1 treatment. For example, a balanced incomplete block design could be used to examine the growth effect of 4 different feed supplements in 3 pigs. There would be 4 blocks (treatment times) so that each pig receives each supplement, and each pair of supplements is provided in the same number of blocks. This method works well when the treatments can be applied in any order and analyzed with mixed-effects models.

Collections of subjects that form entities can be considered single subjects if they are treated and evaluated as a unit. For example, a veterinary practice can be considered a single entity and management or protocol changes within the practice can be evaluated in a single-case study.

Single-case designs are most applicable to studies in which serial measurements are possible before, during, and after interventions. The ABA design would be most appropriate when a treatment can be applied and withdrawn. Examples of ABA studies are drug efficacy evaluations, in which the drug can be withdrawn after the intervention phase. It is important to note that the ABA study design does not imply that the subject will revert to baseline outcome levels

after the intervention is withdrawn. Some treatments, such as surgery, are difficult or impossible to withdraw so that the AB design is appropriate in studies involving such procedures.

One of the original motivations for the development of single-case studies was to develop a mechanism to monitor and evaluate the clinical practice of common procedures in the social sciences. It is also difficult for veterinarians to evaluate professional practice, and single-case designs can be used to compare and evaluate treatments. Procedures may appear to be successful, but it is important to verify that they are better (eg, more efficacious, faster, or safer) than alternatives in the context of a real practice.

-
- a. Reinisch D, Conzemius M. The effect of human interaction on behavioral adaptations in the emu chick. Poster presentation, NIH summer scholars program, 2003.
 - b. S-Plus, Mathsoft Inc, Seattle, Wash.
-

References

1. Howe M. Casework self-evaluation: a single-subject approach. *Soc Serv Rev* 1974;48:1–23.
2. Ayllon T, Micheal J. The psychiatric nurse as behavioral engineer. *J Exp Anal Behav* 1959;2:323–334.
3. Tripodi T. *A primer on single subject design for clinical social workers*. Washington, DC: NASW Press, 1994.
4. Sackett DL, Straus S, Richardson S, et al. *Evidence-based medicine: how to practice and teach EBM*. 2nd ed. New York: Churchill Livingstone Inc, 2000.
5. Todman J, Dugard P. *Single-case and small-n experimental designs: a practical guide to randomization tests*. Mahwah, NJ: Lawrence Erlbaum, 2001.
6. Twisk J. *Applied longitudinal data analysis for epidemiology*. Cambridge, UK: Cambridge University Press, 2003.
7. Franklin R, Allison D, Gorman B. *Design and analysis of single-case research*. Mahwah, NJ: Lawrence Erlbaum, 1996.
8. Conzemius MG, Brown TD, Zhang Y, et al. A new animal model of femoral head osteonecrosis: one that progresses to human-like mechanical failure. *J Orthop Res* 2002;20:303–309.
9. Rumph PF, Steiss JE, West MS, et al. Interday variation in vertical ground reaction force in clinically normal Greyhounds at the trot. *Am J Vet Res* 1999;60:679–683.
10. Conzemius MG, Evans RB, Besancon MF, et al. Effect of surgical technique on limb function after surgery for rupture of the cranial cruciate ligament in dogs. *J Am Vet Med Assoc* 2005;226:232–236.
11. Ramsey F, Shafer D. *The statistical sleuth, a course in methods of data analysis*. Pacific Grove, Calif: Duxbury Press, 2002.
12. Kuehl RO. *Design of experiments: statistical principles of research design and analysis*. Pacific Grove, Calif: Duxbury Press, 2000.