Applied Animal Behaviour Science, 27 (1990) 1-7
Elsevier Science Publishers B.V., Amsterdam

Computer-assisted recording of live and videotaped horse behavior: reliability studies

Sue M. McDonnell and Nancy Kate Diehl
Hofmann Center for Reproductive Studies, University of Pennsylvania, School of Veterinary Medicine, New Bolton Center, Kennett Square, PA 19348 (U.S.A.)

(Accepted for publication 4 April 1990)

ABSTRACT

McDonnell, S.M., and Diehl, N.K., 1990, Computer-assisted recording of live and videotaped horse behavior: reliability studies, *Appl. Anim. Behav. Sci., 27:* 1-7.

Reliability of microcomputer event-recorder and video recording techniques for study of equine behavior was evaluated by Pearson correlation and percentage agreement procedures. Consistently high intra- and inter-observer correlations (most > 0.85), as well as acceptable percentage agreement (60-l 00%) on specific events, indicate that these techniques were reliable. Data recorded live or from video tape yielded similar results, These methods are easily applied to laboratory or field studies of equids.

INTRODUCTION

Computer-assisted recording of live and videotaped stallion sexual behavior is an important research technique in our laboratory. The hand-held microcomputer event recorder permits rapid, time-based entry of a large set of behavior observations in both laboratory and field situations. Video recording allows repeated review of trials, as an adjunct to live observation, as well as continuous observation of undisturbed, stabled horses.

While it might be expected that use of computer and video technology would improve accuracy and consistency of recording, reliability of results obtained using microcomputer event recorder and/or video recording techniques has not been directly addressed in equine behavior research. This study was designed to determine intra-observer reliability for stallion behavior data recorded live and from video, as well as inter-observer reliability of data recorded from video tape. Two behavioral situations were evaluated: (1) timed sexual behavior trials for which observations were recorded live or from video tape; and (2) 24-h continuous video observation of behavior of undisturbed stabled animals. For these two situations, respectively, two experienced ob-

0168-1591/90/$03.50

methods, event recorder, video

servers each recorded and derived multiple data sets for (1) 36 timed sexual behavior trials and (2) eight 6-h samples of video recorded stall behavior. Pearson correlation and percentage agreement procedures were used to estimate reliability of live and video recording of behavioral events, as well as reliability of derivation of the behavioral measures commonly used in our studies of sexual behavior.

TIMED SEXUAL BEHAVIOR TRIALS

Animals, materials and methods

The sexual behavior trials used in this reliability study were from an investigation of the effects of gonadotropin-releasing hormone (GnRH) on sexual behavior of pony geldings (McDonnell et al., 1989). Briefly, sexual behavior trials were conducted three times weekly for each of 12 'geldings. Each trial consisted of placing the test animal in a pen (3.05 X 3.05 m) for 4 min with access to a stimulus mare restrained along the outside of the pen, Sexual behavior responses were primarily recorded live by an experienced observer standing outside the test pen, using a hand-held, battery-operated computer event recorder (OS-3, Observational Systems, Seattle, WA). Each response was entered by the observer as a 2-keystroke code which the event recorder stores, along with time of entry (precise to a tenth of a second). Data were transferred from the event recorder to a desk-top computer or printer by RS-232 serial interface. For the experiment, the following specific behavioral frequency, latency, and duration measures were derived from a printed listing, of each event and its corresponding time of occurrence: sniff female, sniff ground, flehmcn response, vocalization, kick, roll, erection, and mount frequencies; attention, penis drop, erection, and mount latencies and durations. Frequency of a response was a simple count of the number of occurrences; latency to first occurrence of a response and total duration of a behavior were calculated to the nearest second.

The trials were also video recorded using a tripod-mounted video camera and recorder placed outside the pen at an angle that permitted continuous view of the subject and stimulus mare. From the entire experiment, 36 trials (three trials from each of 12 animals) were randomly selected to be reviewed for this reliability study. For these trials, two experienced observers each reviewed the video (played at real time) to record responses and derive behavioral measures as described above. The procedure was repeated l week later, In addition, for one set of 36 trials, each observer derived a data set from the other's printed record, Therefore, the following seven data sets were available for comparison. One of these observers was blind to the treatment or identity of individual animals. The other observer, who had recorded the live trials, was not blind to treatments, .

TABLE 1

Timed sexual behavior trials: Pearson correlation coefficient

	OBSA video 1 vs. video 2	OBS B video 1 vs. video 2	OBS A live vs. video 1	OBS A live vs. OBS B video 1	OBS A video 1 vs. OBS B video1	OBSA video 2 vs. OBS B video 1	OBS B derived OBSA video 1 vs. OBS B video 1	OBS A derived OBS B video 1
Frequency								
Sniff female	0.95	0.95	0.88	0.90	0.92	0.97	1.00	**1.00**
Sniff ground	0.91	0.92	0.83	0.80	0.89	0.94	1.00	1.00
Flehmen	0.96	0.92	0.97	0.96	0.97	0.91	1.00	0.99
Vocalization	0.98	0.97	0.95	0.98	0.93	0.96	1.00	1.00
Kick	0.72	0.47	0.48	0.70	0.69	0.69	0.48	1.00
Roll	1.00	0.97	1.00	0.97	0.97	1.00	0.86	1.00
Erection	0.97	0.93	0.96	0.88	0.93	0.92	1.00	0.97
Mount	0.98	0.98	0.98	0.96	0.98	0.98	0.98	1.00
Latency								
Attention	0.90	0.79	0.80	0.84	0.83	0.93	1.00	0.98
Penis drop	1.00	0.99	0.68	0.67	0.99	1.00	1.00	0.99
Erection	1.00	0.94	1.00	0.92	0.92	0.97	1.00	1.00
Mount	1.00	0.89	1.00	1.00	1.00	0.91	0.47	1.00
Duration								
Attention	0.99	0.92	0.99	**0.97**	**0.95**	0.95	**0.99**	**0.98**
Penis drop	0.93	**0.94**	0.78	**0.75**	**0.92**	0.90	1.00	0.95
Erection	0.99	0.95	0.99	0.98	0.99	0.95	**1.00**	**1.00**
Mount	0.98	0.94	0.94	**1.00**	0.93	0.99	**1.00**	0.95

All coefficients are significant ($P < 0.01, n = 36$).

OBS A live: Observer A recorded sexual behavior responses live during the original behavior study and derived behavioral measures from the time-based printed record.

OBS A video 1: Observer A recorded sexual behavior responses from the videotapes of 36 trials (one pass at real time) and derived behavioral measures from the time-based printed record.

OBS A video 2: Observer A recorded responses from videotaped trials again, 1 week later, and derived behavioral measures.

OBS B video 1: Observer B recorded responses from 36 videotaped trials and derived behavioral measures.

OBS B video 2: Observer B recorded responses from video tape again, 1 week later, and derived behavioral measures.

OBS B derive OBS A
OBS A derive OBS B: One observer derived behavioral measures from the other observer's first printed record from video-taped trials.

A microcomputer statistical analysis program (Statistix, NH Analytical Software, St. Paul, MN) was used to compute a Pearson product-moment correlation matrix (Pearson r) for all data sets for each sexual behavioral measure. All r values were tested for significance with the Student's t-test ($P<0.05$).

Results

Pearson correlation coefficients are shown in Table 1. These represent, for all behavioral measures derived, the meaningful comparisons of data sets. All correlation coefficients were statistically significant.

CONTINUOUS OBSERVATION OF BEHAVIOR IN A STALL

Animals, materials and methods

Video recording is useful for obtaining data concerning activities of undisturbed animals. To assess inter-observer reliability of observations from video recordings, we randomly selected a set of eight 6-h video tapes from a series of studies of spontaneous erection and masturbation in stabled horses and ponies (McDonnell, 1990). The video tapes selected represented 13 animals filmed in the situations in which we routinely use video recording, including different subject types (pony, horse, stallion, or gelding), housing (tie stalls or box stalls), video recording arrangement (1, 2, or 3 animats in view), and time of day, In those studies, 24-h stall activities were video recorded using a low light rated (1-6 lx) wide angle lens which permitted continuous view of the subject. Supplemental stall lighting was used 24 h per day. Video tapes were scanned by a trained observer at 20× real time to identify occurrence of each erection episode. These episodes were then viewed at real time for the duration of penis drop, Behavioral events were recorded using the event recorder described above. Behaviors recorded included erection, masturbation episodes, three types of penis movements, pelvic thrusts, occurrence of glans penis erection and flare, ejaculation, beginning and ending of sternal and lateral recumbency, activities preceding and following spontaneous erection, and apparent level of alertness during spontaneous erection and masturbation. Frequency, duration and latency behavioral measures were derived from the computer event recorder printed record. For the present study, a second observer independently viewed the video tapes, as well as recorded and derived

the behavioral measures. Observers were blind to each other's results and to any treatments of the animals.

Pearson correlation and percentage agreement of data from the two observers were calculated for a representative subset of the behavioral measures listed above. Pearson correlation coefficients were calculated for total erection and masturbation frequencies and durations; erection and masturbation episode durations; penis movement and pelvic thrust frequencies per episode; and interval from last period of recumbency to erection for erections that occurred within 3 min of standing, A percentage agreement of the two observers on the occurrence of each specific erection, masturbation, penis flare, and period of recumbency per 6-h sample was calculated as:

$$\frac{\text{Number detected by both OBS A and OBS B}}{\text{Total detected: by OBS A only+ OBS B only+ by both OBS A and OBS B}}$$

Calculated in this manner, percentage agreement is a more conservative estimate of reliability than the traditional percentage agreement based on checksheet data where non-occurrence of an event is also recorded and included in the analysis (Martin and Bateson, 1986, pp. 86-97).

TABLE 2

Continuous observation of behavior in a stall

		r	No.
A. Pearson correlation coefficients (OBS A vs. OBS B)			
Erection frequency per 6-h sample		0.95	13
Total erection duration per 6-h sample		0.87	10
Erection duration per episode		0.62	39
Masturbation frequency per 6-h sample		0.94	13
Total masturbation duration per 6-h sample		0.90	13
Masturbation duration per episode		0.67	26
Total number of penis movements per episode		0.95	27
Without belly contact		0.68	27
With bounce against belly		0.89	27
With press against belly		0.73	27
Pelvic thrust frequency per episode		1.00	27
Interval from recumbency to erection		0.95	9
B. Percentage agreement			
Erection episode	77		
Masturbation episode	67		
Penis flare occurrence	61		
Recumbency episode	100		

All coefficients are significant ($P < 0.01$).

Results

Table 2 summarizes Pearson correlation and percentage agreement results. Reliability estimates were extremely high ($r = 0.87–1.00$) for erection and masturbation frequencies, total erection and masturbation durations, total number of penis movements per episode, number of penis bounces per episode, frequency of pelvic thrusts per episode and interval from recumbency to erection. For erection and masturbation episode durations, as well as for the remaining two types of penis movements, reliability estimates were lower, but marginally acceptable ($r = 0.62–0.73$). Agreement of the two observers on specific events ranged from 61 to 100%.

DISCUSSION

These results indicate high intra- and inter-observer reliability of behavioral data obtained using a hand-held computerized event recorder from live or videotaped behavior trials, as well as from videotaped behavior of animals in stalls. In the majority of our behavior studies, we measure relative differences among groups associated with treatment. Therefore, the Pearson correlation coefficient adequately estimates relative agreement between observers and across various types of observational situations. In some situations, as in our studies of stall behavior, the actual temporal pattern and association of specific events are also of interest, The percentage agreement procedure, as we conservatively applied it here, suggests lower agreement (61–71%) on specific erection and masturbation events detected by rapid scanning of video tape. Closer examination of the data revealed that the greatest disagreement occurred when 3 animals were simultaneously in view on one video tape. On tapes with 1 or 2 animals in view, percentage agreement increased from 77 to 88% for erection occurrence, 67 to 81% for masturbation, and 61 to 71% for penis flare. In general this would suggest that more than 2 animals in view on one tape should be avaided. However, in situations where this is unavoidable, multiple or slower scannings would no doubt improve results. The 100% agreement on occurrence of recumbency suggests very high reliability for conspicuous long-term behavioral events using the same rapid scanning method, even when 3 animals are in view on the same tape,

In the sexual behavior trials, we expected reliability estimates to be higher with discrete, easily viewed sexual behavior responses (e.g. mount), than with the more inconspicuous responses or those more subject to an obstructed view on video tape (e.g. penis drop, specific types of penis movement) or to observer interpretation (e.g. attention to mare). However, no consistently lower correlation coefficients occurred within any cluster of responses. Kick frequency was one behavioral measure with lower correlation coefficients across all comparisons, possibly as a result of its overall low frequency of occurrence

and its occurrence simultaneously with other responses which were inadvertently given higher priority for recording.

The single lowest correlation (0.47) involved mount latency data recorded by one observer and derived by the other. This result is probably due to the fact that in this experiment, mounts were a rare event. Alternatively, this result may be due to differences between these two observers in their systems for recording and deriving this behavioral measure, It is curious that when the recording and deriving roles of the two observers were reversed, correlation was 1 .O for the same behavioral measure.

We find the microcomputer event recorder has been simple to learn and use compared with previously employed stopwatch-checksheet or audio tape-recorder techniques. The rapid entry of a practically limitless number of responses and the portability of the event recorder have been important advantages for deriving precise time-based behavioral measures for many different types of field and laboratory studies. Intra- and inter-observer reliability of this technique indicates that different observers remain consistent through all phases of data acquisition: the definition, interpretation, and recording of observed events as well as derivation of data from the printed record.

The high reliability between data derived from both live and video recorded sexual behavior trials is evidence that the two observation techniques yield comparable results. Different situations and research protocols lend themselves better to the live or to the video recorded method and from this study we feel more confident that results are comparable. The video tape has proven beneficial as a permanent record of events which may be used to correct or retrieve additional information from that which was observed under live conditions.

ACKNOWLEDGMENTS

This work was sponsored by the Dorothy Russell Havemeyer Foundation Portions of these data were presented at the 1989 Annual Meeting of the Animal Behavior Society.

REFERENCES

Martin, P. and Bateson, P., 1986. Measuring Behavior. Cambridge University Press, New York, pp. 86-97.

McDonnell, S.M., 1990. Masturbation in equids. Proc. 1989 Annu. Meet. Am. Assoc. Equine Pract., Boston, MA, pp. 567480.

McDonnell, SM., Diehl, N.K., Garcia, M.C. and Kenney, R.M., 1989. Gonadotropin releasing hormone (GnRH) affects precopulatory behavior in testosterone-treated geldings, Physiol. Behav., 45(1): 145-149.