**Optimizing Behavioural Observations: A Comparative Approach to Simulated Sampling Methods**

James Edward Brereton1, Jonathan Tuke2 and Eduardo J Fernandez3

1University Centre Sparsholt, Westley Lane, Sparsholt, Winchester, Hampshire, SO21 2NF United Kingdom

2School of Mathematical Sciences, The University of Adelaide, SA 5005, Australia

3School of Animal and Veterinary Sciences, The University of Adelaide, SA 5005, Australia

\*Correspondence: +447748354279; [James.Brereton@sparsholt.ac.uk](mailto:James.Brereton@sparsholt.ac.uk)

Original Research paper

**Running title:** Optimizing behavioural observations

**Abstract:**

*Keywords:* continuous recording, pinpoint sampling, one-zero sampling, interval recording, behavioural measurement, sampling methods

**INTRODUCTION**

The measurement of behaviour is important for those involved in the scientific study of animal behaviour. Behavioural studies are used as a tool to measure captive animal welfare and are used more often than other welfare indicators such as glucocorticoid analysis (Fraser, 2009; Sands & Creel, 2004). For captive animals, behavioural research may also be used to investigate the prevalence of positive behaviours, such as foraging, or negative behaviours, such as stereotypies (Carlstead et al. 1991; Fernandez & Timberlake, 2008; Ward et al. 2018). Studies of behaviour are also frequently conducted for wild animal populations and to better understand natural history or investigate the impact of human disturbance (Lehner, 1998; Sand & Creel, 2004). Research on animal behaviour is now so well recognised that there are numerous journals dedicated to its study, for instance: *Animal Behaviour,* *Applied Animal Behaviour Science, and Ethology.*

The methods used in animal behavior research can be traced back to human studies. Scientists during the mid-twentieth Century often used a mixture of both human and animal models to answer questions in the field of behavioral psychology (Domjan, 2014; Pierce & Cheney, 2013). Based on the range of different techniques that were generated by earlier studies, Altmann (1974) summarised the behavioural research methods available. This paper became fundamentally important to those interested in behavioural research, and remains a keystone paper for researchers, with at least 16,100 citations, according to a search on Google Scholar (2020). Whilst other authors, such as Martin and Bateson (2007) further refined the behavioural methods and their definitions, Altmann’s work is still regularly cited.

Since this initial review of behavioural methods, some behavioural sampling techniques became increasingly popular in animal literature, whereas others are rarely used. Several behaviour measurement techniques have received criticism in terms of their repeatability (Bernstein, 1991). For example, ad libitum (qualitative) sampling may be useful for developing ethograms and for pilot studies but has methodological flaws with regards to its lack of standardisation (Martin & Bateson, 2007; Rhine & Ender, 1983). However*, ad libitum* sampling is still used in animal behaviour literature, with a review by Mann (1999) identifying that between 53% and 59% of cetacean studies published in *Marine Mammal Science* used this sampling technique.

Continuous recording, or focal sampling, is considered the gold standard for behavioural sampling, as this method records all occurrences of behaviour and their durations (Hämäläinen et al. 2016). In the past, this made continuous recording challenging for researchers, as an active animal that rapidly changed behaviour would have been difficult to observe and record (Tyler, 1979). Similarly, measurement of multiple animals using a continuous method would have been incredibly challenging to document accurately, hence why the method is considered synonymous with focal sampling of one individual (Altmann, 1974; Martin & Bateson, 2007). Use of modern technology has in part ameliorated some of these issues by allowing behaviour to be recorded and analysed later (Amato et al. 2013). However, continuous recording may remain a challenge, even with camera availability. As a result, several sampling methods have been developed to measure multiple animals at one time (scan sampling), as well in a non-continuous fashion.

The use of pinpoint sampling, also referred to as instantaneous or momentary time sampling, is a commonly used method for observational study (Fernandez, Kinley & Timberlake, 2019; Lehner, 1998; Stevens et al. 2013). With pinpoint sampling, one or more responses are recorded at preselected moments in time (e.g., every 15 s for an hour). The benefits of pinpoint sampling are that it is less intensive than continuous sampling, and therefore may be more feasible for researchers to conduct (Grenier et al. 1999; Martin & Bateson, 2007; Gilby et al. 2010). The methods are also more versatile, allowing researchers to make decisions as to how long intervals should be spaced. For example, some researchers might choose to use 15-second intervals, particularly when studying an active animal or when conducting observations of a key time period, such as when enrichment is provided (Fernandez & Timberlake, 2019). On the other hand, observers might choose to use much longer intervals, such as one-, two- or five-minute intervals when their subjects are inactive or if they are observing for long time periods (Shora et al. 2020; Teixeira et al. 2017). It has been noted by some authors that shorter intervals tend to result in behavioural values that match more closely the continuous behaviour scores (Pullins et al. 2017).

One-zero or interval sampling involves choosing specific intervals of time, like pinpoint sampling, but instead recording whether one or more responses occur (or conversely, do not occur) within that interval of time (Bailey & Burch, 2017; Bakeman & Quera, 2012; Lehner, 1998). While popular with both human and non-human primate research, one-zero sampling seems to receive less representation than pinpoint sampling in most animal behavior studies and has been criticised by previous researchers (Altmann, 1974; Rhine & Flanigon, 1978). However, one-zero sampling has some of the same benefits of instantaneous sampling in that interval length can be tailored in line with the requirements of the study. Additionally, one-zero sampling has the potential to collect more behaviours during a stated period, as multiple behaviours can be recorded during each interval (Altmann, 1974). Leger (1977) identified good agreement with continuous behaviour measures when using one-zero sampling at 15-second intervals for chimpanzees (*Pan troglodytes)*. Likewise, Rhine and Flanigon (1978) found similar levels of occurrence when comparing continuous, pinpoint, and one-zero sampling methods with a colony group of stumptail macaques (*Macaca arctoides*). As noted above, one-zero (interval) sampling is also frequently used in studies on human behaviour, for example in the classroom (Dunkerton, 1981; Omark et al. 1976).

Both pinpoint and one-zero sampling overcome some of the issues associated with continuous recording by reducing the amount of input required by the researcher, while still aiming to keep the sample representative of the animal’s behavioural repertoire (Mitlöhner et al. 2001; Simpson & Simpson, 1977). However, one key question is how closely these techniques correlate with continuous recording? Pinpoint sampling is reported to lose information in terms of behaviour duration and is potentially less likely to pick up any behaviours of short duration (events) (Martin & Bateson, 2007; Xiao et al. 2005). By contrast, one-zero sampling is better at recording all observable behaviours, but both behavioural frequency and duration could be easily misrepresented: there is no way to identify whether a behaviour recorded as present for one interval was seen once or thirty times during that time period (Saibaba et al. 1996).

The following study proposes to compare simulated occurrence of both low, medium, and high frequency/duration behaviours, as well as similar observation periods for pinpoint and one-zero sampling methods. We hypothesized two results: (1) one-zero sampling would be better suited for detecting the occurrence of low frequency (event) behaviours, particularly when comparing less frequent pinpoint and one-zero observation methods (e.g., 500 second observation periods), and (2) pinpoint sampling would provide a more accurate representation of percentages of occurrence for both low, medium, and high duration (state) behaviours than one-zero sampling.

**METHODS**

For all simulations, continuous recording methods were generated for both frequency of occurrence and percentage of occurrence, with two different non-continuous sampling methods directly compared: pinpoint (instantaneous) and one-zero (interval) sampling. Independent variables were split into two categories: Response frequency (to measure the ability of both behaviour methods to detect short, event behaviours) and response duration (to measure the ability of the methods in assessing long-term, state behaviours). Three levels for response frequency and response duration were determined, based on an arbitrarily level of occurrence: 3 s, 30 s, and 300 s. The interval lengths for both pinpoint and one-zero sampling were set at 5 s, 50 s, and 500 s, in order to compare the effect of interval length on test accuracy.

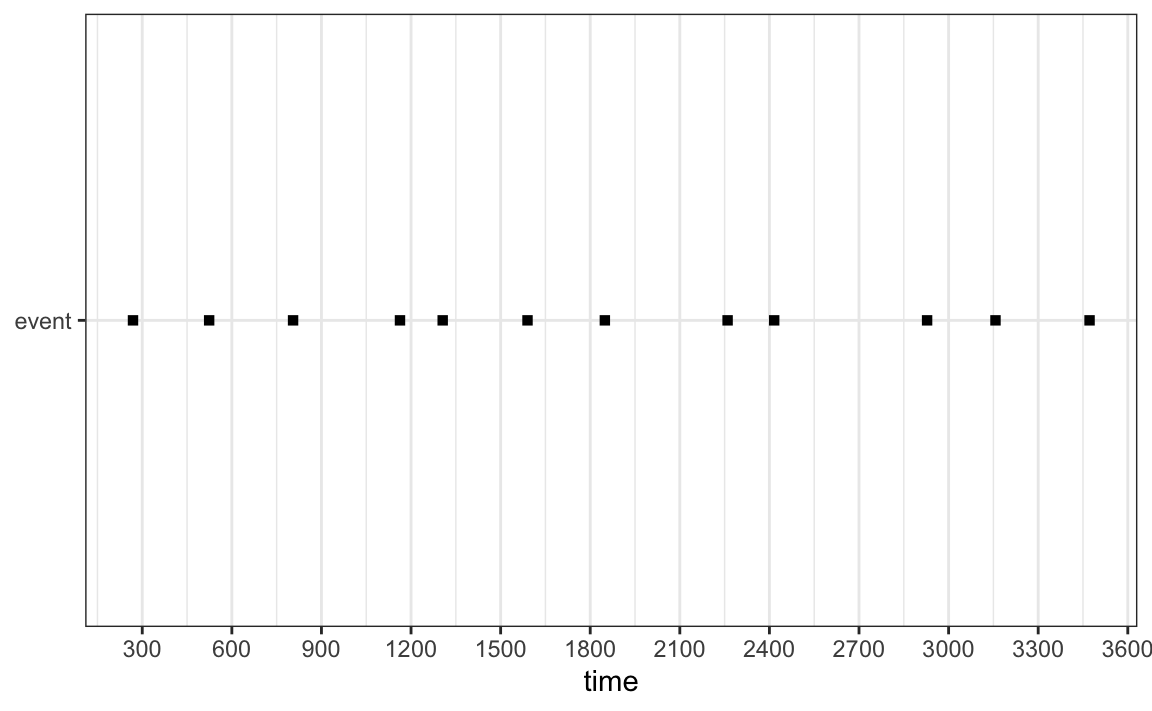
***Simulations***

All of the simulations were done in the R computing language using the GUI RStudio (code publicly available at https://github.com/jonotuke/animal\_simulation\_2020). For the response frequency simulations, for a frequency of every k seconds, we split the entire time period into b blocks of length k seconds. We then randomly selected one second within each block using a discrete uniform distribution and designated that as the time the event occurred. An example of a simulation with k = 300 is given in Figure 1.

For the response duration simulations, of duration d, we split the entire time period into 6 blocks, each of length 600 seconds. For each block, we then chose random number using a discrete uniform on the numbers from 1 to (600 – d), this was then designated as the start time of the event in the block, and the end time of the event was this time plus d. Each block had its own random starting time. An example of a simulation with d = 300 is given in Figure 2.

***Response frequency***

This simulation focused on the recording of event behaviours: behaviours of very short duration (Martin & Bateson, 2007). For the purpose of the simulation, the duration of all event behaviours was set to exactly one second. Next, three different frequencies of event behaviour were selected: high (3 s), medium (30 s) and low (300 s) frequency of occurrence. Simulated data sets for the observation period for all three behavioural frequencies. The observation period was 1 hour in length (3600 seconds). A total of 100 simulated data sets were generated for each of the three response frequencies. The exact time that each event occurred within the 3, 30 or 300 second period was randomised.



***Figure 1.*** *Example of simulated data set to show how each event was presented. This figure shows the location of each event when events were set to low frequency (occurs once per 300 seconds). The exact location of each event within its 300 second window was selected at random.*

The real (continuous) occurrence of each simulated response frequency was determined by calculating the number of seconds of each event that were possible in a simulated hour of data (observation period divided by frequency of occurrence; high frequency = 1200 s; medium frequency = 120 s; and low frequency = 12 s). The event behaviour seconds were then transformed into a percentage of total time (as is often shown in behaviour studies in the form of an activity budget), as well as frequency of occurrence.

To compare against this real (continuous) measurement, one-zero and pinpoint sampling were used on the simulated data sets. Three interval lengths (5 s, 50 s, and 500 s) were used for both pinpoint and one-zero sampling. This resulted in nine-hundred data sets (nine combinations of simulation parameters and sampling parameters, each combination simulated 100 times) being developed.

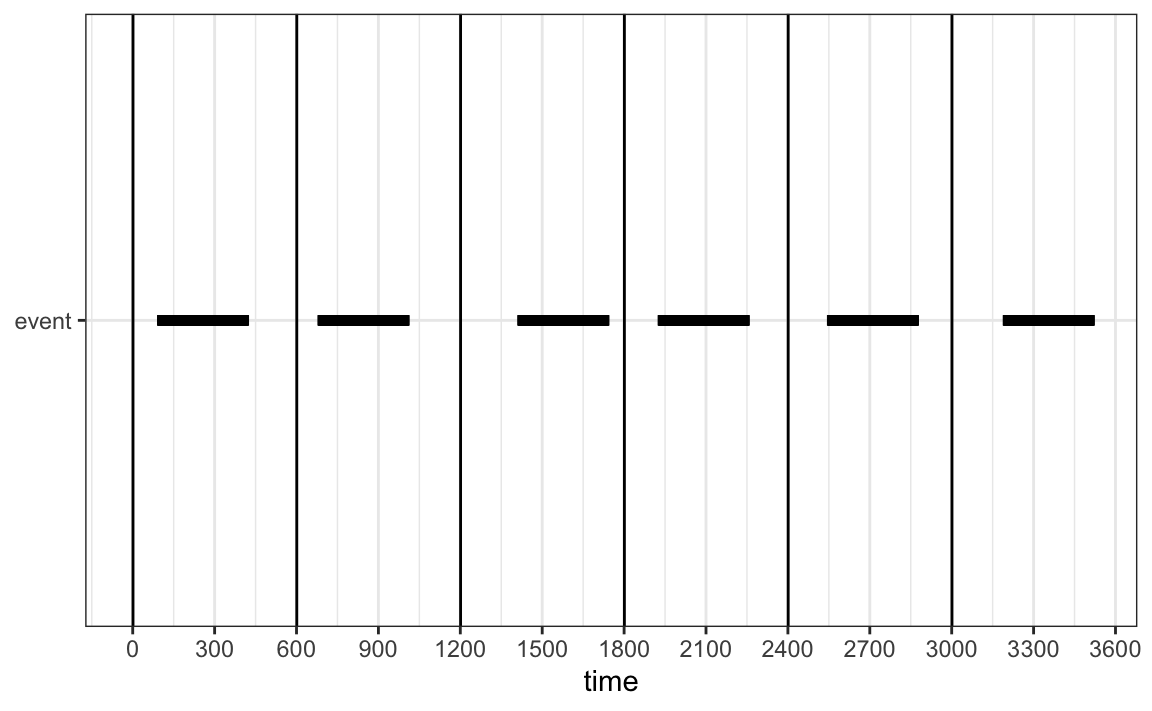
The data generated from the pinpoint and one-zero sampling was then converted into percentages to compare against the continuous data. The error rates for one-zero and pinpoint sampling were calculated for each of their three interval lengths.

***Response duration***

This simulation was developed for long-duration or state behaviours. In the literature, state behaviours can be of variable length, lasting anywhere from seconds to minutes or hours. In order to accommodate this, three levels of behavioural duration were selected. These durations were set as short (3 s), medium (30 s) and long (300 s) durations of occurrence. Each of these states were treated separately (only short, medium or long behaviours occurred in each simulation). As per the *Response frequency* investigation, the observation period was set to one hour in length (3600 seconds). Each behavioural duration simulation was repeated 100 times.

The chosen behaviour occurred once per 600 s period. The exact time that each behaviour occurred within its respective 600s period was selected at random (though the behaviour was not allowed to slip into the next period of 600 s). Continuous data sets were developed by using the raw, simulated data and transforming this into percentages. This meant that each behaviour occurred six times during each hour simulation, with the long duration occurring 50% of the hour, the medium state occurring 5%, and the short state occurring 0.5% of the time.

Each of the three behaviour durations (short, medium, and long) were measured using one-zero and pinpoint sampling. Three interval lengths, again consisting of 5 s, 50 s and 500 s, as had been selected for the *Response frequency* investigations. These interval lengths were used for both the one-zero and the pinpoint sampling. Once complete, the results were then transformed into percentages and compared to the continuous data to determine the level of error.



***Figure 2.*** *Example of simulated data for the response duration for long (300 s states). The location of each state was selected at random within its 600 s period. This results in the state occurring for exactly 50% of the hour simulation.*

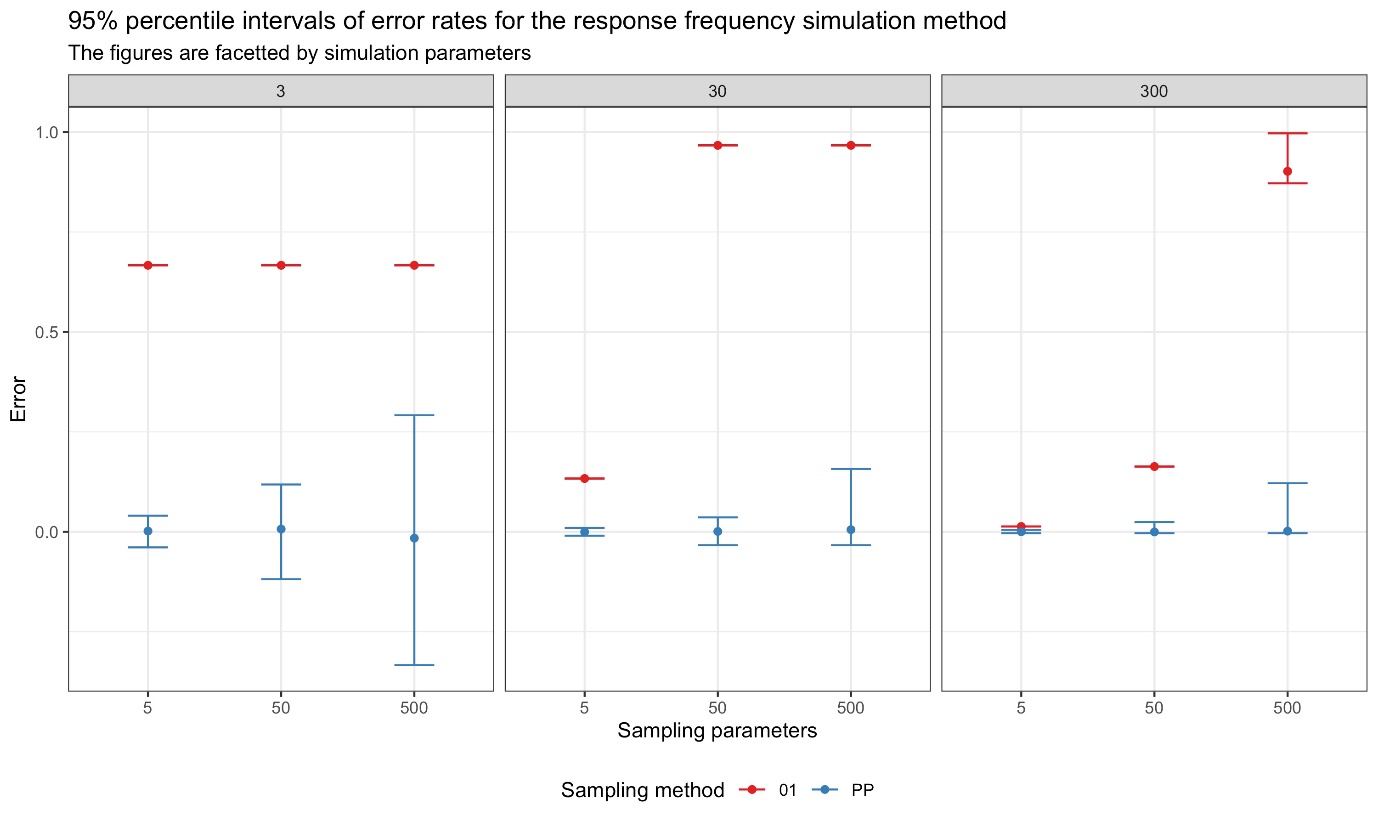
***Statistical Analysis***

Statistical analyses were conducted on the mean error scores for the one-zero and pinpoint sampling at each respective interval length. The Friedman test was used to investigate whether there was a statistically significant effect of sampling method on the estimation error. The sampling / simulation combination was used as a blocking factor. The non-parametric Friedman test was used due to the non-normality of the errors and the observed heteroscedascity.

**RESULTS**

***Response frequency***

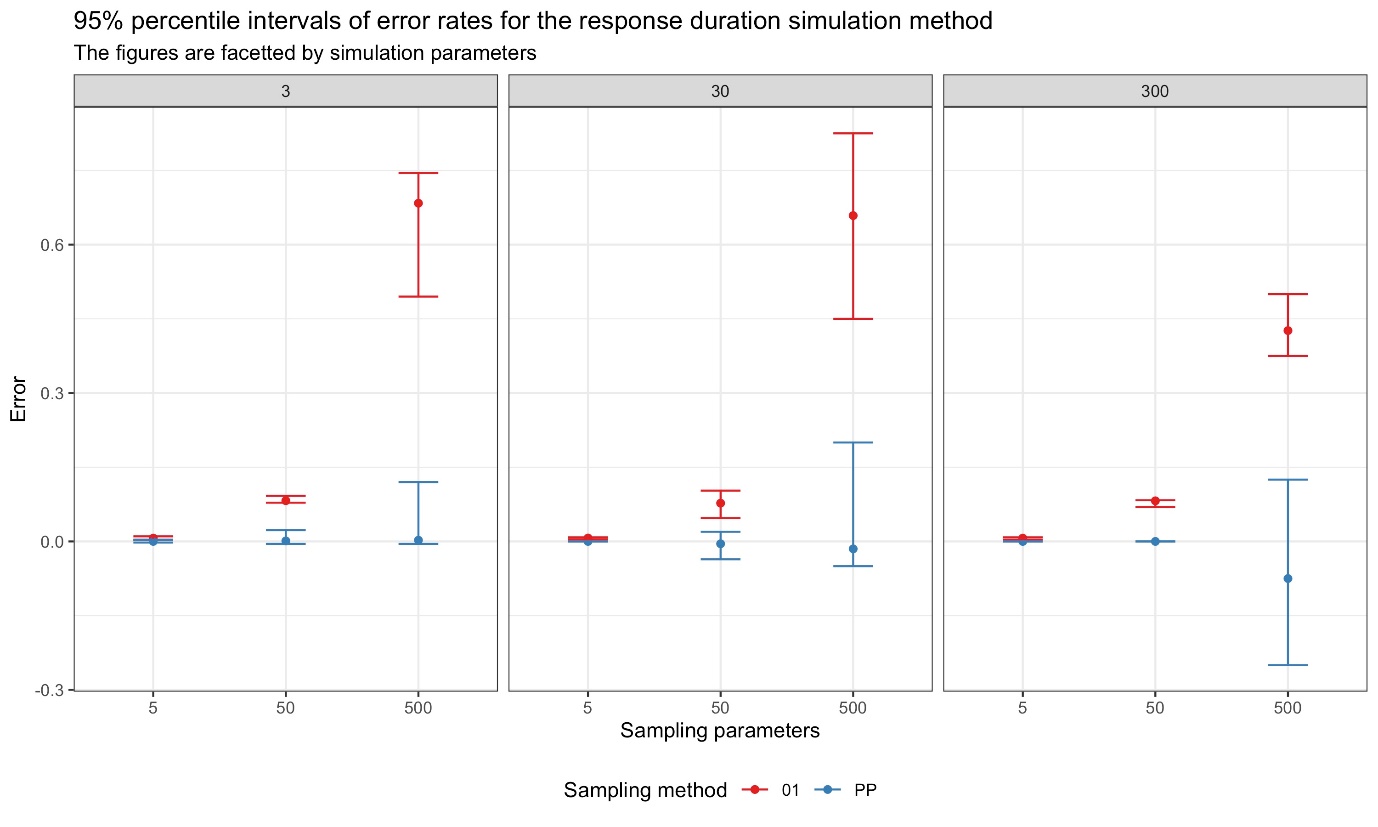
The accuracy of both one-zero and pinpoint sampling was calculated for each interval length and each of the three behavioural frequencies (Table 1 and Figure 3). Overall, mean error rates were consistently lower for the pinpoint sampling method in comparison to the one-zero sampling method. (we’ll give p-values here, as well as confidence intervals). For both behavioural sampling methods, error rates increased as the interval length increased, with the 500 s interval showing the largest error rates and variation for both one-zero and pinpoint sampling.



***Figure 3.*** *Mean error rates for the response frequency simulation and the pinpoint and one-zero sampling methods for the 5s, 50 s and 500 s observation intervals. Error rates were higher for the longer interval periods and were consistently higher for the one-zero sampling method. PP: Pinpoint sampling, 01: One-zero sampling.*

***Response duration***

The accuracy of both one-zero and pinpoint sampling was calculated for each interval length and all three behavioural durations (short, medium, and long) (Table 2 and Figure 4). For the short duration behaviours, one-zero sampling had a consistently high error rate. One-zero sampling was in fact accurate only when interval length was short (5 s) and the behavioural duration was medium or long. Pinpoint sampling error rates were generally lower than those from one-zero sampling, though variation in scores tended to increase as interval lengths became longer.



***Figure 4.*** *Mean error rates for the response duration simulation and the one-zero and pinpoint sampling methods for the 5 s, 50 s and 500 s observation intervals. There was more variation in error rates for the pinpoint sampling as interval length grew longer, but average mean error rates were still small. However, error rates for one-zero sampling increased as interval length increased. PP: Pinpoint sampling, 01: One-zero sampling.*

***Table 1.*** *Mean error rates for each sampling method under 5 s, 50 s and 500 interval lengths for the response frequency simulation.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Simulation parameters | Sampling parameters | Proportion of time event occurs | Mean error | Lower 95% percentile of error | Upper 95% percentile of error |
| **One-zero** |  |  |  |  |  |
| 3 | 5 | 0.333 | 0.667 | 0.667 | 0.667 |
| 3 | 50 | 0.333 | 0.667 | 0.667 | 0.667 |
| 3 | 500 | 0.333 | 0.667 | 0.667 | 0.667 |
| 30 | 5 | 0.033 | 0.133 | 0.133 | 0.133 |
| 30 | 50 | 0.033 | 0.967 | 0.967 | 0.967 |
| 30 | 500 | 0.033 | 0.967 | 0.967 | 0.967 |
| 300 | 5 | 0.003 | 0.013 | 0.0133 | 0.013 |
| 300 | 50 | 0.003 | 0.163 | 0.163 | 0.163 |
| 300 | 500 | 0.003 | 0.902 | 0.872 | 0.997 |
| **Pinpoint** |  |  |  |  |  |
| 3 | 5 | 0.333 | 0.002 | -0.039 | 0.0403 |
| 3 | 50 | 0.333 | 0.007 | -0.118 | 0.118 |
| 3 | 500 | 0.333 | -0.0158 | -0.003 | 0.292 |
| 30 | 5 | 0.033 | -0.0004 | -0.010 | 0.010 |
| 30 | 50 | 0.033 | 0.0010 | -0.033 | 0.036 |
| 30 | 500 | 0.033 | 0.005 | -0.033 | 0.157 |
| 300 | 5 | 0.003 | 0.0003 | -0.003 | 0.005 |
| 300 | 50 | 0.003 | -0.0003 | -0.003 | 0.024 |
| 300 | 500 | 0.003 | 0.0017 | -0.003 | 0.122 |

***Table 2.*** *Mean error rates for each sampling method under 5 s, 50 s and 500 interval lengths for the response duration simulation.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Simulation parameters | Sampling parameters | Proportion of time event occurs | Mean error | Lower 95% percentile of error | Upper 95% percentile of error |
| **One-zero** |  |  |  |  |  |
| 3 | 5 | 0.005 | 0.006 | 0.003 | 0.010 |
| 3 | 50 | 0.005 | 0.082 | 0.078 | 0.092 |
| 3 | 500 | 0.005 | 0.684 | 0.495 | 0.745 |
| 30 | 5 | 0.050 | 0.007 | 0.004 | 0.008 |
| 30 | 50 | 0.050 | 0.008 | 0.047 | 0.103 |
| 30 | 500 | 0.050 | 0.659 | 0.450 | 0.825 |
| 300 | 5 | 0.500 | 0.007 | 0.003 | 0.008 |
| 300 | 50 | 0.500 | 0.082 | 0.069 | 0.083 |
| 300 | 500 | 0.500 | 0.426 | 0.375 | 0.500 |
| **Pinpoint** |  |  |  |  |  |
| 3 | 5 | 0.005 | <0.001 | -0.002 | 0.003 |
| 3 | 50 | 0.005 | 0.001 | -0.005 | 0.023 |
| 3 | 500 | 0.005 | 0.002 | -0.005 | 0.120 |
| 30 | 5 | 0.050 | <0.001 | <0.001 | <0.001 |
| 30 | 50 | 0.050 | -0.004 | -0.036 | 0.019 |
| 30 | 500 | 0.050 | -0.015 | -0.050 | 0.200 |
| 300 | 5 | 0.500 | <0.001 | <0.001 | <0.001 |
| 300 | 50 | 0.500 | <0.001 | <0.001 | <0.001 |
| 300 | 500 | 0.500 | -0.075 | -0.250 | 0.125 |

**DISCUSSION**

Overall, our simulations did not conclusively show that one-zero sampling is better able to identify short duration behaviours, even when they occurred at low frequencies However the simulations did show that for behaviours of longer duration, pinpoint sampling tends to outperform one-zero sampling in terms of accuracy. For both sampling methods, increasing the interval length appeared to significantly reduce the accuracy of the test results (in comparison to the ‘actual’ or continuously recorded data.

***Previous studies***

Prior to this study, researchers have compared differences between pinpoint and one-zero sampling methods. Early simulations lacked the precision and/or ability to run extensive repetitions of their simulations to accurately assess sampling method differences (Griffin & Adams, 1983; Harrop & Daniels, 1986; Repp et al. 1976). Fewer repetitions of simulations may have resulted in larger error margins when comparing behavioural methods. Other researchers have attempted to make similar methodological comparisons via the data collection of actual behavioural occurrences (Gardenier et al. 2004; Leger, 1977; Murphy & Harrop, 1994; Rhine & Flanigon, 1978). While the results of differences in sampling methods for real occurrences of behaviour varied, caution should be used in making determinations of the validity of any result based on specific examples, as exceptions to any rule can and do occur.

**(I’ll have a second paragraph on the six studies here tomorrow).**

***Which behavioural method is most appropriate for my study?***

Pinpoint sampling has not been recommended for measuring frequency (event) responses, particularly those of low occurrence (Altmann, 1974; Lehner, 1998). However, in our simulation this method was accurately able to detect low occurrence (<1%) frequencies. Therefore, the use of pinpoint sampling to measure any event responses, regardless of their frequency of occurrence, appears to be a viable option. Similarly, one-zero sampling methods are often preferred as an observational method because of the ease with which behaviours can be observed, recorded, and assessed for Interobserver Agreement (IOA; Cooper et al. 2019; Poling et al. 1995). The same can also be said for pinpoint sampling, which provides an equally user-friendly research method when compared to continuous (focal) recordings. In addition, researchers attempting to account for under- or over-estimates of one-zero recordings have devised different sampling methods, including partial, whole, occurrence, and non-occurrence interval (one-zero) recordings. Still, the difficulty here is that, if pinpoint sampling provides a more accurate representation of behavioral occurrence, then the solution should be to adopt this method rather than adjust a less accurate one-zero recording method.

An added benefit of using either pinpoint or one-zero sampling methods over continuous recordings are they negate the difficulty in making comparisons between frequency (event) versus duration (state) behaviours. For instance, if a researcher were assessing the impact of pacing on the welfare of some animal, how would they compare 10 instances of 6-second paces to one instance of a 60-second pace? Lehner (1998) suggests that the former could be assessed as a bout of event responses, but it is still not clear how to evaluate the difference between a bout of responses to less frequent but longer duration behaviours. Pinpoint and one-zero sampling methods avoid this problem by only recording whether the response occurred during some observation period, regardless of the frequency or duration of the recorded response. This makes these observation methods valuable in circumstances where presence or absence of a particular behaviour is more important than the measurement of its frequency or duration, such as in studies of courtship or reproduction (Fraser, 2009).

***Sampling methods concluded***

Historically, a major factor in determining behavioural observation methodology has been the prevalence of that sampling method within some field/observational species. For instance, Mann (1999) found that over half of all cetacean studies used *ad lib* sampling, even though such sampling methods are recognized to be both less quantitative and systematic. Likewise, one-zero sampling methods are typically used by primatologists and behaviour analysts for the study of non-human primate and human behaviour, respectively (Cooper et al 2019; Doran, 1992; Merrell 2001; Omark, 1976; Seyfarth et al. 1977; Rhine et al. 1985). The concept of using methodology passed down from previous studies and labs has been referred to as “laboratory lore” and is an asset to the cultural transmission of scientific knowledge (Buskist & Johnston, 1988; Johnston & Pennypacker, 2010). Nonetheless, the selection of behavioral observation methods, like all aspects of scientific research, should be based on the efficacy of the methodology used. In the case of selecting between pinpoint or one-zero sampling methods to estimate behavioural occurrences, our study indicates that pinpoint sampling outperforms one-zero sampling on all frequency and duration measures simulated.

Thus, laboratory lore aside, pinpoint sampling seems to be the better option for measuring some aspect of behavioral prevalence when compared to one-zero sampling methods.

***The effects of interval length***

For both one-zero and pinpoint sampling, increasing the interval length reduced the accuracy of data collected. Increasing interval length appeared to affect the accuracy and variance of one-zero more than it affected pinpoint sampling, though both sampling methods were affects. The most accurate behaviour method and interval length combination was the pinpoint, 5 second sampling technique, when looking at response duration (though one-zero was also relatively accurate at this interval). It is possible that variance levels could be reduced even further with a shorter interval length, but this is likely to be impractical for researchers in the field.

The 50-second interval length still produced low error levels for response duration, regardless as to whether the behaviour was short, medium or long. At this interval length, however, error levels already appear quite pronounced. This may suggest that if a medium or longer-interval length is required for a project, pinpoint sampling is more appropriate as it is more likely to represent the actual animal behaviour seen.

However, consideration must also be paid to the error levels, which increase considerably for pinpoint sampling alongside interval length (Murphy & Harrop, 1994). This suggests that the data recorded could deviate considerably from the actual data, particularly in studies where minimal data has been collected. In this particular study, 100 replications were conducted for each test, which is equivalent to 100 hours of observations. Error levels, had only a couple of simulations been conducted, could result in data that does not meaningfully represent the actual data.

***Future directions***

Our simulations (and the subsequent assessment of error levels and variance) could be used to guide methodological design for future research projects. For example, once a pilot study or previous research have been consulted, a researcher might know the average duration of a behaviour of interest. This average behavioural duration could be mapped against the behavioural duration data (FigureXXX), to identify how much error might occur in their project depending on the interval length they choose. For a species that rapidly switches between behaviours (.e.g. cotton-top marmoset (*Saguinus oedipus),* this may result in the researcher picking very short interval lengths, such as 20 seconds (Hämäläinen et al. 2016). For animals that spend more time inactive, they may lose less behavioural data when picking the longer interval lengths (e.g. royal python (*Python regius)*.

**Acknowledgements**

The authors would like to thank Mrs S Brereton for proof reading. For further information about this research, the authors can be contacted at James.Brereton@sparsholt.ac.uk

**References**

Altmann, J. (1974). Observational study of behavior: sampling methods. *Behaviour*, *49*(3-4), 227-266. https://doi.org/10.1163/156853974X00534

Amato, K. R., Van Belle, S., & Wilkinson, B. (2013). A comparison of scan and focal sampling for the description of wild primate activity, diet and intragroup spatial relationships. *Folia Primatologica*, *84*(2), 87-101. https://doi.org/10.1159/000348305

Bailey, J. S., & Burch, M. R. (2017). *Research methods in applied behavior analysis*. Routledge.

Bakeman, R., & Quera, V. (2012). *Behavioral observation.* In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbooks in psychology®. APA handbook of research methods in psychology, Vol. 1. Foundations, planning, measures, and psychometrics* (p. 207–225). American Psychological Association. https://doi.org/10.1037/13619-013

Bernstein, I. S. (1991). An empirical comparison of focal and ad libitum scoring with commentary on instantaneous scans, all occurrence and one-zero techniques. *Animal Behaviour*, *42*(5), 721-728. https://doi.org/10.1016/S0003-3472(05)80118-6

Buskist, W., & Johnston, J. M. (1988). Laboratory lore and research practices in the experimental analysis of human behavior. *The Behavior Analyst*, *11*(1), 41-42.

Carlstead, K., Seidensticker, J., & Baldwin, R. (1991). Environmental enrichment for zoo bears. *Zoo Biology*, *10*(1), 3-16. https://doi.org/10.1002/zoo.1430100103|

Devine, S. L., Rapp, J. T., Testa, J. R., Henrickson, M. L., & Schnerch, G. (2011). Detecting changes in simulated events using partial‐interval recording and momentary time sampling III: Evaluating sensitivity as a function of session length. *Behavioral Interventions*, *26*(2), 103-124.

Domjan, M. (2014). *The principles of learning and behavior*. Nelson Education.

Doran, D. M. (1992). Comparison of instantaneous and locomotor bout sampling methods: a case study of adult male chimpanzee locomotor behavior and substrate use. *American Journal of Physical Anthropology*, *89*(1), 85-99. https://doi.org/10.1002/ajpa.1330890108|

Dunkerton, J. (1981). Should classroom observation be quantitative? *Educational Research*, *23*(2), 144-151. https://doi.org/10.1080/0013188810230208

Fernandez, E. J., Kinley, R. C., & Timberlake, W. (2019). Training penguins to interact with enrichment devices for lasting effects. *Zoo Biology*, *38*(6), 4-489. https://doi.org/10.1002/zoo.21510

Fernandez, E. J., & Timberlake, W. (2008). Mutual benefits of research collaborations between zoos and academic institutions. *Zoo Biology*, *27*(6), 470-487. https://doi.org/10.1002/zoo.20215

Fernandez, E. J., & Timberlake, W. (2019). Foraging devices as enrichment in captive walruses (*Odobenus rosmarus*). *Behavioural Processes*, *168*, 103943. https://doi.org/10.1016/j.beproc.2019.103943

Fraser, D. (2009). Animal behaviour, animal welfare and the scientific study of affect. *Applied Animal Behaviour Science*, *118*(3-4), 108-117. https://doi.org/10.1016/j.applanim.2009.02.020

Gardenier, N. C., MacDonald, R., & Green, G. (2004). Comparison of direct observational methods for measuring stereotypic behavior in children with autism spectrum disorders. *Research in Developmental Disabilities*, *25*(2), 99-118.

Gilby, I. C., Pokempner, A. A., & Wrangham, R. W. (2010). A direct comparison of scan and focal sampling methods for measuring wild chimpanzee feeding behaviour. *Folia Primatologica*, *81*(5), 254-264. https://doi.org/10.1159/000322354

Google Scholar. (2020). *Observational study of behavior: sampling methods.* Retrieved 11 November, 2020, from https://scholar.google.com/scholar?cites=15059125966377192598&as\_sdt=2005&sciodt=0,5&hl=en

Gouzoules, S., Gouzoules, H., & Marler, P. (1984). Rhesus monkey (*Macaca mulatta*) screams: representational signalling in the recruitment of agonistic aid. *Animal Behaviour*, *32*(1), 182-193. https://doi.org/10.1016/S0003-3472(84)80336-X

Grenier, D., Barrette, C., & Crête, M. (1999). Food access by white-tailed deer (*Odocoileus virginianus*) at winter feeding sites in eastern Québec. *Applied Animal Behaviour Science*, *63*(4), 323-337. https://doi.org/10.1016/S0168-1591(99)00017-9

Griffin, B., & Adams, R. (1983). A parametric model for estimating prevalence, incidence, and mean bout duration from point sampling. *American journal of Primatology*, *4*(3), 261-271.

Hämäläinen, W., Ruuska, S., Kokkonen, T., Orkola, S., & Mononen, J. (2016). Measuring behaviour accurately with instantaneous sampling: A new tool for selecting appropriate sampling intervals. *Applied Animal Behaviour Science*, *180*, 166-173. https://doi.org/10.1016/j.applanim.2016.04.006

Harrop, A., & Daniels, M. (1986). Methods of time sampling: A reappraisal of momentary time sampling and partial interval recording. *Journal of applied behavior analysis*, *19*(1), 73-77.

Hartsock, T. G., & Barczewski, R. A. (1997). Prepartum behavior in swine: effects of pen size. *Journal of Animal Science*, *75*(11), 2899-2904. https://doi.org/10.1016/S0168-1591(99)00017-9

Jauhiainen, L., & Korhonen, H. T. (2005). Optimal behaviour sampling and autocorrelation curve: modelling data of farmed foxes. *Acta Ethologica*, *8*(1), 13-21. https://doi.org/10.1007/s10211-004-0105-1

Johnston, J. M., & Pennypacker, H. S. (2010). *Strategies and Tactics of Behavioral*

*Research*. Routledge.

Leger, D. W. (1977). An empirical evaluation of instantaneous and one-zero sampling of chimpanzee behavior. *Primates*, *18*(2), 387-393. https://doi.org/10.1007/BF02383116

Lehner, P. N. (1998). *Handbook of ethological methods*. Cambridge University Press.

Mann, J. (1999). Behavioral sampling methods for cetaceans: a review and critique. *Marine Mammal Science*, *15*(1), 102-122. https://doi.org/10.1111/j.1748-7692.1999.tb00784.x|

Martin, P., & Bateson, P. (2007). *Recording methods. In ‘Measuring Behaviour: An Introductory Guide’*. Cambridge University Press.

Meany‐Daboul, M. G., Roscoe, E. M., Bourret, J. C., & Ahearn, W. H. (2007). A comparison of momentary time sampling and partial‐interval recording for evaluating functional relations. *Journal of applied behavior analysis*, *40*(3), 501-514.

Merrell, K. (2001). Assessment of children’s social skills: Recent developments, best practices, and new directions. Exceptionality, 9, 3–18.

Mitlöhner, F. M., Morrow-Tesch, J. L., Wilson, S. C., Dailey, J. W., & McGlone, J. J. (2001). Behavioral sampling techniques for feedlot cattle. *Journal of Animal Science*, *79*(5), 1189-1193. https://doi.org/10.1016/S0168-1591(99)00017-9

Murphy, M. J., & Harrop, A. (1994). Observer error in the use of momentary time sampling and partial interval recording. *British Journal of Psychology*, 85(2), 169-179. https://doi.org/10.1111/j.2044-8295.1994.tb02517.x

Omark, D. R., Fiedler, M. L., & Marvin, R. S. (1976). Dominance hierarchies: observational techniques applied to the study of children at play. *Instructional Science*, *5*(4), 403-423. https://doi.org/10.1007/BF00051807

Pierce, W. D., & Cheney, C. D. (2013). *Behavior analysis and learning*. Psychology Press.

Poling, A., Methot, L. L., & LeSage, M. G. (1995). *Fundamentals of behavior analytic research*. Springer Science & Business Media.

Pullin, A. N., Pairis-Garcia, M. D., Campbell, B. J., Campler, M. R., & Proudfoot, K. L. (2017). Instantaneous sampling intervals validated from continuous video observation for behavioral recording of feedlot lambs. *Journal of Animal Science*, *95*(11), 4703-4707. https://doi.org/10.2527/jas2017.1835

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Radley, K. C., O'Handley, R. D., & Labrot, Z. C. (2015). A comparison of momentary time sampling and partial‐interval recording for assessment of effects of social skills training. *Psychology in the Schools*, *52*(4), 363-378.

Rapp, J. T., Colby, A. M., Vollmer, T. R., Roane, H. S., Lomas, J., & Britton, L. N. (2007). Interval recording for duration events: a re‐evaluation. *Behavioral Interventions: Theory & Practice in Residential & Community‐Based Clinical Programs*, *22*(4), 319-345.

Rapp, J. T., Colby‐Dirksen, A. M., Michalski, D. N., Carroll, R. A., & Lindenberg, A. M. (2008). Detecting changes in simulated events using partial‐interval recording and momentary time sampling. *Behavioral Interventions: Theory & Practice in Residential & Community‐Based Clinical Programs*, *23*(4), 237-269.

Repp, A. C., Roberts, D. M., Slack, D. J., Repp, C. F., & Berkler, M. S. (1976). A comparison of frequency, interval, and time‐sampling methods of data collection. *Journal of Applied Behavior Analysis*, *9*(4), 501-508. https://doi.org/10.1901/jaba.1976.9-501

Rhine, R. J., & Ender, P. B. (1983). Comparability of methods used in the sampling of primate behavior. *American Journal of Primatology*, *5*(1), 1-15. https://doi.org/10.1002/ajp.1350050102|

Rhine, R. J., & Flanigon, M. (1978). An empirical comparison of one-zero, focal-animal, and instantaneous methods of sampling spontaneous primate social behavior. *Primates*, *19*(2), 353-361. https://doi.org/10.1007/BF02382803

Rhine, R. J., & Linville, A. K. (1980). Properties of one-zero scores in observational studies of primate social behavior: The effect of assumptions on empirical analyses. *Primates*, *21*(1), 111-122. https://doi.org/10.1007/BF02383828

Rhine, R. J., Norton, G. W., Wynn, G. M., & Wynn, R. D. (1985). Weaning of free-ranging infant baboons (*Papio cynocephalus*) as indicated by one-zero and instantaneous sampling of feeding. *International Journal of Primatology*, *6*(5), 491-499. https://doi.org/10.1007/BF02735572

Saibaba, P., Sales, G. D., Stodulski, G., & Hau, J. (1996). Behaviour of rats in their home cages: daytime variations and effects of routine husbandry procedures analysed by time sampling techniques. *Laboratory Animals*, *30*(1), 13-21. https://doi.org/10.1258/002367796780744875

Sands, J., & Creel, S. (2004). Social dominance, aggression and faecal glucocorticoid levels in a wild population of wolves, *Canis lupus*. *Animal Behaviour*, *67*(3), 387-396. https://doi.org/10.1016/j.anbehav.2003.03.019

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Vervet monkey alarm calls: semantic communication in a free-ranging primate. *Animal Behaviour*, *28*(4), 1070-1094.

Simpson, M. J. A., & Simpson, A. E. (1977). One-zero and scan methods for sampling behaviour. *Animal Behaviour*, *25*, 726-731. https://doi.org/10.1016/0003-3472(77)90122-1

Shora, J., Myhill, M., & Brereton, J. E. (2018). Should zoo foods be coati chopped. *Journal of Zoo and Aquarium Research*, *6*(1), 22-25. https://doi.org/10.19227/jzar.v6i1.309

Smith, P. K. (1985). The Reliability and Validity of One‐zero Sampling: misconceived criticisms and unacknowledged assumptions. *British Educational Research Journal*, *11*(3), 215-220. https://doi.org/10.1080/0141192850110302|

Stevens, J., Thyssen, A., Laevens, H., & Vervaecke, H. (2013). The influence of zoo visitor numbers on the behaviour of harbour seals (*Phoca vitulina*). *Journal of Zoo and Aquarium Research*, *1*(1), 31-34. https://doi.org/10.19227/jzar.v1i1.20

Suen, H. K. (1986). On the utility of a Post Hoc correction procedure for one-zero sampling duration estimates. *Primates*, *27*(2), 237-244. https://doi.org/10.1007/BF02382602

Suen, H. K., & Ary, D. (1984). Variables influencing one-zero and instantaneous time sampling outcomes. *Primates*, *25*(1), 89-94. https://doi.org/10.1007/BF02382298

Tamaki, N., Morisaka, T., & Taki, M. (2006). Does body contact contribute towards repairing relationships? The association between flipper-rubbing and aggressive behavior in captive bottlenose dolphins. *Behavioural Processes*, *73*(2), 209-215. https://doi.org/10.1016/j.beproc.2006.05.010

Teixeira, D. L., Machado Filho, L. C. P., Hötzel, M. J., & Enríquez-Hidalgo, D. (2017). Effects of instantaneous stocking rate, paddock shape and fence with electric shock on dairy cows' behaviour. *Livestock Science*, *198*, 170-173. https://doi.org/10.1016/j.livsci.2017.01.007

Tyler, S. (1979). Time-sampling: a matter of convention. *Animal Behaviour*, *27*, 801-810. https://doi.org/10.1016/0003-3472(79)90016-2

Ward, S. J., Sherwen, S., & Clark, F. E. (2018). Advances in applied zoo animal welfare science. *Journal of Applied Animal Welfare Science*, *21*(1), 23-33. https://doi.org/10.1016/0003-3472(79)90016-2

Wirth, O., Slaven, J., & Taylor, M. A. (2014). Interval sampling methods and measurement error: A computer simulation. *Journal of Applied Behavior Analysis*, *47*(1), 83-100. https://doi.org/10.1002/jaba.93|

Xiao, J., Wang, K., & Wang, D. (2005). Diurnal changes of behavior and respiration of Yangtze finless porpoises (*Neophocaena phocaenoides asiaeorientalis*) in captivity. *Zoo Biology*, *24*(6), 531-541. https://doi.org/10.1002/zoo.20070