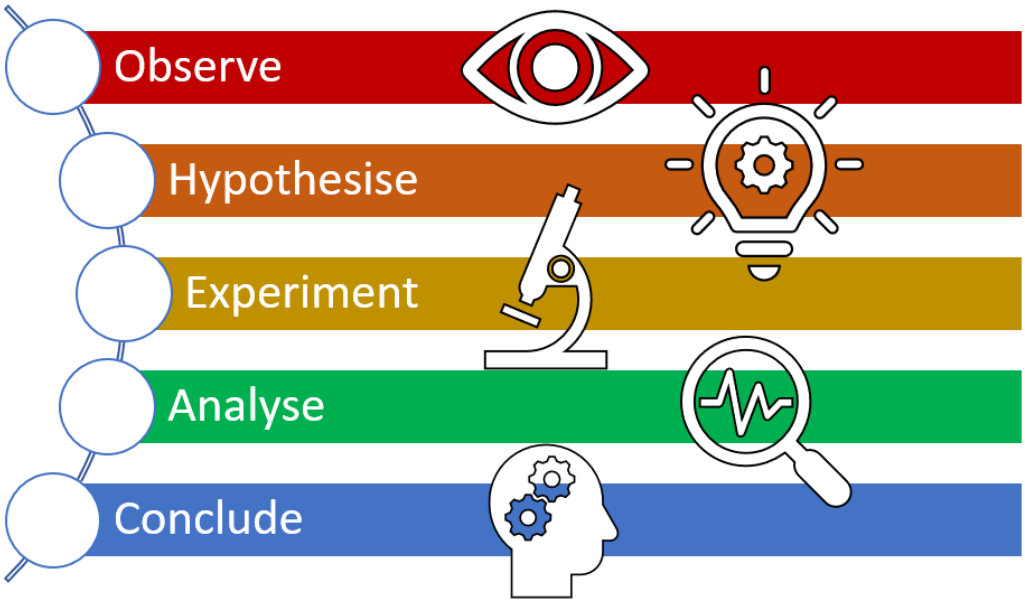
A Horse Owner's Guide to Interpreting Research Evidence

Part 4: Observation and The Scientific Method

~ Tracy Bye MSc

In the first three instalments in this series we learnt about a specific type of research study known as an **intervention study**. This is a study where we change something and then measure the effects of this change. We also talked about lots of things that we need to consider when evaluating intervention studies, like the outcome variables (things we choose to measure), validity, reliability, potential sources of bias, and the placebo effect. What we have been describing is what researchers call the **scientific method**.

The scientific method is the process of testing an idea or **hypothesis** via a properly designed **experiment** and using the measurements (**results**) to inform a logical **conclusion**. The development modern scientific method is credited to Sir Francis Bacon in 1620. It was this concept which informed the founding of the Royal Society [1] in 1660, an organisation dedicated to promoting excellence in science and to demonstrating the importance of science to all. The motto of the Royal Society is *Nullius in verba* which is taken to mean 'take nobody's word for it', thus reinforcing that all new ideas, or hypotheses, should be subject to verification via experiment.



Diagrammatic representation of the scientific method.

So, what does all of this mean for equine research? Well, one thing that we see a lot of within the equine industry is observational learning. We can learn a lot about our horses by observing them; we learn about how they interact with other horses in the field, we learn about their eating and drinking habits, what tack and equipment they prefer, how the feel about the vet, whether they mind getting their feet wet, and much more. This gives us a wealth of really useful information about our own horse(s), or those horses we work with. We may even see patterns in how horses behave, and think

of reasons for why these patterns exist. For example, if none of the horses you look after will eat a particular new feed, then you would be perfectly reasonable to think 'this feed doesn't taste nice and horses don't like it'. But what you have done here is made an **observation** (my horses have not eaten this feed) and formed a **hypothesis** (horses do not like this feed). You can see from the diagram above, that this is just the first two steps from the scientific method. To come to a proper conclusion about ALL horses you would need to set up an experiment to test your hypothesis in a controlled setting.

Equestrianism has been around for hundreds of years, with some of the earliest teachings dating back to the work of the Greek philosopher and horseman Xenophon in approximately 400 BCE [2], long before the development of the modern scientific method. So, it stands to reason that many of the 'rules' of horsemanship have not been scientifically tested. This means that sometimes scientific investigation of long held ideas might show that some of these ideas are not correct. A useful example of this is the use of the 'skin pinch test' for dehydration. Many of us were taught to pinch the skin on the horse's neck and if it didn't immediately flatten out again, this meant the horse was dehydrated. A research study in 2008 discovered that this skin pinch test was not a valid or reliable measured of dehydration (see part 3 of this series for more on validity and reliability). There was no relationship between plasma osmolality (water content in the blood plasma) and the results of the skin pinch test, and what's more the skin pinch test results were affected by which side of the neck was used, whether the coat itself was wet or dry, and by the age of the horse [3]. This is one example, but it does not mean that all 'rules' of horsemanship based on observation are wrong, and it certainly does not mean we should forgo common sense. In fact, many observations are proven to be correct, for example the case of 'Monday morning disease'. We now know this as equine exertional rhabdomyolysis or 'tying up'. The name Monday morning disease came from the observation that horses which were given a day off on a Sunday and returned to work on a Monday were more likely to suffer from this disorder. This idea has since been supported, with scientific research providing numerous exercise management suggestions for horses prone to tying up, including keeping days off to a minimum [4].

Research studies have shown that the skin pinch test shown here is not a reliable indictor of dehydration in horses, and the results can vary if the horse's coat is wet (such as here) or dry.

The important thing to take away from this is that making an observation and coming up with a hypothesis is only a part of the process. Lots of horse owners make observations based on a limited sample of one or two horses, which provides what we call **anecdotal evidence**. We know that horses are individuals, so often this anecdotal evidence can be contradictory. For example, one person can say 'this product helped my horse' whereas another may say 'this product is useless'. And don't forget that these observations can be impacted on by things like unconscious bias and the caregiver placebo effect, as discussed in previous articles in this series. So ultimately, evidence from well-designed scientific investigations should be accepted as having greater value than anecdotal evidence based on a small number of observations made in an uncontrolled situation.

There are some scenarios when we may want to use observation as part of the scientific method, if the thing we are interested in cannot be measured using an objective variable. One of the most common examples of this is measuring behaviour. Any observation is by definition a subjective measure, but we try to standardise the method as much as possible by using things such as **ethograms** or **Qualitative Behavioural Assessment (QBA)** scoring systems.

An ethogram is a list of definitions or descriptions of behaviours we are looking for. This can be an event behaviour, such as bucking or blinking, and the observer just needs to identify how many times this happened in the observation period. Or it can be a state behaviour, such as lying down, and the observer might want to record how long this behaviour is shown for. A QBA is a list of definitions of behavioural states such as 'tense', 'excited', or 'relaxed'. This type of measure involves assessing the demeanour of the whole horse, rather than looking for specific behaviours, but what each behavioural state looks like is still clearly defined before the study starts. The two approaches differ in how the behaviour is interpreted, for example the bucking horse identified by the ethogram could be described as either 'tense' or 'excited' in a QBA study, depending on context and other behavioural indicators shown.

As with any subjective outcome measure, the application of an ethogram or QBA is still open to individual interpretation. For example, different people may come up with different values for how many times a horse performed a particular behaviour, as they may place the 'cut-off' for performing or not performing the behaviour at different points. This is tackled by having multiple people doing the scoring, making sure all observers have training on identifying the behaviours before the study starts and having a defined way of coming to agreement between them if there is disagreement (e.g. taking an average score). If you are reading a piece of research that is based on observations, they should explain in their methods, how and by whom the observations were done, what the criteria were for specific behaviours or scores, and how they handled any lack of agreement in the results. They may also include measures of **inter-rater reliability**. This tells you how closely the different observers (raters) agreed with each other in their scoring, and can be used as an indicator of how well-designed and applicable the measurement system (ethogram or QBA) is.

We have covered many of the factors which make a good experiment in previous articles of the series, but one of the things I have mentioned a couple of times in this article is the number of horses being observed. Anecdotal evidence is usually based on a very small sample of horses. The number of horses needed for a proper research study is typically much larger, and this is determined by the statistical

analysis which will be conducted on the data collected. In the next instalment we will discuss how we decide on a sample of horses for our study and how we can use statistical analysis to decide if we agree with our hypothesis.

References

1. The Royal Society- https://royalsociety.org/
2. Xenophon. The Art of Horsemanship (edited and translated by M. H. Morgan in 2006). New York: Dover Publications.
3. Pritchard, J.C., Barr, A.R.S. and Whay, H.R., 2006. Validity of a behavioural measure of heat stress and a skin tent test for dehydration in working horses and donkeys. Equine Veterinary Journal, 38(5), pp.433-438.
4. Valberg, S.J. 2010. The management of tying up in sport horses: challenges and successes. Proceedings of the 2010 Kentucky Equine Research Nutrition Conference, Lexington, Kentucky, United States of America 16-17 April 2010. pp. 82-93.