Saddle Research Trust

A Horse Owner's Guide to Interpreting Research Evidence

Part 3: Measurement, Validity, and Reliability

By Tracy Bye

In the first two instalments of this series we introduced some of the things researchers consider when they are designing a study, such as how to avoid pitfalls like individual differences, unconscious bias, or the placebo effect impacting on the results. One of the most important things to consider is what we are going to measure, our 'outcome' variable.

We can group outcome variables into two broad categories, **quantitative**, which are numerical values, such as heart rate, angle of a joint, or dressage test score, and **qualitative**, which are descriptions, such as comments from a dressage judge, or answers in an interview. Qualitative measures are generally **subjective**, meaning they are based on someone's opinion and open to interpretation. This doesn't mean they are an inferior way of conducting research, it just means that the researchers need to be really careful about how they interpret qualitative data and must go through the data systematically so they don't miss anything. Ideally, more than one person will review and interpret the data and then they will come to an agreement on the main messages. A systematic approach to analysing qualitative data is really important to avoid 'cherry picking' the bits which support the researchers' hypothesis.

Quantitative data can be either subjective or **objective**, depending on how it is derived. An example of subjective quantitative data is a dressage test score, which is a numerical value. It is not open to different interpretations like the qualitative data, but it is still based on someone's opinion, and if a different person did it then it might be different. Objective quantitative measures are values which are not open to interpretation, such as heart rate or joint angles. Often these are measured by some type of equipment to limit the effect of human error, for example using a heart rate monitor rather than taking a pulse.

Head-neck angle is an example of an objective, quantitative outcome variable. If you were to describe this horse's behaviour then that would be subjective and qualitative. If you were to give a behaviour score based on an agreed scale that would be subjective and quantitative.
Photo from Robinson and Bye (2021).

When researchers are deciding on what the outcome variable is going to be, they will also think about the **validity** and **reliability** of this measure. Validity is how well the outcome measure actually represents the thing we want it to measure. So, if we want to measure stress, for example, we cannot just ask a horse how stressed it is, but we can measure its behaviour, and physiological things that have been linked with stress, such as heart rate variability [1], blink rate [2] or eye temperature [3] (check out the references if you want to learn more about how these different measures work). A study with good validity controls any other factors which could affect the outcome variable other than the intervention. So, if we wanted to use heart rate variability as the outcome measure to look at the effect of a calming supplement on stress levels, we would control for anything else that could affect heart rate variability, such as exercise.

A study with good reliability makes sure that the measures are taken in the same way each time, so they should always yield the same results. They do this by having a strict experimental protocol to **standardise** the intervention and the measurement, and making sure to follow it exactly with every horse every time they measure it. This means the only thing that changes is the intervention, and any differences seen can be attributed to that intervention, and not the chance that something has been done differently.

Photo by Kenny Eliason on Unsplash

You can think of the ideas of validity and reliability in research in terms of shooting arrows at a target. If your measure is **valid** then you will hit the bullseye and you are measuring what you want to measure. If your measure is **reliable**, you will hit the same spot on the target every time. A measure can be reliable without being valid. i.e. you could get the same results (hit the same spot) every time but the measure might not test the thing you want (hit the bullseye).

You may have already picked up on a problem with all of this **standardisation**. How do we know that these interventions will work in the real world, with horses that are not in strict

experimental conditions? This is the problem of **external validity**, or how well a study represents the real world. The type of validity we have been talking about so far is also termed **internal validity**. Internal and external validity are on a continuum, so a study cannot be high in both. The more controlled a situation is, the less true to life it is, and vice versa.

Usually when scientists trial new interventions, we want the internal validity and reliability to be as high as possible, so we will test things in a very controlled situation to start with. For example, we may test a new training aid on a sample of ex-racehorses, all of a similar age, conformation, and training background, working on a treadmill all going at the same speed. This scenario has high internal validity, as any differences we see in the horses' biomechanics are very likely to be due to the training aid, and not a change in speed or surface, and a homogenous sample like this is less likely to show differing effects based on how the training aid interacts with the horse's conformation or posture, or to elicit different effects based on how the horse was previously trained. Once something has been tested in a very controlled situation like this, then researchers will tend to try out different and more varied situations to see what happens when specific differences are added in. So, if you are looking at research on a new intervention, you may see different studies looking at it in different situations, with different levels of standardisation in the methodology. This allows us to build up a picture of how the intervention might work in a range of scenarios, but without compromising the internal validity and reliability needed in the early stages of researching a new intervention.

It is easy to understand how we can make objective, quantitative studies valid and reliable, but this needs a little more thinking about when you are considering qualitative data. We have talked a little here about subjective, qualitative data, and observation as a scientific approach, but there is a lot more we can learn about how to do this type of research. Horsemen were using observation to work out how to manage horses for centuries before the scientific method was even developed. **In the next instalment in this series we will be talking about where observation turns into observational research, and what we can learn from this.**

References

1. Stucke, D., Ruse, M.G. and Lebelt, D., 2015. Measuring heart rate variability in horses to investigate the autonomic nervous system activity–Pros and cons of different methods. Applied animal behaviour science, 166, pp.1-10.

2. Mott, R.O., Hawthorne, S.J. and McBride, S.D., 2020. Blink rate as a measure of stress and attention in the domestic horse (Equus caballus). Scientific reports, 10(1), pp.1-8.

3. Yarnell, K., Hall, C. and Billett, E., 2013. An assessment of the aversive nature of an animal management procedure (clipping) using behavioral and physiological measures. Physiology & behavior, 118, pp.32-39.