Are the Results of Animal Therapy Studies Unreliable?

Harold Herzog, Western Carolina University

Available at: https://works.bepress.com/harold-herzog/94/
Are the Results of Animal Therapy Studies Unreliable?

Most animal therapy studies do not have enough subjects to be valid.

Posted Aug 4, 2016

Animal-assisted therapy is a growth industry. According to a survey conducted by the Human Animal Bond Research Initiative, 69% of family practice physicians have worked with animals in medical settings. And Yale researchers reported that when it comes to the treatment of children with behavior problems, the public views animal-based therapies about as acceptable as psychotherapy and much more acceptable than drug treatments.

These interventions are attracting the attention of a growing number of investigators. The graph below shows the number of research articles published on animal-assisted therapies between 1990 and 2015. Research in this area is increasing at an exponential rate, with the number of published papers doubling every 3.6 years. That’s the good news.

What We Can Learn from Studies of “The Love Hormone”

The bad news is that most studies of the effectiveness of therapies using creatures like dogs, horses, and dolphins are flawed. Among the most common problems are inadequate control groups, lack of long-term follow-ups, insufficient controls for researcher expectations, and no standardized treatment procedures. But here I want to focus on the fact that most animal therapy studies do not have enough subjects to produce valid results.
I became aware of the “sample size problem” when I came across a surprising headline in the magazine *New Scientist*—“Everything You’ve Heard About Sniffing Oxytocin Might Be Wrong.” Oxytocin, of course, is the neurochemical often referred to as the “love hormone.” Over 1,000 studies have linked oxytocin to a host of behaviors, including maternal care, trust, sexual responses, and even our connections to pets. The author of the *New Scientist* article, Simon Oxenham, described a series of publications which have called these results into question. A 2015 paper, for example, reported that many studies showing that oxytocin increases trust have not been replicated by other researchers.

And, like other areas of science, oxytocin studies are prone to “the file drawer effect”—a bias against publishing negative results. This issue was brought home by a group of researchers in Belgium. Stung by the failure to replicate one of their own studies, they dug back into their file drawer for failed experiments. They were shocked to find that 24 out of 25 of their previous studies found oxytocin had little or no impact on the behaviors they were interested in.

**The Love Hormone’s Sample Size Problem**

According to the Belgian researchers, one reason oxytocin studies are often unreliable is there are too few subjects in the experiments. Hasse Walum of Emory University and his colleagues Irwin Waldman and Larry Young did the math and, in a paper recently published in the journal *Biological Psychiatry*, reported the Belgians were right. To understand their analysis, you need to know a couple of basic statistical concepts (bear with me for a minute—this is not rocket science).

The **Effect Size** of a study is an index of the magnitude of the differences obtained between the treatment group in an experiment and the control group. It is usually reported as “Cohen’s *d*.” The higher *d* is, the bigger the impact of the treatment. If *d* is .20, the treatment is said to have had little impact; a *d* of around .50 or so is regarded as indicating a medium size impact, and a *d* of .80 or greater signifies a large treatment effect.

The **Statistical Power** of a study is the likelihood it has enough subjects to detect the impact of treatment condition…if there really is one. In most studies, researchers aim for a statistical power of .80. This means you have enough subjects so that 80% of the time, your experiment would uncover true effects. While statistical power is affected by the number of subjects, it is also influenced by the effect size. If a treatment effect is small, you need more subjects to detect it. Studies with too few subjects are said to be “underpowered.” Paradoxically, underpowered research is both more likely to miss real effects and also to obtain positive results that later turn out to be false. The reason is that these studies tend to overestimate effect sizes when they are found.

Walum’s group examined the statistical power of a set of experiments which examined the impact of sniffing oxytocin on human behavior. Together, these studies involved 57 comparisons between a treatment group and a control group. The average effect size in these comparisons was .28, which is on the small side. The studies included, on average, 49 subjects. Walum calculated that the average statistical power of these studies was only 16%. These findings suggests that at least 84% of the time, the oxytocin experiments produced incorrect results even if a true effect of the hormone really exists.

**Animal-Assisted Therapy Studies Have the Same Problem**

Intrigued, I contacted Hasse to see if his analysis could be applied to animal-assisted therapy studies. He said yes, so I sent him some numbers. To estimate the sample size of the typical study, I used the trials described in 10 research reviews. The median sample size of these studies was 24.5 subjects. To estimate the effect size of animal therapy clinical trials I turned to a meta-analysis of 49 studies by Janelle
Nimer and Brad Lundahl. The average effect size among these trials was .46. Plugging in these numbers, I was surprised to find that the statistical power of the animal-assisted therapy studies was only about 20%—way short of the desired 80% mark.

Here is the bottom line. Most research on the effectiveness of animal assisted therapy is seriously underpowered. As a result, the majority of studies on the use of animals as therapists either do not detect a real therapeutic impact of or obtain false or inflated positive results. The obvious question is, how many subjects would it take to get to a statistical power of 80%? The magic number is 152. Only a handful of the hundreds of animal therapy studies have nearly this many subjects.

I admit these findings are depressing. But, while it is of little consolation, things are even worse in other areas. For example, the average statistical power of brain imaging research is only 8%. This means that the conclusions drawn from these studies are likely to be wrong 90% of the time.

Based on their analysis of the number of subjects in oxytocin studies, Walum and his colleagues wrote, “There is a high probability that most of the published intranasal oxytocin findings do not represent true effect.”

Is this also true for the results of animal-assisted therapy studies?

Thanks to Hasse Walum for his comments on this post.

References


