

Forced Swim Test Fails to Measure Human Depression

From [Marc Bekoff, Psychology Today / Animal Emotions](#)

September 2023

The forced swim test is a controversial animal experiment used to study depression. The forced swim test has poor face, construct, and predictive validity. It should be replaced with more humane and reliable methods of studying depression.



Animal models of human disease have extremely mixed results, and all too often they do not work well and accomplish their [goals](#) of helping humans.¹

One such procedure, the [forced swim test](#), is a prime example of an experimental procedure that has all sorts of problems. I'm pleased researcher Stephen Farghali of Black Hills State University could answer a few questions about why it doesn't help us learn more about human [depression](#).

Marc Bekoff: Can you tell readers why, after 45 years of the forced swim test, one of the most prevalent animal experiments fails to measure up?

Stephen Farghali: Imagine you want to know whether a substance, say zinc, alleviates depression. How should you scientifically test this? One good option would be to give zinc supplements to human participants for a length of time while assessing their depression.¹

Another would be to grab a lab mouse by the back of the neck and inject either zinc or a [placebo](#) down the esophagus. Then drop the mouse into a beaker of tap water and set a timer, watching as the mouse struggles to escape. Eventually, they give up. Instead of trying to escape, the mouse resorts to floating. If mice who get zinc struggle longer than those who don't, a leap of faith is made to conclude that zinc might alleviate human depression.

The latter option is called the forced swim test, and it's one of the most [common](#) animal experiments researchers use to study a substance for antidepressant properties. It rests on interpreting that floating behavior as "despair." A mouse who's given up trying to escape, the thinking goes, is analogous to a person suffering from depression. But this interpretation has been heavily criticized in the [decades](#) since it was originally [published](#) in 1977.

But the impact of the test's validity extends beyond methodological and ethical debates among researchers. To the extent that scientists rely on it as a model for human depressive disorders, any new treatments will be subject to the test's ability to translate from rodents to humans.

If the animal test indicates therapeutic potential when there really is none in humans, scientists end up wasting their time and resources, and the time of patients, who sign up for clinical trials on investigative medications that are doomed to fail. Such failures occur [in 95% percent of drug trials](#), meaning therapies that appeared safe and effective in animal experiments later turn out to be unsafe or ineffective in humans.

At the same time, if the test is inaccurate it can fail by providing a false negative, indicating no benefit when there actually is a benefit to humans. False negatives are harder to measure because the trials do not continue.

The crucial point is if the forced swim test fails to accurately predict and model human outcomes, then our efforts to develop new therapeutics are left to chance.

MB: What about issues of validity?

SF: Scientists generally evaluate animal models by three measures of validity: face, construct, and predictive.

Face validity relates to the similarity of symptoms between human illness and the animal experiment. Face validity remains a controversial goal in all animal experiments aimed at understanding human depression. Even proponents of animal use [have stated](#), "The laboratory rat or mouse will never exhibit depression under any circumstances." It's yet more problematic to argue that the forced swim test has face validity, as researchers have [demonstrated](#) that "there is little similarity between the clinical symptoms of depression in humans and the behaviors measured in the test." Even R. D. Porsolt, the developer of the test, [has acknowledged](#), "There is no obvious induction of a 'depressive state.' "

Construct validity is the second measure; it addresses how well a model replicates the underlying mechanisms of the illness. Here, too, the swim test falls short. Human depression is, by definition, a chronic condition, which [medication](#) takes weeks or months to affect, [if at all](#). The swim test creates an acute stressor—being dropped into an inescapable container of water—claims to alleviate the condition within hours, and with just one dose of medication.

Proponents of the swim test point to the third measure, predictive validity, as its redeeming quality. This is the model's ability to accurately identify an effective antidepressant. However, the test [has failed](#) to accurately identify some of the most widely used [antidepressants](#), including selective serotonin reuptake inhibitors (SSRIs), such as fluoxetine, and [dopamine](#) reuptake inhibitors (NDRIs) like bupropion.

And while some reviews point to the test's predictive accuracy with some known antidepressants, there are often caveats, detailing different effects with different strains of rodents. Essentially, then, the best argument for the test comes with the qualifier that results don't reliably generalize from one strain of mouse or rat to another, or even from one supplier of a given strain to another.

These results do not build [confidence](#) in the likelihood of the test translating well to human outcomes.

MB: How can we move past the forced swim test?

SF: Leaders in antidepressant development, including Pfizer, Johnson & Johnson, GlaxoSmithKline, and Bayer, have stated they will no longer use the forced swim test. Pfizer has [acknowledged](#) that “none of the compounds tested by Pfizer since 1989 using the [forced swim test] are currently approved to treat human depression, which means that the test did not lead to [marketing](#) these compounds as new medications.”

Similarly, in 2019, the director of the U.S. National Institute of Mental Health [wrote](#) that the forced swim and other similar tests "have largely failed to reveal translatable [neural](#) mechanisms, and lack specificity from a pharmacologic-validity perspective.”

Yet even as many labs move away from the forced swim test, its use as a screen for antidepressant properties is increasing in some sectors. Researchers still use the test to assess foods, herbs, and other commonly consumed, safe substances as treatments for depression. For example, academic researchers in France [recently sought](#) to assess the benefits of saffron for depression. Instead of recruiting human volunteers, they used the forced swim test.

Meanwhile, in the U.S., one group of researchers [used](#) the forced swim test to measure “depressive-like” behavior resulting from sleep apnea by waking mice up every two minutes with a horizontal bar sweeping across the floor of their cage. As with the saffron study, this could have been accomplished with humans, measuring depression using methods that have already been established with human participants.

With pay-to-publish journals becoming increasingly prevalent, there is no shortage of journals willing to publish questionable animal experiments like these.

A 2021 [survey](#) of journal editors showed that even those who are aware of the forced swim test’s failures are disinclined to reject papers that rely on it, citing their hesitance to “micromanage” the interpretation of methods. Yet, editors are not only within their means to reject papers for poor methodology, doing so is indeed one of their primary functions. Furthermore, the widely-accepted, NIH-endorsed "3Rs" [ethics framework](#) on using animals in research, which calls for nonanimal methods to replace animals wherever possible, gives cause for rejection.

While some industries are moving away from the forced swim test, others continue, despite its failure to serve as a valid measure of depression. Evidence of the test's theoretical and predictive failures suggests the public, whose tax dollars partially fund this research, and who rely on accuracy in drug development, would be better served if researchers stopped using the forced swim test for good.

References

In conversation with Stephen Farghali, who works for Physicians Committee for Responsible Medicine and teaches psychology at Black Hills State University.

1) ["Rat Trap": Why Animal Models of Human Disease Must Be Replaced; Can We Move Away from Animal Models of Human Disease Research?; It's Time to Move on From Nonhuman Animal Models.](#)