

# Dolphin-Assisted Therapy: More Flawed Data and More Flawed Conclusions

**Lori Marino\* and Scott O. Lilienfeld†**

*\*Neuroscience and Behavioral Biology Program, Emory University, USA*

*†Department of Psychology, Emory University, USA*

*Address for correspondence:*

*Lori Marino, Neuroscience  
and Behavioral Biology  
Program, Emory University,  
1462 Clifton Road, Suite 304,  
Atlanta, GA 30322, USA.  
E-mail: lmarino@emory.edu*

**ABSTRACT** Dolphin-Assisted Therapy (DAT) is an increasingly popular choice of treatment for illness and developmental disabilities by providing participants with the opportunity to swim or interact with live captive dolphins. Two reviews of DAT (Marino and Lilienfeld [1998] and Humphries [2003]) concluded that there is no credible scientific evidence for the effectiveness of this intervention. In this paper, we offer an update of the methodological status of DAT by reviewing five peer-reviewed DAT studies published in the last eight years. We found that all five studies were methodologically flawed and plagued by several threats to both internal and construct validity. We conclude that nearly a decade following our initial review, there remains no compelling evidence that DAT is a legitimate therapy or that it affords any more than fleeting improvements in mood.

**Keywords:** DAT, dolphin-assisted therapy, dolphins, therapy, validity



Dolphin-Assisted Therapy (DAT) is an increasingly popular choice of treatment for illness, disability, and psychopathology in children and adults (Whale and Dolphin Conservation Society 2006). It involves swimming and interaction with dolphins, typically in captivity. DAT formally began in the 1970s and over the years has grown into a highly lucrative business with facilities all over the world, including the United States, Mexico, Israel, Russia, Japan, China, and the Bahamas, to name only a few countries. The claims made by these facilities have been subject to little or no scientific scrutiny. Moreover, there has been no significant increase in the rate of peer-reviewed papers on DAT from the 1970s to the present. Yet DAT programs continue to proliferate. As a consequence, DAT's popularity greatly outstrips its meager research base.

Eight years ago we published a review (Marino and Lilienfeld 1998) of the available peer-reviewed DAT literature at the time, focusing on two papers by David Nathanson and his colleagues (Nathanson et al. 1997; Nathanson 1998). These authors advanced several extremely strong claims concerning the efficacy of DAT for treating severely disabled children: (1) DAT significantly increases attention span, motivation, and language skills; (2) DAT achieves

these results more rapidly and more cost-effectively than conventional therapies; and (3) DAT produces positive treatment effects that are maintained over a long-term period (i.e., at least one year).

In our original paper (1998) we presented a methodological analysis of these two studies and the claims derived from them by applying standard criteria for scientific validity from four established sources (Kazdin and Wilson 1978; Cook and Campbell 1979; Kendall and Norton-Ford 1982; Shaughnessy and Zechmeister 1994). We found no fewer than eleven independent methodological weaknesses in both studies that seriously undermined their scientific validity. These shortcomings included the absence of adequate comparison or control groups, unreliable, subjective and potentially biased raters, and analytic methods that did not allow the reader to ascertain whether any children were harmed by DAT. We concluded:

“...serious threats to validity and flawed data analytic procedures render the findings of Nathanson and colleagues uninterpretable and their conclusions unwarranted and premature... the current evidence for the efficacy of DAT can at best be described as thoroughly unconvincing” (Marino and Lilienfeld 1998, p. 199).

At the time of our review, these two studies were the only peer-reviewed investigations of the therapeutic effects of DAT. Therefore, as of 1998 there was no credible scientific evidence for the effectiveness of this intervention.

A more recent and similar critical review of DAT was published five years later by Humphries (2003). In her excellent analysis of the peer-reviewed DAT literature, Humphries evaluated six DAT studies and found that all six lacked experimental controls and neglected to control adequately for major threats to validity and alternative explanations for the results. She concluded that:

“...the available research evidence, as examined in this synthesis, does not conclusively support the claims that DAT is effective for improving the behaviors of young children with disabilities. More specifically, the results of the synthesis do not support the notion that using interactions with dolphins is any more effective than other reinforcers for improving child learning or social-emotional development.” (Humphries 2003, p. 6)

Finally, Brensing, Linke and Todt (2003) evaluated the claim that DAT works through healing by ultrasound. The authors conducted an observational study of the behavior of dolphins in two DAT programs to determine if their behavior was consistent with the hypothesis that they were echolocating on human participants (and thus ostensibly providing “healing energy” through ultrasound). They concluded that the dolphins’ behavior was not consistent with this hypothesis and therefore did not meet the minimal requirements for common ultrasound therapies. Therefore, there is no scientific evidence that echolocation can heal, nor that dolphins echolocate on humans in a manner even minimally consistent with that claim.

In this paper, we update the DAT literature by reviewing the five peer-reviewed DAT studies published since 1998. We focus on peer-reviewed papers because these articles presumably represent the “best” evidence for the efficacy of DAT. We argue that an updated analysis of DAT is warranted given the increasing popularity of this activity. Humphries (2003) reviewed peer-reviewed studies that ranged from 1989 to 1999, two of which had already been evaluated by Marino and Lilienfeld (1998). Brensing, Linke and Todt (2003) based their conclusions on data taken from their studies of DAT at two facilities and did not provide a methodological evaluation of the peer-reviewed literature. In the present article, we pick up where our 1998 paper left off by evaluating the methodological validity of the five peer-reviewed DAT studies that have been published since then.

## Methods

We examined a total of five papers describing studies using DAT. These are Antonioli and Reveley (2005), Iikura et al. (2001), Lukina (1999), Servais (1999), and Webb and Drummond (2001). The authors of four of the studies reported improved behavioral outcomes for participants receiving DAT.

The lone exception was Servais (1999), who reported positive outcomes in only one of the two experimental groups in her study. Beyond the ambiguous results reported by Servais (1999), we found no other published studies in which negative findings were noted.

We searched for peer-reviewed studies on DAT in several ways. First, we submitted the search terms “dolphin assisted therapy” and “dolphin therapy,” to the Google Scholar search engine. Second, we used these same terms to search Google for the websites of DAT facilities and searched any bibliographies listed on these websites. Third, we used these terms in a search of the *Current Contents* database. Fourth, we conducted a comprehensive search of all papers on DAT from 1999 to the present in the following peer-review journals: *Anthrozoös*, *Society & Animals*, *Applied Animal Behaviour Science*, and *Zoo Biology*. Fifth, we reviewed the reference sections of articles identified through the above means for further relevant studies.

As in our 1998 paper, we assessed the validity of each study according to the standard criteria put forth by four sources: Cook and Campbell (1979), Shadish, Cook and Campbell (2002), Kendall and Norton-Ford (1982), and Shaughnessy and Zechmeister (1994). These sources describe a set of threats to experimental validity that should be avoided in experimental research. The presence of even one major threat to validity can render a study’s findings difficult, or in some cases even impossible, to interpret.

Table 1 displays several important threats to validity, their definition, and whether they are present in each of the five studies. Most of these threats relate to either *internal validity*, that is, the methodological soundness of the study, or to *construct validity*, that is, the soundness of the measures as indicators of the constructs examined by the investigators. In the interest of space, we limit ourselves here to the most serious threats to validity.

## Results

Table 1 shows that each of the five studies we examined violated several important criteria for validity. Moreover, because of inadequate experimental control, two threats to construct validity, that is, nonspecific effects and construct confounding, plagued all of these studies. Because these threats to validity are so ubiquitous in the DAT literature, we discuss them in depth in separate sections and follow with more specific points concerning each of the individual studies.

### *Nonspecific Effects*

Nonspecific effects are improvements from influences that are not specific to the intended treatment, and that are shared by a wide variety of other treatments. They are generic effects of the treatment rather than the result of the intended therapeutic ingredient(s). Two relevant subcategories of nonspecific effects are placebo effects and novelty effects. The placebo effect is the well-documented but little understood improvement that derives from participants’ expectation of improvement. Novelty effects are the general energizing and uplifting effects of a new, exciting experience. Because of its nature, DAT is particularly prone to these two nonspecific effects. Furthermore, many nonspecific effects are notoriously transient. Therefore, any study of DAT must incorporate rigorous procedures for minimizing these nonspecific effects and, in particular, must include ways to assess longer-term effects of the treatment.

DAT is vulnerable to placebo effects in part because DAT is typically marketed to participants and their family members as highly efficacious and in part because the nature of the treatment is evident to participants. Moreover, none of the studies we reviewed used a control that eliminates or substantially minimizes this effect by eliminating cues to treatment condition. DAT is vulnerable to novelty effects because of the obviously new and exciting experience of swimming with and interacting with a large, intelligent, charismatic animal. The proper control for novelty would be exposure of the control group to another novel, attractive animal, while keeping all else equal. In this way, both groups would have similar reason to believe that they had received the relevant treatment, and both would be subject to the excitement of interacting with an exotic animal. At the very least,

**Table 1.** Major threats to validity in each of the five studies.

Validity Threat	Definition	Antonioli and Reveley (2005)	Ikura et al. (2001)	Lukina (1999)	Servais (1999)	Webb and Drummond (2001)
<b>Construct Validity</b>						
<i>Nonspecific Effects (Improvement from effects not specific to the intended treatment)</i>						
Placebo	Improvement from expectation of improvement	X	X	X	X	X
Novelty	Effects of energy, excitement, and enthusiasm not specific to the intended treatment	X	X	X	X	X
Construct Confounding	Failure to take into account the fact that the procedure may include more than one active ingredient	X	X	X	X	X
Resentful Demoralization	Participants aware of not receiving the active treatment may be resentful and respond more negatively than the treatment group	X				
Demand Characteristics	Tendency of participants to alter their responses in accord with their suspicions about the research hypothesis	X		X		X
Experimenter Expectancy Effects	Tendency for experimenter to unintentionally bias the results in accordance with the hypothesis		X	X	X	
<b>Internal Validity</b>						
History	Occurrence of potentially therapeutic events other than the intended treatment during the course of the study		X	X		
Testing	Improvements due to testing itself (e.g., practice effects)			X	X	
Regression	Tendency of extreme scores to become less extreme on re-testing		X		X	
Instrumentation	Changes in the dependent measure at different times in the study		X		X	
Multiple Intervention Interference	Administration of treatments other than the intended treatment during the course of the study		X	X		
Maturation	Changes over time due to natural developmental effects		X	X		
Informant Bias	Tendency of informants to selectively recall improvement in accord with their hopes and expectations (retrospective bias) or unintentional distortion of improvement due to effort justification	X				X

DAT should be compared with other animal-assisted therapies in addition to a no-treatment control group. If one found differential effects of dolphins versus other large charismatic animals, it would be important to seek possible specific therapeutic ingredients inherent in DAT. In contrast, if both dolphin and other animal groups improved equally, this would suggest a generic characteristic of DAT as an example of a temporary “feel-good effect” (activating effect) received from any animal therapies—and indeed any therapies involving exciting and novel features.

### ***Construct Confounding***

Construct confounding occurs when there is a failure to take into account the fact that the experimental procedure may involve more than one active (effective) ingredient. In DAT, the experimental treatment typically consists of a complex assortment of ingredients in addition to interacting with a dolphin per se, such as swimming in water, being near water, being outside, and receiving attention from human professionals. Moreover, the dolphin itself is a complex stimulus that can be deconstructed into various potentially therapeutic components, such as the size and touch of the animal and the opportunity for interaction with the animal. Because none of the DAT studies we examined adequately controlled for these possibilities, they are all subject to construct confounding. In the psychotherapy literature, construct confounding is typically decomposed by means of dismantling studies (Kazdin 1994), which separate the potential effects of different treatment ingredients by creating different experimental conditions containing these effects. And although we are not suggesting there is one single, ideal control for DAT, no DAT study has included an adequate subset of the many control groups that would be required for even a minimally effective dismantling strategy.

### ***Antonoli and Reveley (2005)***

Antonoli and Reveley (2005) conducted a single, blind, randomized controlled experiment to determine if swimming and interacting with dolphins at a captive facility in Honduras lowered scores on depression and anxiety scales. The total sample consisted of 30 men and women (prior to participant drop outs) with mild to moderate depression scores on the ICD-10 (i.e., the International Classification of Diseases, 10th edition) who also scored at least 11 on the modified Hamilton rating scale for depression at baseline after four weeks without medications. The experimental group (the animal care program) consisted of 13 participants who played with, swam with, and “took care of” dolphins while in the water with them. The control group (the outdoor nature program) consisted of 12 participants who swam and snorkeled in the barrier reef and experienced a similar degree of individualized human contact as those in the experimental group, but in the absence of dolphins. Both programs ran simultaneously from Monday to Friday for one hour a day for two weeks. The authors conducted pre-treatment and post-treatment ratings of depression with the Hamilton Rating Scale and the Beck Depression Inventory, and measured anxiety with the Zung Self Rating Anxiety Scale.

The authors found a significantly greater improvement in depression scores in the experimental group than in the control group. They also found no differences in the anxiety scores (although the authors argued the lack of statistical significance could be due to the fact that only a subset of the sample was clinically anxious prior to treatment). It was concluded that DAT is more effective than “water” therapy for treating mild to moderate depression. However, this conclusion is compromised by several shortcomings.

First, the authors acknowledged that a limitation of their study is that the participants were not blind to treatment. Therefore, demand characteristics might have influenced participants’ responses. The authors, did, to some extent, guard against potential demand characteristics by emphasizing to participants that they were taking part in a research study rather than a clinical intervention and should therefore not expect any improvement. Still, it is not known whether participants may have discerned the true nature of the study despite this information.

The second major shortcoming is that the control condition did not account for the potential effects of interacting with any charismatic animal, in this case, a dolphin. The experimental group swam with dolphins, whereas the control group did not interact with any other analogous animal.

Participants in the control group swam along a coral reef and interacted with the experimental personnel. However, there was apparently no introduction of other large mobile animals, such as fish, that could have served as a salient stimulus control for the dolphin. Therefore, because of the lack of an appropriate control group, participant outcomes could have been due to nonspecific effects, including placebo and novelty effects. Therefore, the authors' inference that the dolphin was the effective therapeutic ingredient is premature. The third potential threat to validity in this study is informant bias. Because the authors' relied on self-report measures of depression, the participants, who were not blind to treatment, could have selectively recalled the amount of their improvement as a consequence of expectations, hopes, and even effort justification (i.e., the psychological need to justify to oneself the time, expense, and energy invested in the treatment).

Fourth, the authors acknowledged that they did not perform a follow-up study of the post-treatment ratings. Therefore, the only justifiable conclusion they can draw is that there was a difference between the experimental and control groups immediately following treatment (and even that conclusion is suspect given the flaws discussed above). It is not known whether this difference existed even one day after the post-treatment assessment and, moreover, whether any participants deteriorated. This lack of follow-up assessment is a major shortcoming, as it renders the results susceptible to novelty or short-term activating effects.

Related to this point, the authors touted the benefits of DAT over conventional drug therapy or psychotherapy by claiming that depression is relieved after only two weeks of dolphin therapy but requires at least four weeks to improve with mainstream therapies. The authors neglected to point out, however, that the conventional treatments often exert long-term effects. In contrast, the authors provided no evidence that dolphin treatment produces any effects beyond the immediate aftermath of the treatment.

Finally, Shadish, Cook, and Campbell (2002) refer to a threat to construct validity known as "resentful demoralization," which may occur when members of a control group become aware that they are receiving a less desirable treatment or no treatment, thereby becoming disappointed and resentful. These negative emotions can then affect their responses. Resentful demoralization may be responsible for negative control group outcomes that falsely produce the appearance that the treatment was effective. Antonioli and Reveley appeared to be aware of this possibility and attempted to mitigate it by allowing the control group to swim with dolphins at the conclusion of the treatment. However, because the control group was allowed to swim with the dolphins only *after* the final evaluation, the possibility of resentful demoralization was not adequately eliminated.

### *likura et al. (2001)*

likura et al. (2001) sought to determine whether the presence of dolphins enhanced the effectiveness of seawater therapy for atopic dermatitis. Thirty-six patients with atopic dermatitis were subjected to seawater therapy (bathing in seawater at a beach) with dolphins present for six days, and 27 received the seawater treatment without dolphins for six days. No other details were offered about the seawater therapy, the condition of the patients, the nature of exposure to the dolphins, or any other methodological components of the study. Yet the authors concluded that all parameters of the atopic dermatitis improved in both groups and, more important, that the dolphin group experienced "less stress and pain" during the seawater therapy than the non-dolphin group. Furthermore, the authors claimed that the "patients also enjoyed swimming with the dolphins." They concluded that dolphins were "very useful as a pain reliever during therapy..." (p. 390).

It is difficult to evaluate the methodology and conclusions of likura et al. given the marked paucity of information in the paper regarding how the study was conducted. Moreover, the authors' conclusions are vague and subjective. In the absence of appropriate methodology, it can only be concluded that the likura et al. study is wholly uninterpretable and offers no credible support for the efficacy of DAT. Therefore, as far as can be determined, most of the potential threats to validity described in Table 1 apply to likura et al.

### *Lukina (1999)*

Lukina (1999) employed a single-group pretest-posttest design to determine the effects of DAT on “psychoneurological” functioning in healthy children and those with various diseases. The participants were 57 healthy children, 30 with “infantile neurosis,” 25 with mental retardation and autism, and 35 with other unspecified diseases. All of the children swam with and interacted with captive dolphins for 10–15 minute periods over 5–10 sessions. Although Lukina indicated that a number of psychological tasks were presented to the children during these sessions, she did not describe them. Likewise, although she stated that her outcome measures included many tests and observations, none were described in sufficient detail. Therefore, most of the threats to validity in Table 1 cannot be ruled out.

Lukina (1999) reported that the variability of cardiac rhythms in all groups increased after the dolphin exposure and claimed that “this confirms the redistribution of the psychoemotional dominants in the course of contacts with the dolphin, a fact that opens possibilities for rehabilitation measures and psychotherapy” (p. 678), but it is unclear what the author meant by “psychoemotional dominants” or how they are related to cardiac rhythms. Furthermore, the author claimed that after the dolphin sessions, parents and facility workers noted the emergence of “new-individual personal qualities,” such as kindness and self-control. Nevertheless, these observations were subjective and unquantified. Also, Lukina claimed that many disease symptoms in the “infantile neurosis” group, such as depression, night phobias, hysteria, and enuresis, diminished and that all the children responded “positively” to the dolphin sessions. Nevertheless, she offers no data or description of the assessment instruments to support these purported outcomes. Finally, because psychotherapy was a part of the system of treatment, it is possible that any improvement in the participants was due to interventions other than the dolphin component per se, thereby making multiple interference intervention a potential threat to validity.

In addition to these flaws, the pivotal weakness of the Lukina study is the absence of a control group consisting of children who did not swim with dolphins. Therefore, the study does not meet the minimal criteria for basic experimental design. This flaw alone renders the Lukina study difficult to interpret even without the myriad other threats to validity. Although sophisticated single-subject designs for drawing causal conclusions exist (Shadish, Cook and Campbell 2002), simple A-B (pretest-posttest) designs are typically extremely limited in internal validity.

### *Servais (1999)*

The design of this study included two experiments, lasting 16 and 14 months, respectively. The first included three groups of 3 autistic children each. The dolphin group was taught a cognitive task while interacting with a dolphin and trainer at dockside. The two control groups were a classroom group and a computer group in which the same cognitive task was taught in their respective settings. The second experiment included a dolphin group and a classroom group. Servais did not report the actual age of the participants but indicated that the developmental ages were 1–3 years. All sessions were held on an individual basis and lasted 15–20 minutes. The dolphin and computer groups were given 10–35 “habituation sessions” to familiarize themselves with the new setting and were required to master some simple behaviors before moving on to the next phase of the study. All groups received pre-tests followed by 10–15 “learning sessions” in which other cognitive tasks were given, and then post-tests.

It is not clear whether the pre-tests and the post-tests were the same across or within groups. Therefore, instrumentation effects (the effects of changing dependent measures at different times in the study) might be present. Besides post-test performance, the other outcome measure was a rating of attention based on videotaped observations by the author. Because the author was the only person to code behavioral outcomes, it is possible that experimenter expectancy effects influenced the findings.

Servais (1999) found that the first dolphin group learned the tasks better than the other groups, but reported no other differences between groups. In particular, the second dolphin group did not

perform better than the control group. Servais concluded that the positive results were due to the emotional interaction between the experimenters and children and that better-designed studies may “make the ‘animal effect’ disappear” (p. 14).

Despite Servais’ appropriate cautions about her findings and the effects of DAT, it is worth noting several threats to validity. In addition to experimenter expectancy effects and potential instrumentation effects, demand characteristics are impossible to rule out. Moreover, the authors’ conclusion that the dolphin may not have been as important for improvement as other components of the treatment points saliently to the possibility that other aspects of the treatment procedure (such as swimming outdoors) besides the dolphin interaction were responsible for the reported effects.

### *Webb and Drummond (2001)*

Webb and Drummond (2001) investigated the psychological effects of swimming with dolphins at a marine park in Australia. Participants consisted of a wide age range of male and female teenagers and adults without known or, at least, measured clinical pathology. They were paying participants and had been on a waiting list for up to six months. In the well-being study, the experimental group consisted of 74 females and 25 males who swam with four dolphins in groups of four for 25–30 minutes. The control group consisted of 14 females and 15 males who swam at a beach adjacent to the marina in the absence of dolphins. Both groups self-reported their feelings of well-being pre-treatment and post-treatment. In the anxiety study, self-reported anxiety ratings were obtained from 12 females and 7 males who swam at a beach resort for 20–30 minutes in the presence of wild dolphins (the experimental group). The control group consisted of 13 females and 8 males who swam at the same beach after the dolphins left.

In the well-being study, participants completed a self-report questionnaire designed to examine their perceived levels of well-being immediately before and after their swim. Psychological well-being was defined as how “positive” each participant felt at the moment. Physiological well-being was defined as how energetic each participant felt. The participants reported these feelings on a scale of 1 to 100 for each questionnaire item. In the anxiety study, participants completed the “state” component of the State-Trait Anxiety Inventory immediately before and after their swim.

Webb and Drummond found that, in the well-being study, the experimental group rated their well-being significantly higher than the control group prior to the *treatment*. This finding was presumably due to the positive anticipation of swimming with dolphins in the experimental group. The well-being ratings of both groups increased post-treatment, but ratings were significantly higher for the experimental group than the control group. However, the authors found, following an analysis of covariance, that the pre-treatment difference between the groups accounted for the differences that persisted after swimming. Therefore, there was no significant effect of swimming with dolphins on well-being. In the anxiety study, there were no significant pre-treatment differences in anxiety between the experimental and control groups. The authors found a significant decrease in self-reported anxiety in the experimental group, but not in the control group. From these results, Webb and Drummond concluded that anticipation of a new and exciting experience, and swimming itself, increase well-being. In addition, they concluded that swimming specifically with dolphins may lower anxiety.

Because, of the two studies, only the anxiety study revealed significant differences between the experimental and control groups, we focus on the anxiety study here. However, all of the threats to validity in the anxiety study also apply to the well-being study. In the anxiety study, participants swam in the ocean during a visit by wild dolphins. Prior to swimming with the dolphins, the participants were told that dolphins appeared at the site on a daily basis. Therefore, the participants were led to anticipate a swim with dolphins as they waited on the beach along with members of the public who were also hoping for a dolphin encounter. The control group swam in the same area after the dolphins left and did not expect to see dolphins because they were swimming after the period when dolphins usually arrived.



Nevertheless, there are several problems with comparisons of the experimental and control groups. The authors acknowledged that there was no limit on the number of people allowed in the water while the dolphins were present. Given that they reported groups of people waiting on the beach to swim with the dolphins, it seems plausible, if not likely, that there were more people in the water accompanying the experimental group than the control group. Moreover, there was no apparent effort to make the experimental and control group comparable on that dimension. Therefore, decreased anxiety in the experimental condition could have been attributable to interacting with more people.

In addition, if all the participants were swimming together during the dolphin encounter, then it is likely that they experienced the emotional contagion of positive feelings and excitement that accompany seeing dolphins. This implies that the “therapeutic agent” in this condition may have been the experience of swimming with other happy and excited people. The reported effects may have had nothing to do with dolphins *per se*. Because the authors did not control for this confounding factor (or, at least did not offer evidence that they did), we cannot conclude that the significant decrease in anxiety in the experimental group was due to the presence of the dolphins. Moreover, the authors offered no information concerning how participants interacted with the wild dolphins or even if any interacted directly with a dolphin. Because this information was not provided, there is no means of determining to what “treatment” the experimental group was actually exposed.

As in Antonioli and Reveley (2005), participants were not blind to condition and were asked to self-report on their anxiety level. Consequently, it is not possible to rule out either demand characteristics or informant bias. Informant bias based on effort justification may have played a significant role because the participants were visitors to a local marine park who paid to participate in a dolphin swim experience; moreover, some waited six months for the opportunity. If the participants in the experimental group were those who waited longer for their opportunity to participate than those in the control group, then bias due to effort justification becomes possible.

## Conclusions and Summary

In summary, the abundance of serious threats to validity in the five studies we examined renders each of their conclusions questionable at best, and entirely unwarranted at worst. All of the studies are vulnerable to nonspecific effects, including placebo and novelty effects, and construct confounding. In addition, each of the studies contained several other methodological weaknesses that render the conclusions doubtful.

However, these five studies varied considerably in methodological rigor. Of the studies reviewed, Antonioli and Reveley (2005) is the most methodologically rigorous, as it includes controls for more potential confounds than the other studies. This study implemented a number of commendable methodological features, including randomized assignment of participants to conditions, pre- and post-treatment blind raters of outcomes, procedures to minimize demand characteristics, the use of standard validated assessment instruments (the Hamilton Depression Rating Scale, the Beck Depression Inventory, and the Zung Self-Rating Anxiety Scale), and the application of appropriate statistical tests. In this respect, Antonioli and Reveley’s study is clearly methodologically superior to Nathanson et al. (1997) and Nathanson (1998). However, even the laudable attempts of Antonioli and Reveley do not eliminate the presence of important validity threats, such as resentful demoralization and informant bias, nor the ever-present problems of nonspecific effects and construct confounding. Therefore, Antonioli and Reveley (2005) falls short of providing a valid test of DAT efficacy and stands as an important reminder of how far the DAT literature must progress to meet minimal standards of methodological quality.

Of the remaining four studies, none come close to the methodological quality of Antonioli and Reveley. Most of these studies are plagued by potential experimenter expectancy effects that could have been eliminated by the inclusion of raters blind to experimental condition. In addition, these studies were bedeviled by a host of other threats to validity (e.g., history, maturation, testing, regression),

as well as a paucity of reported information concerning basic methods and procedures. Moreover, none of the studies included a follow-up assessment of the reported short-term improvements.

Most problematic are construct confounding and nonspecific effects (including placebo and novelty effects), both of which appear to be ubiquitous in the DAT literature. None of the five studies incorporated adequate controls for these two validity threats. Minimization or elimination of construct confounding would require that both the experimental group and the control group be exposed to the same or at least highly similar procedures and stimuli with only the key ingredient—the dolphin *per se*—as the differential treatment component between groups.

Placebo effects can be minimized or controlled by a blind study in which participants are not afforded any information that would provide them with clues regarding their assignment to treatment condition. Novelty effects can be controlled for by exposure of the control group to another novel, attractive animal (e.g., a horse, dog, or another aquatic mammal), while keeping as many other variables as possible equal. Therefore, before concluding that DAT possesses specific efficacy that could not be attained with a variety of alternative treatment conditions, authors must offer sufficient evidence that the same results would not be achieved with another large, charismatic, attractive animal used in the same procedures. If the proper tests were conducted and the control and treatment groups were found to improve equally, one would need to conclude that there is nothing special or necessary about the dolphin *per se* in DAT. If the DAT group improved significantly more than the control group, then perhaps DAT could be considered a potentially therapeutically interesting intervention, although pragmatic issues of accessibility, expense, and risk would remain. So far, however, no studies have met this challenge and compared DAT with an appropriate control group—one exposed to identical procedures using a similar animal—within the same study.

Given that dolphins are highly attractive and interesting animals to most people, the likelihood of novelty effects raises particularly troubling concerns regarding the DAT literature. Most worrisome is the conspicuous absence of evidence for long-term improvement from DAT. Despite DAT's extensive promotion to the general public, the evidence that it produces enduring improvements in the core symptoms of any psychological disorder is nil. Occam's razor suggests that it is probably most parsimonious to interpret improvements from DAT as a temporary "feel good effect" of a highly positive and exciting experience. From this perspective, there is little reason to believe that DAT is a legitimate therapy or that it constitutes much more than entertainment.

The surprising paucity of scientific evidence for the long-term effects of DAT raises profoundly troubling ethical questions regarding its widespread use and promotion. There is abundant evidence for injuries sustained by participants in DAT programs (Frohoff and Packard 1995; Samuels and Spradlin 1995; Webster, Neil and Madden 1998). Moreover, interactions between dolphins and humans carry a significant risk of infections and parasitism for both humans and dolphins (Geraci and Ridgway 1991). Therefore, DAT poses important ethical questions from the standpoint of human and captive dolphin welfare. At the very least, we believe that DAT practitioners should be required to inform parents and, when relevant, participants, of the absence of evidence for DAT's enduring effects on psychological symptoms. Only then can consumers of DAT make adequately informed decisions regarding the costs and benefits of this unsubstantiated intervention.

## References

- Antonoli, C. and Reveley, M. A. 2005. Randomized controlled trial of animal facilitated therapy with dolphins in the treatment of depression. *British Medical Journal* 331: 1231–1234.
- Brensing, K., Linke, K. and Todt, D. 2003. Can dolphins heal by ultrasound. *Journal of Theoretical Biology* 225: 99–105.
- Cook, T. D. and Campbell, D. T. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston, MA: Houghton Mifflin.
- Frohoff, T. G. and Packard, J. M. 1995. Interactions between humans and free-ranging and captive bottlenose dolphins. *Anthrozoös* 8: 44–54.

- Geraci, J. R. and Ridgway, S. H. 1991. On disease transmission between cetaceans and humans. *Marine Mammal Science* 7: 191–194.
- Humphries, T. L. 2003. Effectiveness of dolphin-assisted therapy as a behavioral intervention for young children with disabilities. *Bridges: Practice-Based Research Synthesis* 1: 1–9.
- Iikura, Y., Sakamoto, Y., Imai, T., Akai, L., Matsuoaka, T., Sugihara, K., Utumi, M. and Tomikawa, M. 2001. Dolphin-assisted seawater therapy for severe atopic dermatitis: an immunological and psychological study. *Archives of Allergy and Immunology* 124: 389–390.
- Kazdin, A. E. 1994. Methodology, design, and evaluation in psychotherapy research. In *Handbook of Psychotherapy and Behavior Change*. 4th edn, 19–71, ed. A. E. Bergin and S. L. Garfield. New York: Wiley.
- Kazdin, A. E. and Wilson, G. T. 1978. *Evaluation of Behavior Therapy: Issues, Evidence and Research Strategies*. Cambridge, MA: Ballinger.
- Kendall, P. C. and Norton-Ford, J. D. 1982. Therapy outcome research methods. In *Handbook of Research Methods in Clinical Psychology*, 429–460, ed. P. C. Kendall and J. N. Butcher. New York: John Wiley and Sons.
- Lukina, L. N. 1999. Influence of dolphin-assisted therapy sessions on the functional state of children with psychoneurological symptoms of diseases. *Human Physiology* 25: 676–679.
- Marino, L. and Lilienfeld, S. O. 1998. Dolphin-assisted therapy: flawed data, flawed conclusions. *Anthrozoös* 11: 194–200.
- Nathanson, D. E. 1998. Long-term effectiveness of dolphin-assisted therapy for children with severe disabilities. *Anthrozoös* 11: 22–32.
- Nathanson, D. E., de Castro, D., Friend, H., and McMahon, M. 1997. Effectiveness of short-term dolphin-assisted therapy for children with severe disabilities. *Anthrozoös* 10: 90–100.
- Samuels, A. and Spradlin, T. 1995. Quantitative behavioral study of bottlenose dolphins in swim-with-the-dolphin programs in the United States. *Marine Mammal Science* 11: 520–544.
- Servais, V. 1999. Some comments on context embodiment in zootherapy: the case of the Autodolfijn project. *Anthrozoös* 12: 5–15.
- Shadish, W. R., Cook, T. D. and Campbell, D. T. 2002. *Experimental and Quasi Experimental Designs for Generalized Causal Inference*. Boston: Houghton: Mifflin.
- Shaughnessy, J. J. and Zechmeister, E. B. 1994. *Research Methods in Psychology*. New York: McGraw-Hill.
- Webb, N. L. and Drummond, P. D. 2001. The effect of swimming with dolphins on human well-being and anxiety. *Anthrozoös* 14: 81–85.
- Webster, L. S., Neil, D. T. and Madden, C. A. 1998. Dolphin-initiated inter- and intra-specific contact and aggression during provisioning at Tangalooma. Special Topic report, Department of Geographical Sciences and Planning and School of Marine Science, The University of Queensland.
- Whale and Dolphin Conservation Society. 2006. Dolphin therapy in the headlines. <<http://www.wdcs.org.au>>