Received: 17 July 2020 | Revised: 11 December 2020 | Accepted: 20 December 2020

DOI: 10.1002/jclp.23110

REVIEW ARTICLE

WILEY

Third time's the charm or three strikes you're out? An updated review of the efficacy of dolphin-assisted therapy for autism and developmental disabilities

Lori Marino[1] | Scott O. Lilienfeld[2]

[1]Whale Sanctuary Project, Kimmela Center for Animal Advocacy, Kanab, Utah, USA

[2]Department of Psychology, Emory University, Atlanta, Georgia, USA

Correspondence
Lori Marino, Whale Sanctuary Project, Kimmela Center for Animal Advocacy, Kanab, UT 84741, USA.
Email: Marinolori@outlook.com

Abstract

Context: Dolphin-assisted therapy (DAT) is a popular form of animal-assisted therapy for autism spectrum disorders and other psychological conditions.

Objective: In this review, our third, we analyze the most recent DAT studies in terms of construct and internal validity criteria to determine if there is empirical support for DAT.

Method: To ensure a systematic review, we searched for peer-reviewed studies on DAT by submitting relevant search terms to *Google Scholar* from 2007 to 2020, conducted a further search of all DAT papers in several peer-reviewed journals, and reviewed reference sections of DAT articles to ensure a thorough review of the literature between 2007 and the present.

Results: The DAT literature continues to be marked by several weaknesses in both internal and construct validity that preclude confident inferences regarding the intervention's efficacy.

Conclusion: There is still insufficient evidence that DAT has therapeutic value.

KEYWORDS
animal, autism, construct validity, dolphin, internal validity

1 | INTRODUCTION

Dolphin-assisted therapy (DAT) is a popular form of animal-assisted therapy (AAT) in which live captive dolphins (usually bottlenose dolphins) are involved in a therapeutic session with children or adults who suffer from various conditions, especially psychological and neurological disorders. DAT is a form of AAT, a broader, goal oriented, planned, and structured intervention directed and/or delivered by health, education, or human service professionals, including psychologists and social workers. In AAT, intervention progress is measured (https://iahaio.org/wp/wp-content/uploads/2020/07/iahaio_wp_updated-2020-aai-adjust-1.pdf). DAT engages the patient in traditional verbal, motor, or attentional tasks in the presence of a dolphin. The involvement of the dolphin ranges from situations in which the patient simply sits poolside observing the dolphin, to touching or feeding the dolphin, to entering the water and swimming with or being pulled around by the dolphin. Although DAT is advertized as a therapy for a wide range of conditions, from depression (Antonioli & Reveley, 2005), Down's syndrome (Griffioen & Enders-Slegers, 2014), and encephalopathy (Kohn & Oerter, 2013) to atopic dermatitis (Ikura et al., 2001), it continues to be most often applied to individuals identified as having autism spectrum disorders (ASD), commonly known as autism.

DAT originated with the work of anthropologist Betsy Smith at Florida International University in 1971 (Marine Connection, 2009), who later disavowed it. DAT programs are offered across the world, including Europe, the Middle East, Asia, United States, the Caribbean, Mexico, Israel, Bahamas, and South America. These programs typically involve travel to a facility from the patient's home and are extremely expensive, with prices in the thousands of dollars, which may cause patients and their families to make a choice between DAT and more effective, empirically based treatment options. The length and frequency of sessions varies across facilities but typically involves several sessions over a week to a month. Therefore, participants and their families are often required to be away from home for an extended period.

The International Association of Human-Animal Interaction Organization (IAHAIO) disallows the use of wild animals in human-animal interactions and explicitly discourages direct human-dolphin interactions (https://iahaio.org/best-practice/white-paper-on-animal-assisted-interventions/), as they pose a risk of injury and disease transmission. Moreover, DAT promotes the confinement of wild animals, namely, dolphins, coerced to interact with humans in artificial tanks, often contributing to poor well-being and shortened lives for the dolphins themselves (Stewart & Marino, 2009). These caveats notwithstanding, DAT continues to be offered globally. The average consumer may find it difficult to distinguish DAT from the ubiquitous "swim-with-dolphins" offerings at entertainment parks and tourist attractions. These programs are superficially similar not only in the way they are conducted but, in the claims, made by the entertainment facilities themselves, which in both cases promote the mythology surrounding dolphins' purported healing properties.

Even though the practice of DAT has expanded globally, relatively few peer-reviewed scientific articles have addressed its efficacy as a treatment for ASD or other disorders. In 1998, we published a methodological review (Marino & Lilienfeld, 1998) of the available peer-reviewed DAT literature at the time, focusing on several papers by Nathanson and his colleagues (Nathanson, 1998; Nathanson & de Faria, 1993; Nathanson et al., 1997). We found that a plethora of serious threats to various types of validity and flawed data analytic methods rendered these studies essentially uninterpretable and unable to support the authors' strong claims regarding the effectiveness of DAT. Nine years later, in 2007, we revisited this literature by conducting a review of the five DAT papers published between 1998 and 2007. Although we found some improvement in the methodological rigor of some of the studies, we again reported serious threats to validity in all of them. We concluded that the aggregate findings were again difficult to interpret at best.

Now, more than a decade after our second paper, we return to the question of the validity of DAT with an updated methodological review of the findings since 2007. Because all programs of research take time to develop, we argue that a detailed analysis of the validity of these more recent articles is essential for ascertaining whether and which DAT claims have received more compelling empirical support over the past decade and a half. Hence, we

pick up where we left off in 2007 and apply the same foundational criteria for assessing validity as we have used in our two previous reviews. We complement this analysis with a detailed narrative review of each study.

1.1 | What are the theoretical rationales for DAT?

Proponents of DAT have advanced a variety of explanations for its purported effectiveness. One feature shared across most or all of these explanations is that they are typically couched in language that is designed to sound scientific, even when the scientific foundations of these explanations may be weak.

One of the most widespread and longstanding claims of DAT is that echolocation (commonly known as sonar), the means by which dolphins use sound to perceive their world, possesses special curative powers. Many facilities suggest or directly claim in their online promotions that the sounds made by dolphins possess healing properties (e.g., https://balidolphintherapy.com/; https://www.upledgerclinic.com/therapies/dolphin-assisted.php). Echolocation is the use of high frequency reflected sound waves to perceive objects and the surrounding environment. The assumption underlying this hypothesis is that when a dolphin echolocates on a person in the water, the clicks exert a mechanical effect on tissue, changing the body and healing it (Birch, 1997; Cole, 1996). Nevertheless, there is no scientific support for this effect. The dolphins used in DAT do not always echolocate on the patient and, if they do, the ways in which they use echolocation are inconsistent with any known mechanisms for how ultrasound is used to heal the body. They are also not consistent with any known mechanism for treating either the core or ancillary symptoms of neurodevelopmental disorders such as ASDs (Brensing et al., 2003).

Another common claim regarding the efficacy of DAT is that the technique is correlated with changes in brain waves as measured by electroencephalography (EEG). The few studies using EEG suggest that, at best, DAT may produce a nonspecific relaxation effect (Brensing et al., 2003). One study revealed "increased interhemispheric coherence" with DAT in children with ASD, but it is unclear how the changes were related to any specific therapeutic effect for that condition (Ortiz-Sánchez et al., 2018). Even if some elements of DAT were found to produce a temporary relaxation effect in some patients, it is unclear how such a finding should be interpreted given that there is no evidence that a core or ancillary deficit in ASDs is an inability to relax.

Some proponents of DAT suggest that it increases attention to external stimuli. Best known for this idea are Nathanson and co-authors, who suggested in early articles that the fundamental problem in people with a wide range of disorders (e.g., autism, Cri-du-chat syndrome, cerebral palsy, and tuberous sclerosis) is their inability to attend to the important details of stimuli and that DAT operates by addressing this deficit (Nathanson et al., 1997; Nathanson, 1998). Nevertheless, this hypothesis is inconsistent with our current understanding of most of the disorders afflicting the participants in their studies. For example, there is no evidence that the core deficits in neurological disorders such as cerebral palsy and Cri-du-chat syndrome are attentional in nature. Moreover, although there is a link between attentional deficits and autism, such as circumscribed attention to specific objects (Sasson et al., 2011), the minimal nature of the intervention of swimming with dolphins strains credulity as an adequate treatment for such a profound, complex, and lifelong disorder as ASD. Despite their claim that attention deficits underlie their participants disabilities and the effectiveness of DAT, Nathanson et al. (1997) did not assess attention in their participants as a mediator variable or examine whether improvements in attention were correlated with improvement on the dependent measures. Their claim that DAT is effective for ASD because it increases attention lacks compelling empirical support.

Other theoretical assertions regarding DAT range from biophilia (the implied positive effects of being in contact with nature; Antonioni & Reveley, 2005; Yerbury & Boyd, 2018) to the calming effects of being in water (Granger & Kogan, 2000; McKinney et al., 2001) to a deep emotional connection between dolphins and people that brings about healing (DeMares, 2000; Dobbs, 2000; McKinney et al., 2001). Nevertheless, all of these explanations are so nonspecific that they provide little or no practical insight into the mechanisms by which DAT is allegedly helpful for the features of ASD or other disabilities.

Several general reviews and analyses of AAT published since our 2007 review have noted the lack of substantive support for the claims made by the DAT industry (Burton, 2013; Cole, 2009; Friesen, 2010; Herzog, 2014; May et al., 2016; Serpell et al., 2017). A review of the DAT literature by Fiksdal et al. (2012), incorporating the findings of Marino and Lilienfeld (1998, 2007), reiterated our earlier conclusion that "Overall, research studies need to be better designed and threats to validity must be addressed before we deem DAT as an effective intervention for any population" (p. 7). Fiksdal et al. (2012) did not review any of the empirical DAT articles between 2007 and 2012 and based their conclusions on studies analyzed by Marino and Lilienfeld (1998, 2007). Hence, their summary does not overlap in coverage with the present review.

In summary, even setting aside the crucial question of DAT's effectiveness, there remains no scientifically plausible mechanism for such effectiveness, whether for using dolphins or any other animals. Moreover, none of the theoretical proposals is linked with current knowledge regarding the pathogenesis of ASD and related neurodevelopmental disorders. Apart from the lack of a compelling theoretical foundation, there remain significant unanswered questions concerning the validity and methodological soundness of DAT therapeutic claims. In this article, we revisit the peer-reviewed literature on DAT since 2007, analyze the methodological strength of these studies and determine whether there is any substantial evidence that DAT is an efficacious treatment for ASD and other developmental disorders. In this third, updated review, we pick up where we left off in 2007, providing a methodological analysis of specific threats to construct and internal validity in the DAT literature published since Marino and Lilienfeld (2007).

2 | METHODS

To ensure a systematic review, 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, allowing us to capture almost all of the peer-reviewed literature in this area. Scientometric analyses suggest that Google Scholar is the most comprehensive of all major scholarly internet search engines (Gusenbauer, 2019). Second, as a further precaution, we conducted a comprehensive search of all papers on DAT from 2007 to April, 2020 in the following peer-reviewed journals: *Anthrozoös*, *Society & Animals*, *Applied Animal Behaviour Science*, and *Zoo Biology*, where most DAT papers appear, and, to cover the clinical literature beyond the *Google Scholar* Search, *Review of Autism and Developmental Disorders*, *Autism Research and Treatment*. Third, we reviewed the reference sections of those articles identified through the aforementioned means to ensure that we had not missed any relevant studies.

Our inclusion criteria for studies were those that (1) entailed interaction with a live dolphin and (2) measured clinically relevant outcomes, such as attention or depression. Our exclusion criteria for studies were (1) lack of a condition in which a live dolphin interacted with a clinical sample of participants, (2) lack of established outcome measures directly relevant to improvement specific to the participants' condition, and (3) no conclusions claiming support for DAT as a treatment for the conditions represented in the study's sample. Because of exclusion criterion (1), we did not examine virtual reality (VR) studies that employed images of dolphins. Several studies (e.g., Veling et al., 2017) have explored the impact of VR on stress and other psychological symptoms. Nevertheless, we do not regard these studies, even those presenting dolphins or other animals, as equivalent to DAT or other interventions involving actual interactions with a live animal. There are many reasons (e.g., novelty effects; see below) why VR displays might be helpful to participants, but we consider that a separate question from DAT, as there are several qualitatively different experiential components to VR versus reality.

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

of validity: construct validity, that is, the extent to which effects are due to the intended treatment, and internal validity, that is, the soundness of the hypothesized causal inferences. The threats to construct validity we assessed were:

- placebo effects, that is, improvement due to expectation;
- novelty effects, that is, improvement due to the energizing influence of a new experience;
- construct confounding, that is, failure to take into account the fact that the treatment may include more than one active ingredient;
- demand characteristics, that is, tendency to respond to items in accordance with the perceived hypothesis; and
- experimenter expectancy effects, that is, tendency for the experimenter to unintentionally bias results in accordance with the hypothesis.

The threats to internal validity assessed were:

- history, that is, the occurrence of one or more life events during the study that could affect the study's outcome variables, such as a major storm during treatment;
- maturation, that is, changes over time due to natural developmental changes;
- multiple intervention interference, that is, administration of treatments other than the intended treatment during the course of the study;
- informant retrospective bias, that is, tendency of informants to selectively recall improvement in accord with expectations or effort justification, that is, the need to rationalize one's investment in the treatment owing to the time, energy, and resources expended; and
- differential attrition, that is, comparison groups become different because of uneven drop-out across the groups.

We identified eleven empirical peer-reviewed studies of DAT published from 2007 to the present. One was a feasibility study of recording EEG in water and did not purport to assess the efficacy of DAT (Johannes et al., 2016) and was therefore not included in our analysis. Another, by Lu et al. (2017), employed "virtual dolphins" on a gaming device. This study was not comparable to studies using live dolphins and was, therefore, also excluded. Another, by Salgueiro et al. (2012), was an exploratory study and the only one of the investigations reporting no evidence for improvement in autism with DAT. Although the findings of Salgueiro et al. (2012) do not offer support for the efficacy of DAT, it is subject to many of the same methodological weaknesses identified in those studies supporting the effectiveness of DAT, including placebo effects, multiple treatment interference, and other confounds.

Another study, by Homma et al. (2011), examined the effect of touching dolphins on trait anxiety and brain waves in a small sample of six healthy children. Thus, their findings may not be generalizable to clinical populations. The study also lacked a control group. Finally, another exploratory study employed EEG measurement of inter-hemispheric coherence after children with ASD interacted with dolphins compared with a no-intervention control group. The authors found an increase in alpha coherence in the anterior frontal lobes in the former group (Ortiz-Sánchez et al., 2018). However, there were no other outcome measures for ASD and, as the authors acknowledged, no way to determine if the dolphin intervention yielded any beneficial psychological effects.

The remaining six studies all met the inclusion criteria of using live dolphins and reporting evidence for treatment efficacy of DAT in a clinical sample using established outcome measures relevant to assessing improvement in the participants. These were: Breitenbach et al. (2009), Dilts et al. (2011), Griffioen and Enders-Slegers (2014), Griffioen et al. (2019), Kohn and Oerter (2013), and MdYusof and Chia (2012); we later review each study in greater detail. Table 1 displays the participant characteristics and outcome measures for each of these six studies.

TABLE 1 Participant characteristics and assessment measures in each of the six evaluated studies

Study	Participants	Measures
Breitenbach et al. (2009)	*N* = 94 (F and M), 5–8 years age autism, Down syndrome mental and physical disabilities	Parent and staff surveys, video observations of verbaland social behavior
Dilts et al. (2011)	*N* = 37 (19 F, 18 M), children with range of diagnoses, e.g., autism, developmental delay, cerebral palsy, epilepsy, etc. most with multiple diagnoses.	Parent reports using Behavior Dimensions Rating Scale (BDRS)
Griffioen and Enders-Slegers (2014)	*N* = 45 (19 F, 26 M), 7–11 yrs age Down syndrome	Parent reports using Matson Evaluation of Social Skills for Individuals with Severe Retardation (MESSIER)
Griffioen et al. (2019)	5 case studies (1 F, 4 M) 6–8.5 years age, autism spectrum disorder	Student raters of behavioral categories from videos
Kohn and Oerter (2013)	US sample: *N* = 162 (76 F, 86 M), 2–30 years age Israeli sample: *N* = 31 (9 F, 22 M) 5–23 yrs age	Parent surveys and teacher-therapist surveys of a range of behavioral and physical competencies
	Wide range of diagnoses, e.g., autism, general mental retardation, ADHD, Downs syndrome, cerebral palsy, etc.	
MdYusof and Chia (2012)	*N* = 15 (5 F, 10 M), 9–10 yrs autism	Gilliam Autism Rating Scale (GARS) administered by researchers

3 | RESULTS

The six studies assessed in this review vary in methodological quality, but all are characterized by weaknesses that relate to different aspects of validity. Table 2 displays several threats to validity, their definition, and whether they are present in each of the six studies. Most of these threats relate to either *construct validity*, that is, the extent to which effects are due to the intended treatment, or *internal validity*, that is, the methodological soundness of the study. In the interest of space, we limit ourselves to the most serious threats to construct and internal validity.

3.1 | Construct validity

3.1.1 | Nonspecific effects

Nonspecific effects are improvements from influences that are not distinctive 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). We found several nonspecific effects in most of the studies assessed. Three relevant subcategories of nonspecific effects are placebo effects, novelty effects and demand characteristics.

The placebo effect is the well-documented but little understood improvement that derives from participants' expectation of improvement (Linde et al., 2011).

DAT is especially vulnerable to placebo effects in part because this intervention is widely marketed and available online as highly efficacious and in part because the nature of the treatment is evident to participants. Proper control conditions would eliminate or substantially minimize this effect by eliminating cues to treatment

TABLE 2 Major threats to validity in each of the six studies

Threat	Definition	Study					
		1	2	3	4	5	6
Construct Validity	**Nonspecific effects not due to treatment**						
Placebo effects	Improvement due to expectation	✓	✓	✓	✓	✓	✓
Novelty effects	Improvement due to energizing impact of new experiences	✓	✓		✓	✓	✓
Construct Confounding	Failure to take into account the fact that the treatment may include more than one active ingredient	✓	✓	✓	✓	✓	✓
Demand Characteristics	Tendency to respond to items in accordance with perceived hypothesis	✓	✓	✓	✓	✓	✓
Experimenter Expectancy Effects	Tendency for experimenter to unintentionally bias results in accordance with the research hypothesis	✓		✓	✓	✓	✓
Internal Validity	**The soundness of the causal inferences**						
History	Occurrence of potentially therapeutic events other than the intended treatment during the course of the study		✓			✓	✓
Maturation	Changes over time due to natural development				✓	✓	
Multiple Intervention Interference	Administration of treatments other than the intended treatment during the course of the study		✓		✓	✓	✓
Informant Retrospective Bias	Tendency of informants to selectively recall improvement in accord with expectations or effort justification	✓	✓	✓		✓	✓
Differential Attrition	Comparison groups become different because there is uneven drop out across the groups		✓			✓	

Key: Study 1 = Breitenbach et al. (2009), Study 2 = Dilts et al. (2011), Study 3 = Griffioen and Ender-Slegers (2014), Study 4 = Griffioen et al. (2019), Study 5 = Kohn and Oerter (2013), Study 6 = MdYusof and Chia (2012).

condition. None of the six studies included appropriate procedures to do this. A proper control for placebo effects might include adding another session with a charismatic animal (see below) to partially conceal the one treatment being proposed to work.

Novelty effects are the general energizing and uplifting effects of a new, exciting experience. Because of its novel nature, DAT is particularly prone to this nonspecific effect because of the obviously new and exciting experience of swimming with or interacting with a large, intelligent, and charismatic animal. As with placebo effects, a 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 reasons to believe that they had received

the relevant treatment, and both would be subject to the excitement of interacting with an exotic animal. Apart from Griffioen and Ender-Slegers (2014), none of the studies included a control for novelty effects.

Demand characteristics refer to the tendency of participants to respond according to their perceptions about the hypothesis or intention of the study. Several methodological procedures can be used to minimize, although not eliminate, demand characteristics. These include the use of standardized validated assessment instruments designed to detect responses to demand characteristics and the use of independent raters of the participant's performance, that is, not parents or caregivers, who presumably have less of a stake in the children's improvement. All of the six studies we reviewed lacked sufficient controls for demand characteristics for some of the same reasons they were subject to placebo effects.

At the very least, to identify and understand nonspecific effects, 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 could 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 experiences.

3.1.2 | Construct confounding

Construct confounding occurs when there is a failure to account for 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 and other forms of positive reinforcement 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. Although there is no single, ideal control for DAT, no DAT study has included an adequate subset of the many comparison groups that would be required for even a minimally effective dismantling strategy. DAT involves a complex set of components that include, among others, the water, the animal, the therapist(s), and each of these components in turn consists of multiple subcomponents. No one study can dismantle all of the components in DAT, but few DAT studies have provided a control for even the factors that are the least difficult to control, such as the aquatic environment. Indeed, only two of the six studies we analyzed attempted to control for the aquatic environment (see below).

3.2 | Internal validity

3.2.1 | Informant retrospective bias

Among the many threats to internal validity present in several of the studies we assessed, the most prevalent was informant retrospective bias, that is, the tendency of informants (participants or relatives) to selectively recall improvement in accord with expectations or effort justification. This form of bias is one to which DAT is particularly vulnerable given that many, if not most, of the participants in these studies are children with parent respondents who presumably have a strong need to find an effective treatment for ASD and other disorders. Given that many DAT practices involve considerable expense and time commitment on the part of the families of

participants, effort justification is a worrisome possibility. In the case of ASD, the emotional desire to "connect" with one's child can leave parents especially vulnerable to effort justification. Except for Griffioen et al. (2019), the other five studies assessed neglected to adequately control for informant retrospective bias.

3.3 | Strengths and weaknesses of each study

3.3.1 | Study 1

This study (Breitenbach et al., 2009) was designed to determine whether DAT is an effective therapeutic intervention for children with social and communicative disabilities. The authors studied 94 participants aged 5–8 years with a wide range of disabilities, including autism and Down's Syndrome, using parent and staff surveys and video observations as dependent measures in a complex pre-post design with four treatment groups. The authors went to considerable lengths to provide the proper controls for strong internal and construct validity. As such, this study is more methodologically rigorous than most of the others in our review. For instance, the authors controlled for the effects of an aquatic environment as a potential confounding variable.

The findings on which they based their overall conclusions ranged from not significant to marginally significant. Yet, the authors portrayed these highly mixed results as positive, in several cases reporting nonsignificant findings as significant. Also, statistical interactions across treatment groups were not evaluated, making it impossible to ascertain whether the experimental group outperformed the control groups. Finally, the authors tried to, but fell short of, controlling for important nonspecific effects. For instance, it is not clear how much the positive findings were due to the inclusion of the parent in the treatment. Moreover, the samples were not well matched, by the authors' own admission, for the dolphin versus farm animal conditions, making it difficult to interpret and compare the results of these two conditions.

3.3.2 | Study 2

This is a pilot study (Dilts et al., 2011) conducted to assess the effectiveness of a 2-week DAT program at a dolphin swim facility in the Ukraine for 37 children with a wide range of developmental and neurological disorders, which encompassed multiple diagnoses. The authors reported positive behavior change in some of the subscales used in the outcome measure, that is, the Behavior Dimensions Rating Scale Parent Report Form (Bullock & Wilson, 1989), but acknowledged that the study possessed major methodological weaknesses, including an absence of control groups. The study was also vulnerable to differential attrition, as 18 out of 37 participants did not respond to all scoring items and were excluded. In addition, there was a risk of multiple intervention interference as each subject received a relaxation regime called "alpha sphere," which confounded potential effects of the experimental treatment. Moreover, the study was subject to informant retrospective bias, as the outcome measure was completed by parents following the intervention. Dilts et al. (2011) concluded that DAT may result in positive behavior change in children with a range of conditions but conceded that their study leaves this question unanswered.

3.3.3 | Study 3

This study (Griffioen & Enders-Siegers, 2014) investigated the effects of DAT on the development of speech and social behavior in 45 children, Ages 7–11 years, with Down's Syndrome. They used the Matson Evaluation of Social Skills for Individuals with Severe Retardation (Vermeulen, 2001) as the dependent measure. This semi-crossover study employed several conditions to control for nonspecific effects. For instance, to dismantle variables in the

DAT condition, participants in a control group played with a remote-controlled boat in a pool, although the authors acknowledged that the use of another live animal might be a more appropriate control for the dolphin. However, the participants were not randomly assigned to groups, thereby weakening the construct validity of the study. Random assignment of participants to study groups is critical to determine if outcome differences across groups are due to the intervention rather than to pre-existing differences in the participants themselves. Furthermore, placebo effects and demand characteristics could not be ruled out.

The authors reported that the results were mixed across the dependent measures, with significant changes in three out of five variables. These were "verbalization," "recognition of persons," and "impulsiveness." In follow-up studies, improvement in only one of the factors, namely, "verbalization," persisted.

3.3.4 | Study 4

This article comprised five case studies investigating the synchrony in conversations (i.e., turn-taking) between a therapist and five children with ASD during DAT. Although this study (Griffioen et al., 2019) was partly exploratory, the authors set out to not just characterize turn-taking but to explicitly determine whether DAT exerted an impact upon it; hence, it was partly confirmatory as well. They noted that therapies that boost synchrony can increase social skills, which can also be helpful for children with ASD. The authors concluded that "the results of this study seemed only partially in line with our expectations that synchronization of verbalizations…would improve during DAT" (p. 10).

Nevertheless, due to the absence of any experimental controls across these five case studies, there is no way to evaluate the validity of the findings. Moreover, the results were mixed, as turned-taking seemed to increase over time, but only for children who initially exhibited reasonably good verbal skills to begin with.

3.3.5 | Study 5

This study (Kohn & Oerter, 2013) examined whether DAT improved scores on a range of dependent measures of cognitive and behavioral function in a sample of 162 youth/young adults with a wide range of disabilities undergoing two very different DAT regimes in two different locations—Eilat, Israel, and Key Largo, Florida. This was the weakest of the experimental studies we assessed in terms of internal and construct validity, as it met none of the important methodological criteria for these forms of validity. Furthermore, the authors' reasoning behind their conclusions was problematic. Several dependent measures of improvement were assessed by parents and by teachers. Nevertheless, no controls for demand characteristics, experimenter expectancy effects, informant retrospective bias, that is, the tendency of informants or participant representatives to misremember improvement in accord with their desires or effort justification, or other factors were implemented. Because little information concerning the specific procedures at the two DAT locations was reported, the possibility of multiple intervention interference cannot be ruled out. More importantly, there is no way to systematically compare the results from the two locations.

Moreover, the authors reasoned that if improvements in some of the dependent measures were similar across very different conditions (Eilat vs. Key Largo) then they have demonstrated that the dolphin is the key therapeutic agent in DAT given that the dolphin is "the only constant factor in the therapy" (p. 14),. But given that there were no formal controls in either of the DAT locations, this conclusion is unwarranted. In fact, formal data from the Eilat program were not reported. Instead, the authors merely stated that the findings between the two locations were similar. For comparisons of disorder groups, no statistical interactions were tested. Furthermore, there was pronounced differential attrition, only 162 out of 220 participants returned responses returned and no data were

reported on the nonresponders. These methodological flaws render the conclusions of Kohn and Oerter (2013) extremely difficult to interpret.

3.3.6 | Study 6

In this study (MdYusof & Chia, 2012), 15 children, aged 9–10 years, with ASD participated in DAT over 12 months at a facility off the mainland of Singapore. Using the Gilliam Autism Rating Scale (Gilliam, 1995) in a pre/post treatment design, the authors reported a significant reduction in stereotyped behaviors as well as significant improvements in communication and social interaction. But because there were no control groups, the authors acknowledged that these findings might have been influenced by a host of confounding factors and threats to validity, such as history, maturation, differential attrition, changes in instrumentation, and novelty effects. Hence, their results are difficult to interpret with confidence.

4 | CONCLUSIONS

In this third, updated review of the DAT literature, the six studies we assessed varied considerably in methodological rigor. However, they all contained serious threats to validity, rendering each of their conclusions questionable and often unwarranted. Hence, the evidence for the efficacy of DAT appears to be no more compelling than it was in our previous two review articles.

All the studies we reviewed are vulnerable to nonspecific effects, including placebo and novelty effects, and construct confounding, which remain ubiquitous in the DAT literature. Placebo effects could be ideally minimized or controlled by a blinded study in which participants are not afforded any information that would provide them with clues regarding their assignment to treatment condition. This procedure is admittedly difficult to devise in the case of DAT but not impossible and would involve adding a control condition that resembles the actual experimental treatment. Novelty effects, too, can be at least partially controlled in this way by exposure of the control group to another new, attractive animal (e.g., a horse, dog, or another aquatic mammal), while keeping as many other variables as possible equal. Minimization or elimination of construct confounding would require a dismantling strategy in which both the experimental group and the control group(s) are 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/among groups. Likewise, there are procedures for minimizing the other threats to internal validity found in these studies.

Despite the many methodological issues plaguing the current DAT literature, we are encouraged to find that the authors in all six studies acknowledged the threats to validity we identified in our previous papers (Marino & Lilienfeld, 1998, 2007). However, only Breitenbach et al. (2009) and Griffioen and Enders-Slegers (2014) implemented ways to minimize them.

Apart from methodology, there remains a concern about the soundness of the theoretical rationales for DAT treatment. Only Breitenbach et al. (2009) and Griffioen and Enders-Slegers (2014) offered any conjectures regarding this issue. Breitenbach et al. (2009) reported that the positive effects of DAT were not dependent on the patients being in or out of the water. They drew this conclusion because they controlled for the water condition. They noted that this finding "refutes the mystical explanation of the healing effects of ultrasonic waves emitted by dolphins…" (p. 288). Griffioen and Enders-Slegers (2014) appealed to the theoretical "transmission triangle" concept of Malan (1999) in which the interaction of the subject, the dolphin, and therapist creates a three-way connection in which the communication of and toward the child is redirected via the dolphin, which motivates the child to provide correct responses to the therapist. Claims regarding attentional shifts being foundational for the success of DAT are not uncommon but lack robust supportive evidence.

Griffioen and Enders-Slegers (2014) also noted that "improved learning is not restricted to the interaction of the child with the dolphin, but it is the total of factors which determines the context of learning" (Servais, 1999, p. 578). This holistic rationale makes it difficult to falsify the notion that the dolphin is an important therapeutic ingredient (see Ruscio, 2002, for a broader discussion of the "mantra of holism" as an obstacle to evaluating the scientific status of treatments). Moreover, it does not suggest a specific mechanism by which DAT would be effective in treating ASD or other developmental disabilities.

In conclusion, our third review yields no compelling evidence for the validity of DAT as an effective treatment for ASD or other disorders. Given the present state of the evidence, it may be helpful to place DAT in perspective as part of the more general realm of AAT. Although most of the AAT literature has used cats, dogs, horses, and small animals, the broader validity issues remain the same.

Marino (2016) conducted a methodological review of the AAT literature published between 2005 and 2010, assessing 28 studies that used dogs, farm animals, birds, and horses in various therapeutic regimes, focusing on the construct validity of these studies, that is, asking whether a live animal is the key therapeutic ingredient in AAT. Marino found substantial methodological weaknesses in these studies that precluded firm conclusions regarding both the general effectiveness of AAT and, in particular, whether a live animal is necessary. In fact, one study obtained the same results with a robotic dog as with a live dog (Banks et al., 2008). A number of relatively recent meta-analyses and narrative reviews of the literature on general AAT and similar interventions (e.g., equine-assisted therapy) have yielded similar results, suggesting that any improvements from AAT are likely to be modest and temporary at best (Anestis et al., 2014; Chitic et al., 2012; Crossman, 2016; Nimer & Lundahl, 2007; Souter & Miller, 2007). Furthermore, few if any of the improvements appear to reflect changes in the core features of psychological disorders as opposed to more nonspecific and short-term changes in affect. Taken together, there is considerable uncertainty concerning the efficacy of AAT in general. Furthermore, questions remain about the importance of using a live animal and about the robustness of any potential improvements.

In another recent review of validity issues in animal-assisted interventions Crossman (in press) noted that, while the research literature is improving in quality, many of the claims made regarding the effectiveness of AAT outstrip the empirical evidence. They highlighted the lack of consistency across studies in the AAT literature stemming from inadequate sample sizes, substantial heterogeneity in the clinical characteristics of the samples within each study, and lack of evidence for long-term treatment effects, among other issues. They especially singled out weak construct validity in the AAT literature owing to inadequate control conditions.

In this review, we have shown that there remains little sound or even especially suggestive evidence that DAT is effective for ASD or other conditions. Specifically, there are substantial weaknesses in construct and internal validity in the current DAT literature. These findings underscore the need for more rigorous methodological approaches to DAT (and AAT more broadly) to determine whether these approaches to therapy are worth promoting and continuing as well as more conservative claims about its effects pending more convincing evidence.

Finally, one might argue that even if DAT does not have specific therapeutic value, it is essentially a harmless pursuit and one that might, at the very least, afford patients an opportunity to undergo an exciting experience with an attractive and beloved animal. Nevertheless, there are several costs and potential risks to DAT.

Reliance on invalidated therapeutic interventions is often a substantial opportunity cost for parents who are led to believe that DAT is effective. DAT is expensive, with most facilities charging thousands of dollars in fees for a regime that may last a few days to several weeks. This expense in time and money may be incurred at the cost of an opportunity for a more conventional empirically sound treatment. In addition, unlike dogs and guinea pigs, dolphins are large, wild animals. There are numerous reports of patients being severely injured by captive dolphins in these facilities (Stewart & Marino, 2009). Moreover, from the point of view of animal welfare, DAT studies are ethically problematic as many of these dolphins are taken from their families in the wild and forced into small concrete tanks. There, they may endure years of diminished physical and mental well-being (Marino, 2013; Stewart & Marino, 2009). In short, for both scientific and ethical reasons, we strongly suggest that patient advocacy groups,

particularly in the autism community, take special care in recommending and promoting DAT for ASD and other disorders.

DATA AVAILABILITY STATEMENT

The data in this manuscript are available in the published literature.

ORCID

Lori Marino https://orcid.org/0000-0003-4882-8192

REFERENCES

Anestis, M. D., Anestis, J. C., Zawilinski, L. L., Hopkins, T. A., & Lilienfeld, S. O. (2014). Equine-related treatments for mental disorders lack empirical support: A systematic review of empirical investigations. *Journal of Clinical Psychology*, *70*(12), 1115–1132.

Antonioli, C., & Reveley, M. A. (2005). Randomized controlled trial of animal facilitated therapy with dolphins in the treatment of depression. *British Medical Journal*, *331*, 1231–1234.

Banks, M. R., Willoughby, L. M., & Banks, W. A. (2008). Animal-assisted therapy and loneliness in nursing homes: Use of robotic versus living dogs. *Journal of the American Medical Directors Association*, *9*, 173–177.

Birch, S. (1997). *Dolphin–human interaction effects* (Doctoral thesis). Department of Electrical & Computer Systems Engineering, Monash University, Caulfield Campus.

Breitenbach, E., Stumpf, E., Fersen, L. V., & Ebert, H. V. (2009). Dolphin-assisted therapy: Changes in interaction and communication between children with severe disabilities and their caregivers. *Anthrozoös*, *22*, 277–289.

Brensing, K., Linke, K., & Todt, D. (2003). Can dolphins heal by ultrasound? *Journal of Theoretical Biology*, *225*(1), 99–105.

Bullock, L. M., & Wilson, M. J. (1989). *Behavior dimensions rating scale: Examiner's manual*. Riverside Publishing.

Burton, A. (2013). Dolphins, dogs, and robot seals for the treatment of neurological disease. *The Lancet Neurology*, *12*(9), 851–852.

Chitic, V., Rusu, A. S., & Szamoskozi, S. (2012). The effects of animal assisted therapy on communication and social skills: A meta-analysis. *Transylvanian. Journal of Psychology*, *13*(1), 1–17.

Cole, D. M. (1996). Phenomenological effect of dolphin interaction on humans. *International Symposium on Dolphin Healing*. Co-hosted by the AquaThought Foundation, 1–7.

Cole, M. (2009). *Literature review and manual: Animal-assisted therapy* (Doctoral dissertation). University of Lethbridge, Faculty of Education, Lethbridge, Alta.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Houghton Mifflin.

Crossman, M. K.. (in press). The research challenge: Threats to the validity of human-animal interaction intervention studies and suggestions for improvement. In A. Fine (Ed.), *Handbook on animal-assisted therapy: Theoretical foundations and guidelines for practice* (5th ed.). Academic Press.

Crossman, M. K. (2016). Effects of interactions with animals on human psychological distress. *Journal of Clinical Psychology*, *73*(7), 761–784.

DeMares, R. (2000). Human peak experience triggered by encounters with cetaceans. *Anthrozoös*, *13*(2), 89–103.

Dilts, R., Trompisch, N., & Bergquist, T. M. (2011). Dolphin-assisted therapy for children with special needs: A pilot study. *Journal of Creativity in Mental Health*, *6*(1), 56–68.

Dobbs, H. (2000). *Dolphin healing*. Piatkus.

Fiksdal, B. L., Houlihan, D., & Barnes, A. C. (2012). Dolphin assisted therapy: Claims versus evidence. *Autism Research and Treatment*, *839792*, 1–7.

Friesen, L. (2010). Exploring animal-assisted programs with children in school and therapeutic contexts. (2010). *Early Childhood Education Journal*, *37*, 261–267.

Gilliam, J. E. (1995). *Gilliam Autism Rating Scale (GARS)*. Pro-Ed Corp.

Granger, B. P., & Kogan, L. (2000). Animal-assisted therapy in specialized settings. In A. Fine (Ed.), *Animal assisted therapy: Theoretical foundations and guidelines for practice* (pp. 212–234). Academic Press.

Griffioen, R., van der Steen, S., Cox, R. F. A., Verheggen, T., & Enders-Slegers, M. J. (2019). Verbal interactional synchronization between therapist and children with autism spectrum disorder during dolphin assisted therapy: Five case studies. *Animals*, *9*, 716.

Griffioen, R. E., & Enders-Slegers, M. J. (2014). The effect of dolphin-assisted therapy on the cognitive and social development of children with Down syndrome. *Anthrozoös*, *27*(4), 569–580.

Gusenbauer, M. (2019). Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. *Scientometrics*, *118*, 177–214.

Herzog, H. (2014). Does animal-assisted therapy really work. What clinical trials reveal about the effectiveness of four-legged therapists. Available online: https://www.Psychology/today.com/blog/animals-and-us/201411/does-animal-assisted-therapy-really-work

Homma, A., Hara, H., Matzuzaki, K., Sasaki, M., Masaoka, Y., & Homma, I. (2011). The effect of touching a dolphin on the EEG slow waves of children. *Showa University Journal of Medical Sciences*, *23*(2), 115–119.

IAHAIO. Definitions for Animal Assisted Intervention and Guidelines for Wellness of Animals Involved in AAI. https://iahaio.org/wp/wp-content/uploads/2020/07/iahaio_wp_updated-2020-aai-adjust-1.pdf

Ikura, Y., Sakamoto, Y., Imai, T., Akai, L., Matsucka, T., Sugihara, K., Utumi, M., & Tomikawa, M. (2001). Dolphin assistedted seawater therapy for severe atopic dermatitis: An immunological and ppsychological study. *Archives of Allergy and Immunology*, *124*, 389–3909.

Johannes, B., Bernius, P., Lindemann, J., Kraus de Camargo, O., & Oerter, R. (2016). Feasibility study using in-water EEG measurement during dolphin assisted therapy. *International Journal of Clinical Psychiatry*, *4*(1), 17–25.

Kazdin, A. E. (1994). Methodology, design, and evaluation in psychotherapy research. In A. E. Bergin, & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (pp. 19–71). Wiley.

Kendall, P. C., & Norton-Ford, J. D. (1982). Therapy outcome research methods. In P. C. Kendall, & J. N. Butcher (Eds.), *Handbook of research methods in clinical psychology* (pp. 429–460). John Wiley and Sons.

Kohn, N., & Oerter, R. (2013). Dolphin assisted therapy works: Scientific findings from Eilat and Florida. *International Journal of Clinical Psychiatry*, *1*(1), 1–16.

Linde, K., Fassler, M., & Meissner, K. (2011). Placebo interventions, placebo effects and clinical practice. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *366*, 1905–1912.

Lu, A., Chan, S., Cai, Y., Huang, L., Nay, Z. T., & Goei, S. L. (2017). Learning through VR gaming with virtual pink dolphins for children with ASD. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2017.1399149

Malan, D. (1999). *Individual psychotherapy and the science of psychodynamics* (2nd ed.). Butterworth-Heinemann.

Marine Connection: Protecting Dolphins and Whales Worldwide. (2009). "Truth about dolphin assisted therapy". http://www.marineconnection.org/campaigns/captivitydat2006.html

Marino, L. (2013) Dolphins are not healers. AEON Magazine. http://www.aeonmagazine.com/nature-and-cosmos/lori-marino-dolphins-are-not-healers/

Marino, L. (2016). Construct validity of animal-assisted therapy and activities: How important is the animal in AAT? *Anthrozoös*, *25*(Suppl 1), S139–S151.

Marino, L., & Lilienfeld, S. (1998). Dolphin-assisted therapy: Flawed data, flawed conclusions. *Anthrozoos*, *11*(4), 194–199.

Marino, L., & Lilienfeld, S. (2007). Dolphin assisted therapy: More flawed data, more flawed conclusions. *Anthrozoos*, *20*, 239–249.

May, D. K., Seivert, N. P., Cano, A., Casey, R. J., & Johnson, A. (2016). Animal-assisted therapy for youth. A systematic methodological critique. *Human Animal Interaction Bulletin*, *4*(1), 1–18.

McKinney, A., Dustin, D., & Wolff, R. (2001). The promise of dolphin-assisted therapy. *Parks and Recreation*, *36*, 46–50.

MdYusof, M. S. B., & Chia, N. K. H. (2012). Dolphin encounter for special children (DESC) program: Effectiveness of dolphin-assisted therapy for children with autism. *International Journal of Special Education*, *27*(3), 54–67.

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., & McMahon, M. (1997). Effectiveness of short-term dolphin assisted therapy for children with severe disabilities. *Anthrozoös*, *10*, 90–100.

Nathanson, D. E., & de Faria, S. (1993). Cognitive improvement of children in water with and without dolphins. *Anthrozoös*, *6*(1), 17–29.

Nimer, J., & Lundahl, B. (2007). Animal-assisted therapy: A meta-analysis. *Anthrozoös*, *20*(3), 225–238.

Ortiz-Sánchez, P., Mulas, F., Abad-Mas, L., Roca, P., & Gandía-Benetó, R. (2018). Randomized controlled study of the interhemispheric coherence of the electroencephalogram after dolphin assisted therapy in children with autism spectrum disorder. *Review Neurologique*, *66*(Suppl 1), S65–S70.

Ruscio, J. (2002). *Clear thinking with psychology: Separating sense from nonsense*. Wadsworth Publishing Company.

Salguiero, E., Nunes, L., Barros, A., Salguiero, A. I., & dos Santos, M. E. (2012). Effects of a dolphin interaction program on children with autism spectrum disorders: An exploratory research. *BMC Research Notes*, *5*, 199.

Sasson, N. J., Elison, J. T., Turner-Brown, L. M., Dichter, G. S., & Bodfish, J. W. (2011). Brief report: Circumscribed attention in young children with autism. *Journal of Autism and Developmental Disorders*, *41*(2), 242–247.

Serpell, J., McCune, S., Gee, N., & Griffin, J. A. (2017). Current challenges to research on animal-assisted interventions. *Applied Developmental Science*, *21*(3), 223–233.

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., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton: Mifflin.

Shaughnessy, J. J., & Zechmeister, E. B. (1994). *Research methods in psychology*. McGraw-Hill.

Souter, M. A., & Miller, M. D. (2007). Do animal-assisted activities effectively treat depression? A meta- analysis. *Anthrozoös*, *20*(2), 167–180.

Stewart, K. L., & Marino, L. (2009). Dolphin-human interaction programs: Policies, problems, and practical alternatives. *Policy Paper for Animals and Society Institute*, 39.

Veling, W., Sjolemma, M., & Brada, B. C. (2018). Reducing impact of stress in patients with psychiatric disorders: A pilot study on the effects of swimming with wild, free dolphins in virtual reality. *International Journal of Child Health and Human Development*, *11*, 183–187.

Vermeulen, J. (2001). *Het meten van sociale vaardigheden bij mensen met een ernstige of diepe verstandelijke handicap* (Masters thesis). University of Leiden. The Netherlands.

Yerbury, R. M., & Boyd, W. E. (2018). Human–dolphin Interactions: Relationships, connections, and the reinforcement of an ongoing nature relationship. *Anthrozoös*, *31*(4), 443–458.

How to cite this article: Marino L, Lilienfeld SO. Third time's the charm or three strikes you're out? An updated review of the efficacy of dolphin-assisted therapy for autism and developmental disabilities. *J Clin Psychol*. 2021;1–15. https://doi.org/10.1002/jclp.23110