Journal of Veterinary Behavior 15 (2016) 66–77

Contents lists available at ScienceDirect

Journal of Veterinary Behavior

journal homepage: www.journalvetbehavior.com

Canine Research

No better than flipping a coin: Reconsidering canine behavior evaluations in animal shelters

Gary J. Patronek [a,*], Janis Bradley [b]

[a] Center for Animals and Public Policy, Cummings School of Veterinary Medicine, Tufts University, North Grafton, Massachusetts
[b] The National Canine Research Council, Amenia, New York

ARTICLE INFO

Article history:
Received 12 May 2016
Received in revised form
27 July 2016
Accepted 4 August 2016
Available online 13 August 2016

Keywords:
animal shelter
dog behavior evaluation
aggression
dog personality
sensitivity
predictive value

ABSTRACT

Use of behavior evaluations for shelter dogs has progressed despite their lack of scientific validation as reliable diagnostic tools. Yet results of these evaluations are often used to make life-and-death decisions. Despite acknowledging the significant limitations of evaluations, most authors suggest that the solution is to continue to attempt to remedy deficiencies. We take a contrary position and use existing data and principles of diagnostic test evaluation to demonstrate that reliably predicting problematic behaviors in future adoptive homes is vanishingly unlikely, even in theory, much less under the logistical constraints of real-world implementation of these evaluations in shelters. We explain why it would be difficult, if not impossible, to calculate robust values for sensitivity and specificity of a shelter canine behavior evaluation as required for any valid diagnostic test. We further explain the consequences of disregarding the effect of prevalence on the predictive value of a positive test (e.g., eliciting biting or warning behavior from the dog in the behavior evaluation). Finally, we mathematically demonstrate why, for any plausible combination of sensitivity, specificity, and prevalence of biting and warning behaviors, a positive test would at best be not much better than flipping a coin, and often be much worse, because many of the dogs who test positive will be false positives. Shelters already screen from adoption obviously dangerous dogs during the intake process. Subsequent provocative testing of the general population of shelter dogs is predicated on an assumption of risk that is far in excess of existing data and relies on assumptions about dog behavior that may not be supportable. We suggest that instead of striving to bring out the worst in dogs in the stressful and transitional environment of a shelter and devoting scarce resources to inherently flawed formal evaluations that do not increase public safety, it may be far better for dogs, shelters, and communities if effort spent on frequently misleading testing was instead spent in maximizing opportunities to interact with dogs in normal and enjoyable ways that mirror what they are expected to do once adopted (e.g., walking, socializing with people, playgroups with other dogs, games, training). In conjunction with a thorough and objective intake history when available, these more natural types of assessment activities will help identify any additional dogs whose behavior may be of concern. Engaging in the normal repertoire of activities familiar to pet dogs has the additional benefit of enriching dogs' lives and minimizing the adverse effect of being relinquished and confined to a shelter, will be more indicative of the typical personality and behavior of dogs, and may help make dogs better candidates for adoption.

Introduction

Use of formal canine behavior evaluations in animal shelters as a way to assess propensity for various undesirable behaviors in dogs before making them available for adoption to the public has been going on for more than 2 decades. The first published report of a behavior evaluation of shelter dogs appeared in the literature in 1991 (Van der Borg et al., 1991), and various other instruments have

* Address for reprint requests and correspondence: Gary J. Patronek, Center for Animals and Public Policy, Cummings School of Veterinary Medicine, Tufts University, North Grafton, MA 01536.
E-mail address: gary.patronek@tufts.edu (G.J. Patronek).

http://dx.doi.org/10.1016/j.jveb.2016.08.001

been developed since (Haverbeke et al., 2015). These range from very systematic batteries of tests designed by individuals credentialed in animal behavior, to *ad hoc* procedures developed by shelter staff members, to impromptu combinations of both that have been modified and adapted according to the preferences of different users. Although each evaluation is different, they generally include exposing dogs to a series of provocative stimuli (tests) in a semi-controlled environment to determine whether behaviors such as growling, snarling, snapping, lunging, or biting can be elicited, sometimes along with other behaviors that might prove either problematic or even desirable (e.g., trainability) in an adoptive home. In our experience, the resources required to conduct these evaluations are substantial, and shelters may rely on the results to make life-and-death decisions for dogs, so the consequences are significant for all involved.

The extent of use of formal canine behavior evaluations is unknown, but results from one online convenience sample of mostly small, private sheltering organizations indicated that about 25% of the organizations used one, with most of those (60%) using a test of their own design (D'Arpino et al., 2012). Large, public shelters, however, were very underrepresented in that sample. Although we have no systematic information either on why shelters came to adopt this practice or their current reasons for maintaining it, anecdotal reports among people involved in shelter work suggest that they originally emanated from a desire to protect the public from potentially dangerous dogs. In some cases, this has grown to include making the best match between dogs and adopters or trying to identify behavioral issues that may require attention while in the shelter. Another underlying motivation may be to remove or mitigate some of the emotional stress on shelter staff when confronted with making euthanasia decisions to make space for incoming dogs. In these situations, the behavior evaluation process could provide the appearance of a less-arbitrary, more justifiable rationale than number of days in the shelter or workers' opinions about which dogs would be more attractive to adopters. It is also possible that shelter staff or board members may have been influenced by reports in the medical, veterinary, and behavioral literature in which dog bites frequently are framed as an epidemic (despite declines of about 90% in reports of dog bites from the 1970s through the 2000s) (NCRC, undated). Numerous published reports about reasons for relinquishment of dogs to shelters may also have contributed to an impression that shelter dogs are "damaged goods," somehow markedly different from owned dogs. This would be unfortunate because data indicate that human-related factors such as housing, cost of care and/or veterinary treatment, and family problems are important contributors to relinquishment (Weiss et al., 2015; see Coe et al., 2014 for a comprehensive review). Furthermore, being relinquished for a manageable problem (e.g., housetraining) likely reflects more on the owner's commitment and ability than on the dog. The desire on the part of shelters to avoid liability may also play a role, but the question is one that needs study. (Interestingly, legal experts have not come to an agreement about what effect performing such an evaluation would have on a shelter's liability in the event of a bite. They have, however, identified several strategies to reduce liability, such as being sure that ownership of the dog is transferred at the time of adoption and disclosing any information the shelter has regarding prior behavior [Lutz, 2009]).

How do we begin to evaluate the merits of canine behavior evaluations in shelters as valid diagnostic instruments? The goal of a clinical diagnostic is to determine whether a subject has a particular condition or trait. This is seldom straightforward for any diagnostic test, as there is not always a clear biological "black and white" cutpoint for an individual who is positive or negative for a condition. It is even more challenging for a condition requiring a subjective assessment. A good example is the radiology literature, where studies have shown that agreement about the diagnosis of a physical condition or disease state on a radiograph at a single point in time is far from perfect, even among seasoned specialists working under ideal conditions (e.g., Arealis et al., 2014; Khan et al., 2011; Matsunaga et al., 2009). For a canine behavior evaluation, "diagnosis" would involve ascertaining not only whether a dog did or did not exhibit a behavior of interest on one or more tests in the shelter but also that the behavior, if it occurred, constituted a stable trait that would be expressed in other contexts and that it posed a danger. In the unlikely case that the first of these conditions could achieve reliability, the other two remain entirely speculative.

A large body of science has developed around the principles of developing, assessing, validating, and using diagnostic tests. The formulas and principles for evaluating key attributes of diagnostic tests (sensitivity, specificity, predictive value of a positive test, predictive value of a negative test) are well established and fairly straightforward. However, the process of doing so is complicated and costly, and it is unsurprising that no behavior evaluation for shelter dogs has yet been scientifically validated. Given the resource-constrained environments of animal shelters, and a sincere desire to adopt best practices when possible, promotion and use of behavior evaluations for shelter dogs has progressed well ahead of their scientific validation as a reliable diagnostic tool. Indeed, one of the authors has been involved in efforts to develop, implement, and validate such behavioral tests (Gary J. Patronek) and the other (Janis Bradley) has been involved in administering tests. The limitations of canine behavioral evaluations have been well described, although the tendency is that after conceding these points, most authors suggest that the solution is to attempt to remedy the deficiencies (Rayment et al., 2015; King et al., 2012; Mornement et al., 2010; van der Borg et al., 2010; Diesel et al., 2008; Christensen et al., 2007; Diederich & Giffroy, 2006; Taylor & Mills, 2006).

In this article, we take a contrary position and argue that it might be time to step back and ask a more fundamental question—namely, is it even feasible to develop a canine behavioral evaluation that is sufficiently predictive of certain unwanted behaviors in the future home to justify the cost to shelters and dogs? To address that question, we unpack each of the criteria and assumptions for constructing and validating diagnostic tests and examine some conceptual issues related to canine behavior and conducting these tests in a shelter. We will limit the discussion to the evaluation of behaviors considered as dangerous by the test designers because of the emphasis on provoking warning and biting behaviors and because this is consistently the top, sometimes the only, priority of organizations that use behavior evaluations. Finally, we will explain why eliciting warning and biting behaviors (referred to here as a positive finding or positive test) in particular is no better than flipping a coin in terms of informative value for either improving public safety or justifying euthanasia decisions for dogs and make recommendations for moving forward. The simulations described in this article demonstrate how achieving a result better than simple chance with regard to reliably predicting whether dogs will exhibit growling, snarling, snapping, or biting behavior that becomes problematic in their adoptive homes is vanishingly unlikely, even in theory, much less under the logistical constraints of real-world implementation in shelters.

Key attributes of diagnostic tests

Sensitivity and specificity

Every diagnostic test has 2 inherent characteristics, sensitivity and specificity, that play a major role in determining the