ESSAY

Rational use of diagnostic and screening tests

B. A. McKenzie

Adobe Animal Hospital, 4470 El Camino Real, Los Altos, CA 94022, USA

Corresponding author email: mckenzievmd@gmail.com

Veterinarians have a vast and ever-expanding array of diagnostic tests available to them. However, this abundance can be an embarrassment of riches that confounds diagnosis and undermines patient care if we do not make critical and informed decisions about the selection and interpretation of the tests we employ. Effective use of diagnostic tests requires a deliberate and informed approach. We must consider the strengths and weaknesses of the tests themselves and the clinical context, and we must be wary of the many biases that skew our use and interpretation of diagnostic tests. Understanding sensitivity and specificity, likelihood, prevalence and predictive value, the basic principles of Bayesian reasoning, and the cognitive biases that drive inappropriate testing are all critical to ensuring our use of imaging and laboratory testing improves patient outcomes.

Journal of Small Animal Practice (2021), 1–6
DOI: 10.1111/jsap.13393

Accepted: 24 May 2021

INTRODUCTION

Veterinarians have a vast and ever-expanding array of diagnostic tests available to them. However, this abundance can be an embarrassment of riches that confounds diagnosis and undermines patient care if we do not make critical and informed decisions about the selection and interpretation of the tests we employ.

In human medicine, there is evidence that overuse and misuse of diagnostic tests is common (Rao & Levin 2012, Greenberg & Green 2014). The consequences of this include misdiagnosis (incorrect diagnosis of a disease that is not present) and overdiagnosis (correct diagnosis of a disease that will never harm the patient), both of which cause patient harm and increase health care costs. A recognition of these adverse consequences has stimulated efforts to reduce inappropriate and unnecessary diagnostic testing in the medical profession (Welch *et al.* 2011).

Unfortunately, there is no similar effort to discourage inappropriate diagnostic testing in veterinary medicine. The extent to which diagnostic tests are misused is unknown in our field, and the subjects of overdiagnosis and rational diagnostic testing strategies are seldom discussed in the veterinary literature (Kiss & Pierce 2010, McKenzie 2016). However, the causes of inappropriate diagnostic testing in human medicine are well characterised, and these are likely to be relevant to testing decisions made by veterinarians as well. These causes include a lack of understanding of test characteristics, the lack of a consistent strategy for integrating test results into clinical decision making, and various psychological and economic pressures that influence decisions about testing.

WHY TEST?

The goal of any test we run should be obtaining information that allows us to more effectively treat or prevent health problems in our patients. This seems obvious, but it is all too easy to lose sight of this core aim. We may feel obligated to run tests to confirm a diagnosis even when the level of confidence is already high and the outcome of the test will not change what we recommend or what the client chooses to do. We may employ diagnostic tests because of a perceived pressure from clients to take action, even when such action is unlikely to change the outcome for the patient. In some situations, we may be completely confused by a case and run multiple lab tests indiscriminately, hoping for some insight to emerge.

Unfortunately, these practices reduce the reliability of the tests we employ. Effective use of diagnostic tests requires a deliberate and informed approach. We must consider the strengths and weaknesses of the tests themselves and the specific clinical context, and we must be wary of the many biases that skew our use and interpretation of diagnostic tests. The first step in constructing a rational strategy for diagnostic testing is understanding the salient characteristics of our tests and our patients.

JSAP JOURNAL OF SMALL ANIMAL PRACTICE

CHARACTERISTICS OF TESTS AND PATIENTS

Reference intervals

It is common practice for clinicians to identify test results outside of the reference interval (often misleadingly called the "normal range") as indicative of a disease state and then to pursue further diagnosis or treatment of patients based on these results. However, a reference interval is a statistical contrivance, typically defined as the interval which includes the values from 95% of healthy individuals tested during the validation of a specific assay. This range, by definition, excludes 5% of individuals from the test population used to establish it even though they are free of any disease, meaning such "abnormal" values may actually represent normal individual variation (Hyltoft Petersen & Henny 2004).

The reference interval is also a characteristic of the specific population tested and the testing methods used. Individuals with different characteristics from this population (differing in age, sex, breed, health status, etc.) will likely have different "normal" values for a given test. Furthermore, when testing multiple variables simultaneously, as is commonly done, the likelihood of some variables falling outside the reference interval by chance is often quite high even if the variables are independent and the patients tested are healthy (Hyltoft Petersen & Henny 2004).

It is inappropriate, therefore, to view a reference interval in binary terms, delineating normal from abnormal values. The clinical significance of values outside of a reference range must be interpreted in light of both the context of an individual patient and the inherent limitations of how reference ranges are established.

It may be useful to think of reference intervals for clinical laboratory tests in the same way we think about rectal temperatures in our patients. While there is a normal distribution around some mean value for healthy animals that we use to define a "normal" body temperature, there is considerable variation between individuals and from day to day in a given patient. Healthy individuals many consistently have values outside of the reference range that are typical for them. A rambunctious young dog on a hot day is likely to have an elevated rectal temperature, but when interpreted in context this is clearly not an indication of disease. Veterinarians are accustomed to contextualising such physical examination findings, and the same approach is useful for interpreting laboratory test results.

Sensitivity and specificity

Most veterinarians are familiar with sensitivity and specificity, values that compare the performance of a diagnostic test to some gold standard (Fig 1). Sensitivity is the proportion of patients with the disease that correctly test positive, so the higher the sensitivity the less likely the test is to fail to detect an individual with the disease. Specificity is the proportion of patients without the disease that correctly test negative, so the higher the sensitivity the less likely the test is to falsely detect disease in a healthy patient (Drobatz 2009).

Sensitivity and specificity are a rough guide to the accuracy of diagnostic tests. However, they are influenced by the context in which they are determined, including the specific test methods and gold standard used and the characteristics of the population tested. The sensitivity and specificity of a given test are likely to be different in actual clinical use than under the conditions in which the test was validated (Shreffler & Huecker 2021).

Likelihood ratios

More useful to the clinician than sensitivity and specificity are positive and negative likelihood ratios (LRs; Fig 1). These are

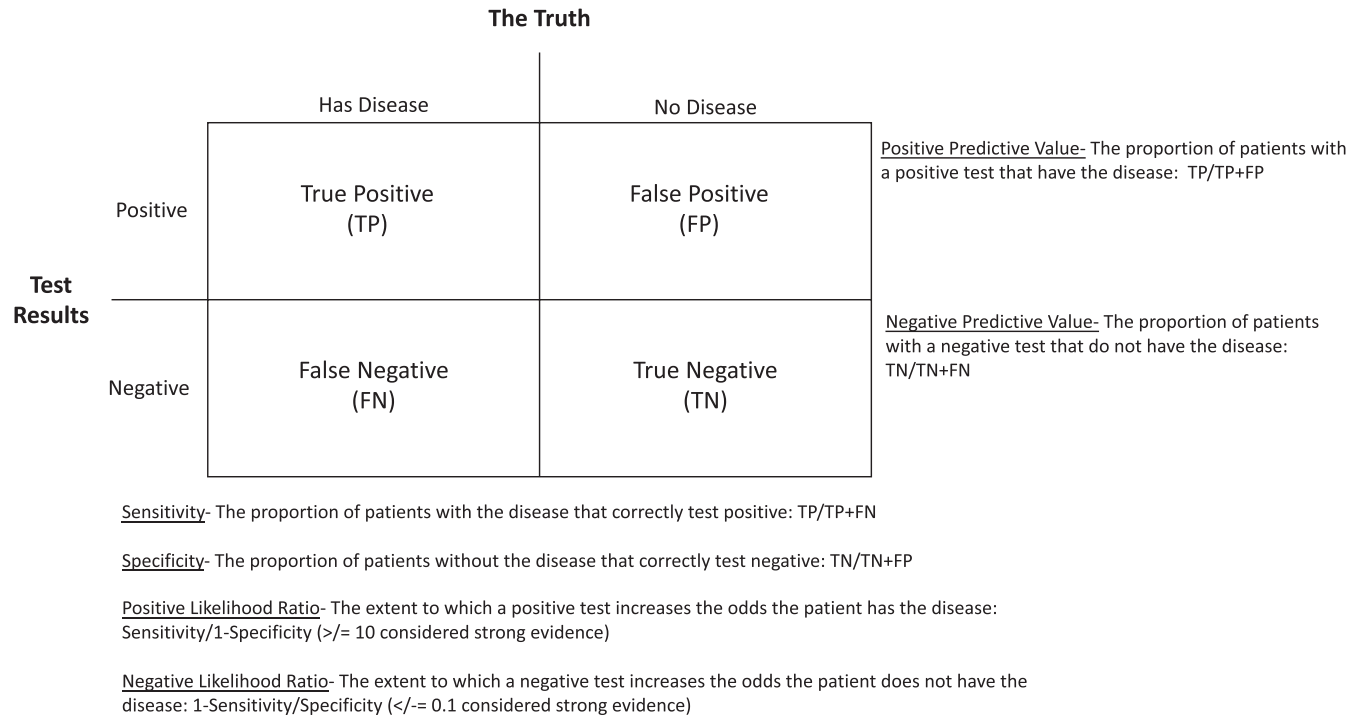

FIG 1. Definitions and calculations of sensitivity, specificity, likelihood ratios and predictive value

calculated proportions that indicate how strongly a test results alters the probability that a patient being tested actually has the disease the clinician is looking for (Lamb 2007).

For example, in a young Labrador retriever that presents with polyuria, polydipsia, and a urine specific gravity of 1.014, our differential diagnosis list may include hyperadrenocorticism (HAC). If the patient has no other signs of this condition in its history or physical examination, however, this disease will fall relatively low on our prioritised differential list. The sensitivity and specificity of a low-dose dexamethasone suppression test has been reported as 96 and 70%, respectively (Bennaim *et al.* 2018). Using these estimates, the positive and negative LRs for this test would be 3.2 and 0.05, respectively. A positive test, then would raise the odds of our theoretical patient having HAC by about threefold, a relatively small change that would not give us much confidence in the diagnosis in a patient already thought to be unlikely to have the condition (a positive LR > 10 strongly alters the probability of true disease). A negative test, however, would lower the odds of the patient having HAC dramatically, making the diagnosis very unlikely (a negative LR < 0.1 strongly alters the probability of true disease). This information is more useful in deciding whether to run such a test and how to interpret the results than sensitivity and specificity.

Positive and negative predictive value

While many clinicians understand that high sensitivity and specificity reflect in some way the accuracy of diagnostic tests, it is less widely understood that these indicators do not directly answer the question of greatest importance for the clinician: how likely is it that a patient with a positive or negative test actually has, or does not have, the disease we are testing for? The best answer to this question requires not only information about the accuracy of the test but also about the prevalence of the disease in the population of which the patient is a member. In general, the more common a disease is (i.e., the higher the prevalence of the disease), the more likely a patient with a positive test actually has the disease. Similarly, the less common the disease, the more likely a patient with a negative test is truly free of it.

We can calculate the positive predictive value (the probability a patient with a positive test has the disease: PPV) and the negative predictive value (the probability a patient with a negative test does *not* have the disease: NPV) (Fig 1). This calculation requires knowing the sensitivity and specificity of the test and the prevalence of the disease in our patient population (Shreffler & Huecker 2021).

For example, suppose we are testing feral cats in a Trap-Neuter-Release programme, with the intention of euthanizing those with feline immunodeficiency virus (FIV). If the colony we are sampling has an FIV prevalence of 2%, two of 100 cats evaluated will test positive with a perfectly sensitive test (sensitivity = 100%). If the test also has a typical specificity of 98%, then about two of 100 cats will test positive even though they do not have FIV.

All of the cats who test negative will be truly free of the disease (a NPV of 100%). However, half (two out of four) of the cats who test positive do *not* actually have FIV (PPV of 50%). This is a significant rate of misdiagnosis, especially if we are planning on euthanizing cats diagnosed with FIV!

Often we do not have reliable estimates of the prevalence of diseases we are testing for. However, understanding the general principles of PPV and NPV and the importance of prevalence is still useful. We should be sceptical of positive test results when a disease is uncommon and require additional confirmatory evidence before making a confident diagnosis. We can be more confident in positive results when testing for common conditions for which we have other corroborating evidence.

The role of prevalence and predictive value in interpreting laboratory test results is nothing more than a formal method for applying a general reasoning strategy already familiar to veterinarians. We are trained to consider common disorders before those that are rare (often with reference to the cliché of assuming hoofbeats can be attributed to horses rather than zebras in parts of the world where zebras are far less prevalent). Establishing a clinical index of suspicion, in which the relative likelihood of the possible diagnoses is assessed based on general knowledge of which conditions are more of less common in a given locale or patient population, is an informal way of estimating prevalence.

We naturally use such informal estimates to prioritise our differential diagnoses and guide our use of diagnostic tests. More formal methods, such as calculating PPV or NPV for a specific test or the process of Bayesian inference described below, are simply more explicit and rigorous ways of employing this familiar clinical reasoning strategy to help us employ and interpret diagnostic tests most effectively. This general strategy, and more formal calculations of prevalence and predictive value, has particular relevance to a special kind of testing we often conduct called screening.

Screening

Screening is the testing of apparently healthy individuals, those without clinical signs or other evidence of disease. The purpose of screening is to detect preclinical disease with the intent of intervening at an earlier stage to obtain a better outcome for the patient (Lamb 2008, Speechley *et al.* 2017). Common screening tests used in veterinary medicine include pre-anaesthetic blood testing for elective surgical procedures, routine parasite testing, and imaging or other wellness testing of healthy geriatric patients (American Animal Hospital Association (AAHA) and American Veterinary Medical Association (AVMA) 2011a,b).

The goal of screening is laudable, but evidence has emergence in recent years in human medicine showing that such testing can do more harm than good (Welch *et al.* 2011, Moynihan *et al.* 2012, Baird 2019). Apparently healthy patients are, of course, less likely to have disease than those with clinical signs, so the prevalence of disease is inherently lower in this population. Because the probability of a patient with a positive test actually having the disease is low when the prevalence of the disease is low (i.e. the PPV is low), false positive results are likely with screening, and misdiagnosis is a significant risk.

Data in humans also indicates that screening commonly leads to overdiagnosis, the correct identification of a disease that is actually present but that will never cause morbidity or mortality (Welch *et al.* 2011). Asymptomatic disease is more likely to be indolent than symptomatic disease, so early detection does not necessarily mean that early intervention will be beneficial.

Many aggressive population screening programmes in humans, such as prostate-specific antigen testing (PSA) and routine mammography, have been scaled back in the face of evidence that more patients are harmed by overdiagnosis and unnecessary follow-up testing and treatment than benefit from early detection (Berry 2013, Tabayoyong & Abouassaly 2015).

The use of PSA testing to screen for prostate cancer in men is a clear example of the harms of overdiagnosis. Most prostate cancer in humans is indolent and unlikely to cause clinical illness. Widespread screening of asymptomatic men led to unnecessary biopsy procedures and prostate cancer treatment. This raised health care costs and caused serious adverse physical and psychological effects in patients without reducing the mortality rate for this disease. Once these harms were understood, PSA testing was significantly curtailed (Welch *et al.* 2011, Tabayoyong & Abouassaly 2015).

Harm from similar unproductive testing and treatment also resulted from the aggressive use of mammography to detect mammary cancer in women. Significant increases in costs, psychological distress and harm from unneeded surgery, chemotherapy and radiation treatment resulted from the progressive expansion of mammography to younger populations. Recognition of these harms subsequently led to tighter criteria for mammography as well as a wider recognition of the dangers of overdiagnosis (Welch *et al.* 2011, Berry 2013).

In companion animal medicine, there is little evidence available to clarify the balance of benefits and harms from screening of asymptomatic individuals. The lessons about overdiagnosis from excessive PSA testing and mammography have not yet significantly impacted the veterinary profession, and many screening tests are still widely recommended (Vogt 2010, Creevy *et al.* 2019).

Pre-anaesthetic blood testing is a commonly employed screening test in veterinary medicine. It is often promoted as beneficial or even the standard of care (Irwin 2003, Lewis 2006, IDEXX Laboratories 2018). A 2015 survey on the internet discussion board Veterinary Information Network found that only 3% of the 2275 veterinarians surveyed did not require blood such testing in apparently healthy surgical patients (Veterinary Information Network Community Quick Poll 2015). In contrast, preoperative blood testing is often not recommended for routine procedures in healthy adult humans because the evidence suggests it is of little value in preventing anaesthetic complications (American Society of Anesthesiologists Task Force on Preanesthetic Evaluation 2012, National Guideline Centre (UK) 2016).

The evidence supporting pre-anaesthetic blood testing for veterinary patients is not robust (Joubert 2007, Alef *et al.* 2008, Itami *et al.* 2017). Such testing often identifies abnormalities that do not alter the anaesthetic or surgical plan. No research has shown that routine pre-anaesthetic testing in healthy patients reduces morbidity or mortality. There is also no research investigating the risk of overdiagnosis or other harms from this testing.

In general, veterinary studies of screening frequently identify lesions or values outside of reference intervals, especially in geriatric animals (Davies 2012, Paepe *et al.* 2013, Willems *et al.* 2017). This is often assumed to mean such testing is useful because occult disease is common in these populations. However, there is no evidence showing that this testing actually improves health or life expectancy in these animals, or that the benefits outweigh the risks of misdiagnosis and overdiagnosis. Choosing to use screening tests without such evidence is simply an act of faith, hoping this testing will do more good than harm. The example of human medicine suggests this is not a sound basis for screening.

CHARACTERISTICS OF CLIENTS AND CLINICIANS

In addition to the characteristics of diagnostic tests and veterinary patients, rational use of such tests requires an understanding of the drivers of ineffective testing practices. These include cognitive biases and economic factors.

Veterinarians are often encouraged to test often and extensively. Clinical practice guidelines, the scientific literature and industry magazines, and marketing materials from companies selling testing products and services promote aggressive diagnostic testing as an essential feature of good medical practice (Lewis 2006, Vogt 2010, Kipperman 2014, IDEXX Laboratories 2018). Such exhortations are rarely accompanied by discussion of the potential harms of excessive testing. Clinicians may perceive that more testing is inherently better for patients and feel obliged to test extensively even though there is a lack of robust evidence to support this.

Several cognitive biases have been shown to increase the use of tests irrespective of the objective evidence for their benefits or harms (McKenzie 2014). One of these is commission bias; the drive to take action without an adequate rationale or evidence for the value of the action. Clients or clinicians may seek testing in order to reduce anxiety or uncertainty about a patient's state of health. Veterinarians may feel a need to order tests to gain the confidence of the client, justify the time and cost of a visit, or protect themselves from accusations of negligence or malpractice. These are understandable feelings, but they a poor foundation for effective use of diagnostic tests.

Availability bias is another cognitive error that can drive inappropriate diagnostic testing. Phenomena which are easily remembered are viewed as more common or more important than phenomena which are less readily recalled. Conditions we have failed to diagnose correctly in the past are readily available to memory, and this encourages us to look for these conditions more aggressively. Any doctor who has missed a diagnosis by failing to test for it is likely to employ such testing more frequently in the future, even if the context and evidence does not support doing so.

Clients can also drive inappropriate testing decisions. Clients may seek tests to gain a feeling of control over their pets' health. Surveys indicate that patients in human medicine believe strongly in the value of screening and associate more testing with better care (Little *et al.* 2004, Schwartz *et al.* 2004, Shaked *et al.* 2019). The drive to reduce anxiety or uncertainty is so powerful that human patients often resist evidence-based recommendations to reduce screening and view screening as beneficial even after having received a false-positive result (Schwartz *et al.* 2004). Physicians respond to patient pressure when ordering tests, and it is

likely that veterinarians experience and respond to similar pressure from clients.

Money can also influence the use of diagnostic tests in human and veterinary medicine (Hughes *et al*. 2010, Kipperman *et al*. 2017). Most commonly in the veterinary field, concerns about cost interfere with recommended testing. However, diagnostic testing is a source of income for veterinarians, and wellness screening is often promoted as an important revenue stream (Noonan 2002, Hill 2016). Physicians have been shown to order more tests when they have a financial interest in testing, and it would be naïve to imagine veterinarians are immune from this type of bias (Bishop *et al*. 2010). While it is appropriate to charge for our expertise and services, we must be vigilant to ensure our use of diagnostic tests is consistent with evidence-based practices regardless of our financial interests.

RATIONAL USE OF DIAGNOSTIC TESTS

The selection and interpretation of diagnostic tests must always occur in the context of an individual case, so general rules about testing cannot be followed rigidly. However, there are some broad strategies for the rational use of diagnostic tests which can help us to increase the value and reduce the harm of the testing we do.

One useful framework is Bayesian reasoning (Gardner 2002, Homwong *et al*. 2015). Briefly, a Bayesian approach begins with estimating the likelihood of a diagnosis based on all of the usual factors we consider (signalment, personal history, prevalence rates, physical exam findings, previous test results, etc.). This clinical index of suspicion is used to set a prior probability for the diagnosis. If this probability is high enough or low enough to definitively make or rule out a diagnosis, no additional testing is needed.

If, however, the prior probability leaves significant uncertainty, we should select a test that will meaningfully raise or lower that probability. We should view the results of this test not as determining the disease to be present or absent, but as a factor that raises or lowers the likelihood of the disease. This new estimate, our posterior probability, then determines whether we have a definitive diagnosis or whether more testing is indicated. The process continues until the probability of a disease being present or absent is sufficient for us to make a reasonably confident diagnosis.

As discussed previously, Bayesian inference is simply an explicit and formal way of ensuring we incorporate our understanding that common diseases are common and rare diseases are rare into our use and interpretation of diagnostic tests. The quantitative nature and apparent precision of many tests can mislead us into too readily accepting results that conflict with a reasonable clinical index of suspicion (our prior probability). Bayesian methods help us to avoid this pitfall.

In our previous example of FIV testing in feral cats, we can consider a middle-aged intact male cat with evidence of numerous wounds suggesting a history of fighting. While the prevalence of FIV in the population as a whole is 2%, prevalence studies have consistently found infection rates are much higher in males than females, and both age and a history of fighting are relevant risk factors. Even without a precise numerical estimate of prevalence, we would consider the prior probability of FIV in this cat much higher than the average in this patient. This makes the PPV of our test correspondingly higher, and we would be more likely to accept a positive result as diagnostic.

If we estimate the prevalence in adult males as 6%, based on available studies, we can calculate our PPV as 75%. This is significantly greater than the 50% PPV based on a general prevalence of 2%, so our confidence that a positive test result indicates FIV is present is also greater. This is a much more reliable method for integrating test results into our clinical decision making than simply accepting a positive or negative result as definitive.

It is possible to simplify our approach even further into several general principles which may reduce the misuse and misinterpretation of test results, if they are applied thoughtfully rather than reflexively.

1. If a test result will not significantly alter your diagnostic or treatment plan, do not run the test.

It is worth remembering that a diagnosis in itself offers little benefit to the patient. If treatment is not possible, not desired by the client, or not appropriate based on an assessment of the overall risks and benefits, testing is unlikely to be useful.

2. If the prior probability of a diagnosis is very high or very low, do not run the test.

In some circumstances, a single test result can dramatically shift our assessment of the probability of a diagnosis. However, in most cases, if the prior probability is extremely high or low, the predictive value of a test result that conflicts with this estimate is very low. Making a definitive diagnosis, or instituting treatment, is not warranted on the basis of a single test result with a low predictive value, so the test is unlikely to facilitate accurate diagnosis or benefit the patient.

3. Avoid screening tests without clear evidence that the probability of detecting actionable disease and benefitting the patient is high.

This is a difficult principle for most veterinarians. The belief that testing for occult disease has great benefits and negligible risks is widespread. The evidence for the harms of overdiagnosis and overtreatment that has reduced the use of screening in humans has not yet been developed in our field. The cognitive bias of feedback sanction also encourages screening.(Croskerry 2000). We are more likely to be aware of cases in which we have failed to make a diagnosis through failure to test than we are to see the harms of unnecessary or misleading testing. Availability bias and feedback sanction bias us towards identifying errors of undertesting more readily than errors of overtesting.

However, the basic principles of statistics and Bayesian reasoning already outlined, and the strong evidence from human medicine, should cause veterinarians to be more careful and judicious in our use of tests in asymptomatic individuals. While being aware of our biases does not fully immunise us against them, this awareness provides an opportunity to reflect on our clinical reasoning and change our strategies to be more rational and evidence-based.

It is harder for veterinarians to choose inaction over action, to recommend not testing rather than testing. And there is no question that diagnostic and screening can provide information that benefits our patients. A philosophy of eschewing testing and relying only on clinical findings to guide diagnosis and treatment is no more rational than a strategy of indiscriminate testing. The balance between testing too much or too little should, ideally, be determined by robust evidence concerning the value of specific tests in specific populations for particular indications.

Sadly, such evidence is often lacking or incomplete. Fortunately, veterinarians are adept at working in a context with limited information and resources, and a little critical, informed thinking about diagnostic tests can go a long way to reducing inappropriate testing and maximising the benefits for our patients of the tests we use.

Conflict of interest

None of the authors of this article has a financial or personal relationship with other people or organisations that could inappropriately influence or bias the content of the paper.

References

Alef, M., von Praun, F. & Oechtering, G. (2008) Is routine pre-anaesthetic haematological and biochemical screening justified in dogs? *Veterinary Anaesthesia and Analgesia* **35**, 132-140. https://doi.org/10.1111/j.1467-2995.2007.00364.x

American Animal Hospital Association (AAHA) and American veterinary Medical Association (AVMA) (2011a) AAHA-AVMA canine preventive healthcare guidelines. https://www.aaha.org/globalassets/02-guidelines/preventive-healthcare/caninepreventiveguidelines_ppph.pdf. Accessed November 25, 2020

American Animal Hospital Association (AAHA) and American Veterinary Medical Association (AVMA) (2011b) AAHA-AVMA feline preventive healthcare guidelines. https://www.aaha.org/globalassets/02-guidelines/preventive-healthcare/felinepreventiveguidelines_ppph.pdf. Accessed November 25, 2020

American Society of Anesthesiologists Task Force on Preanesthetic Evaluation (2012) Practice advisory for preanesthesia evaluation. *Anesthesiology* **116**, 522-538. https://doi.org/10.1097/ALN.0b013e31823c1067

Baird, G. S. (2019) The choosing wisely initiative and laboratory test stewardship. *Diagnosis* **6**, 15-23. https://doi.org/10.1515/dx-2018-0045

Bennaim, M., Shiel, R. E., Forde, C., *et al*. (2018) Evaluation of individual low-dose dexamethasone suppression test patterns in naturally occurring hyperadrenocorticism in dogs. *Journal of Veterinary Internal Medicine* **32**, 967-977. https://doi.org/10.1111/jvim.15079

Berry, D. A. (2013) Breast cancer screening: controversy of impact. *Breast* **22**(Suppl 2), S73-S76. https://doi.org/10.1016/j.breast.2013.07.013

Bishop, T. F., Federman, A. D. & Ross, J. S. (2010) Laboratory test ordering at physician offices with and without on-site laboratories. *Journal of General Internal Medicine* **25**, 1057-1063. https://doi.org/10.1007/s11606-010-1409-7

Creevy, K. E., Grady, J., Little, S. E., *et al*. (2019) 2019 AAHA canine life stage guidelines. *Journal of the American Animal Hospital Association* **55**, 267-290. https://doi.org/10.5326/JAAHA-MS-6999

Croskerry, P. (2000) The feedback sanction. *Academic Emergency Medicine* **7**, 1232-1238. https://doi.org/10.1111/j.1553-2712.2000.tb00468.x

Davies, M. (2012) Geriatric screening in first opinion practice – results from 45 dogs. *The Journal of Small Animal Practice* **53**, 507-513. https://doi.org/10.1111/j.1748-5827.2012.01247.x

Drobatz, K. J. (2009) Measures of accuracy and performance of diagnostic tests. *Journal of Veterinary Cardiology* **11**(Suppl 1), S33-S40. https://doi.org/10.1016/j.jvc.2009.03.004

Gardner, I. A. (2002) 'The utility of Bayes' theorem and Bayesian inference in veterinary clinical practice and research. *Australian Veterinary Journal* **80**, 758-761. https://doi.org/10.1111/j.1751-0813.2002.tb11347.x

Greenberg, J. & Green, J. B. (2014) Overtesting: why more is not better. *American Journal of Medicine* **127**, 362-363. https://doi.org/10.1016/j.amjmed.2013.10.024

Hill, L. (2016) The Economics of the In-house Lab. Veterinary Advantage, Gainesville, FL

Homwong, N., Hunprasit, V., Marthaler, D., *et al*. (2015) A Bayesian approach for inductive reasoning to clinical veterinary medicine: the math of experience. *Journal of Veterinary Medicine and Animal Health* **7**, 308-316. https://doi.org/10.5897/jvmah2015.0409

Hughes, D. R., Bhargavan, M. & Sunshine, J. H. (2010) Imaging self-referral associated with higher costs and limited impact on duration of illness. *Health Affairs (Millwood)* **29**, 2244-2251. https://doi.org/10.1377/hlthaff.2010.0413

Hyltoft Petersen, P. & Henny, J. (2004) Special issue: reference intervals. *Clinical Chemistry and Laboratory Medicine* **42**, 685-876

IDEXX Laboratories (2018) *Preanesthetic testing: don't compromise—change happens fast*

Irwin, J. (2003) Do tests first: pre-surgical blood work may eliminate variety of surprises, *DVM360*, October.

Itami, T., Aida, H., Asakawa, M., *et al*. (2017) Association between preoperative characteristics and risk of anaesthesia-related death in dogs in small-animal referral hospitals in Japan. *Veterinary Anaesthesia and Analgesia* **44**, 461-472. https://doi.org/10.1016/j.vaa.2016.08.007

Joubert, K. E. (2007) Pre-anaesthetic screening of geriatric dogs. *Journal of the South African Veterinary Association* **78**, 31-35. https://doi.org/10.4102/jsava.v78i1.283

Kipperman, B. S. (2014) The demise of the minimum database. *Journal of the American Veterinary Medical Association* **244**, 1368-1370. https://doi.org/10.2460/javma.244.12.1368

Kipperman, B. S., Kass, P. H. & Rishniw, M. (2017) 'Factors that influence small animal veterinarians' opinions and actions regarding cost of care and effects of economic limitations on patient care and outcome and professional career satisfaction and burnout. *Journal of the American Veterinary Medical Association* **250**, 785-794. https://doi.org/10.2460/javma.250.7.785

Kiss, C. M. & Pierce, B. J. (2010) The rational use of screening tests in senior wellness programs. *Clinician's Brief*, 13-16. http://carodog.eu/wp-content/uploads/2014/10/Diagnostics_Geriatrics.pdf. Accessed December 3, 2020

Lamb, C. R. (2007) Statistical briefing: likelihood ratios. *Veterinary Radiology and Ultrasound* **48**, 394-395. https://doi.org/10.1111/j.1740-8261.2007.00263.x

Lamb, C. R. (2008) Statistical briefing: screening for disease. *Veterinary Radiology and Ultrasound* **49**, 109-110. https://doi.org/10.1111/j.1740-8261.2007.00324.x

Lewis, H. B. (2006) Healthy pets benefit from blood work. *Banfield Data Savant*, 18-20. www.banfield.com/getmedia/1216c698-7da1-4899-81a3-24ab549b7a8c/2_1-Healthy-Pets-benefit-from-blood-work. Accessed November 25, 2020

Little, P., Dorward, M., Warner, G., *et al*. (2004) Importance of patient pressure and perceived pressure and perceived medical need for investigations, referral, and prescribing in primary care: nested observational study. *BMJ* **328**, 444. https://doi.org/10.1136/bmj.38013.644086.7C

McKenzie, B. A. (2014) Veterinary clinical decision-making: cognitive biases, external constraints, and strategies for improvement. *Journal of the American Veterinary Medical Association* **244**, 271-276. https://doi.org/10.2460/javma.244.3.271

McKenzie, B. A. (2016) Overdiagnosis. *Journal of the American Veterinary Medical Association* **249**, 884-889. https://doi.org/10.2460/javma.249.8.884

Moynihan, R., Doust, J. & Henry, D. (2012) Preventing overdiagnosis: how to stop harming the healthy. *BMJ* **344**, e3502-e3502. https://doi.org/10.1136/bmj.e3502

National Guideline Centre (UK) (2016) Preoperative Tests (Update), Preoperative Tests (Update): Routine Preoperative Tests for Elective Surgery. National Institute for Health and Care Excellence (UK). http://www.ncbi.nlm.nih.gov/pubmed/27077168. Accessed March 14, 2021

Noonan, S. (2002) Diagnostic testing and imaging – creating a profit centre: blueprints for your bottom line. *Canadian Veterinary Journal* **43**, 609

Paepe, D., Verjans, G., Duchateau, L., *et al*. (2013) Routine health screening: findings in apparently healthy middle-aged and old cats. *Journal of Feline Medicine and Surgery* **15**, 8-19. https://doi.org/10.1177/1098612X12464628

Rao, V. M. & Levin, D. C. (2012) The overuse of diagnostic imaging and the choosing wisely initiative. *Annals of Internal Medicine* **157**, 574. https://doi.org/10.7326/0003-4819-157-8-201210160-00535

Schwartz, L. M., Woloshin, S., Fowler, F. J. Jr., *et al*. (2004) Enthusiasm for cancer screening in the United States. *JAMA* **291**, 71. https://doi.org/10.1001/jama.291.1.71

Shaked, M., Levkovich, I., Adar, T., *et al*. (2019) Perspective of healthy asymptomatic patients requesting general blood tests from their physicians: a qualitative study. *BMC Family Practice* **20**, 51. https://doi.org/10.1186/s12875-019-0940-9

Shreffler, J. & Huecker, M. R. (2021) Diagnostic testing accuracy: sensitivity, specificity, predictive values and likelihood ratios. StatPearls, Treasure Island, FL. https://www.ncbi.nlm.nih.gov/books/NBK557491/

Speechley, M., Kunnilathu, A., Aluckal, E., *et al*. (2017) Screening in public health and clinical care: similarities and differences in definitions, types, and aims – a systematic review. *Journal of Cinical and Diagnostic Research* **11**, LE01-LE04. https://doi.org/10.7860/JCDR/2017/24811.9419

Tabayoyong, W. & Abouassaly, R. (2015) Prostate cancer screening and the associated controversy. *Surgical Clinics of North America* **95**, 1023-1039. https://doi.org/10.1016/j.suc.2015.05.001

Veterinary Information Network Community Quick Poll (2015) An acceptable period of time between screening bloodwork and a general anesthetic procedure

Vogt, A. H. (2010) AAFP-AAHA feline life stage guidelines. *Journal of the American Animal Hospital Association* **46**, 70-85

Welch, H. G., Schwartz, L. M. & Woloshin, S. (2011) Overdiagnosed: Making People Sick in Pursuit of Health. Beacon Press, Boston, MA, USA

Willems, A., Paepe, D., Marynissen, S., *et al*. (2017) Results of screening of apparently healthy senior and geriatric dogs. *Journal of Veterinary Internal Medicine* **31**, 81-92. https://doi.org/10.1111/jvim.14587