Commissioned Review Article for Special Issue
15-01068

Veterinary diagnostic imaging: Probability, accuracy and impact

Christopher R. Lamb *

Department of Clinical Sciences and Services, The Royal Veterinary College, Hawkshead Lane, North Mymms, Hertfordshire, AL9 7TA, UK

*Corresponding author: Tel: +44-1707-666234.
E-mail address: clamb@rvc.ac.uk (C.R. Lamb).

Abstract

Diagnostic imaging is essential for diagnosis and management of many common problems, but imaging is not 100% accurate and does not always benefit the patient in the way intended. When assessing the need for imaging of a patient, the probability that the patient has a morphological lesion, the accuracy of the imaging test, and the likelihood of a beneficial impact on the patient must all be considered. Few imaging tests are sufficiently accurate that they enable a diagnosis to be ruled in or out; instead the result of imaging only modifies the probability of a diagnosis.

Potential problems with excessive use of imaging tests include false positive diagnoses, incidentaloma and overdiagnosis, all of which may contribute to a negative benefit to the patient. Clinicians must be selective in their use of imaging studies for their patients, use existing clinical information when interpreting images and sensibly apply the results of imaging in the context of the needs of individual patients. There is a need for more clinical research to assess the impact of diagnostic imaging studies for veterinary patients with common conditions to help clinicians make decisions conducive to optimal patient care.

Keywords: Accuracy; Diagnostic imaging; Overdiagnosis; Screening; Staging

Introduction

In the 21st century, veterinary radiologists are able to utilise a wider range of diagnostic modalities and their services are in greater demand than ever before. Radiography has been the mainstay of diagnostic imaging for decades, but ultrasonography (US), computed x-ray tomography (CT) and magnetic resonance imaging (MRI) are now in routine use in veterinary referral hospitals throughout the world. These cross-sectional imaging modalities eliminate the problem of superimposition that affects radiography and therefore enable clearer depiction of anatomy and clearer depiction of morphological abnormalities that alter anatomy. As a result, cross-sectional imaging modalities are inherently better detectors of disease than radiography and are useful complementary methods of imaging patients.

To many, it will seem obvious that imaging is essential for diagnosis and management of many common problems, such as a fracture; however, while this is true, it must be recognised that imaging is not 100% accurate and does not always benefit the patient in the way intended. For example, not all fractures are detected by imaging, results of imaging sometimes suggest a fracture when none is present and fractures are not always correctly distinguished from other bone lesions. These limitations in clinical use of diagnostic imaging reflect variations in the nature of disease, imperfections in imaging technology and errors made by those interpreting images. Furthermore, even when each link in the imaging chain is strong, there may be limited benefit to the individual patient because imaging was unnecessary, the abnormalities detected required no treatment, selection of optimal treatment did not depend on the results of imaging or the results of imaging lead to incorrect patient management decisions (Lamb and David, 2012).

In medicine, an ‘indication’ is a valid reason to use a diagnostic test or treatment. There are three questions that a clinician needs to answer in order to determine whether imaging is indicated for a patient (Weinstein et al., 2005):

1. What is the probability that this patient has a morphological lesion?
2. How accurate is the imaging test being considered?
3. Are the results of imaging likely to have a beneficial impact on patient management?

The indication for imaging is strongest when the answers to these questions are high, high, yes; however, multiple combinations of answers are possible. When a patient is considered unlikely to have a morphological lesion, the indication for imaging is weak and serious consideration should be given to not performing imaging.

Probability

An assessment by a clinician of the probability that a patient has an abnormality, condition or specific diagnosis occurs early in a typical clinical encounter. Based on the patient's history and clinical signs, it may be possible to estimate the likelihood of a diagnosis, which is its pre-test probability. Diagnostic tests, including imaging studies, do not generally prove or disprove a diagnosis; instead the result of a test modifies the pre-test probability of a diagnosis, converting it into the post-test probability. When speaking about results of diagnostic tests, a positive result is abnormal and a negative result is normal (Guyatt, 2006). Positive test results increase the probability of diagnosis (post-test probability > pre-test probability) whereas negative test results decrease the probability of diagnosis (post-test probability < pre-test probability) (Fig. 1).

It should be noted that if – based on the history and clinical signs – the pre-test probability of a specific diagnosis is very low, it will remain low even if the result of a diagnostic test for that diagnosis is positive, and if the pre-test probability is very high, it will remain high even if the diagnostic test result is negative. An imaging example of this principle is CT for pulmonary metastasis. CT is clearly a more sensitive test for pulmonary nodules that radiography (Nemanic et al., 2006), but if the pre-test probability of metastasis is high because the patient has a malignant neoplasm with known tendency for pulmonary metastasis (e.g. canine long-bone osteosarcoma) a negative thoracic CT does not rule out metastasis (Fig. 2). Key point: post-test probability partly depends on pre-test probability.

Estimating the pre-test probability is a challenge for clinicians (Attia et al., 2004) and many clinicians do not take prevalence of disease into account when interpreting test results (Agoritsas et al., 2011). Clinicians tend to rely on their perceptions of what conditions are more likely but, in theory, it should be possible to determine the prevalence of all the important conditions in the population of animals that are registered with a particular veterinary practice, and to use that information to estimate pre-test probability because, at the start of a consultation, the likelihood that the patient has disease X (pre-test probability) will be equal to the prevalence of disease X in the population of animals that use that practice. With computerised medical records, these data are retrievable and work to do this has started (O'Neill et al., 2014a, b).

Diagnostic imaging modalities (with the exception of scintigraphy) depict morphology and enable detection of diseases that alter normal morphology. When considering pre-test probability of disease as a prelude to selection of an imaging test, it is the likelihood of morphological lesions that is most relevant. Although signs of certain functional

disorders may sometimes be detected by imaging, many animals with functional disorders such as endocrinopathies, immune-mediated conditions, renal insufficiency or diarrhoea have none or non-specific morphological changes, hence the results of imaging are likely to be negative (Leib et al., 2012).

Role of clinical history

For all types of diagnostic testing, the pre-test probability partly determines the post-test probability. In the case of diagnostic imaging, pre-test probability is also liable to influence the result because radiologists use their estimate of the pre-test probability when interpreting diagnostic images. Although it is possible to report radiographs, CT or MRI studies without knowledge of the patient, this is not advisable in a clinical setting. Only with knowledge of the patient and their clinical signs can the radiologist judge the adequacy of the images obtained, account for anatomical variants (which is particularly important in veterinary medicine), interpret the likely meaning of a negative study and answer any specific questions raised by the primary clinician. Furthermore, knowing the history makes it more likely that a radiologist will detect a relevant abnormality and less likely that they will overinterpret a normal feature of the images (Berbaum et al., 1986; Berbaum et al., 1993; Peterson, 1999; Loy and Irwig, 2004). Radiologists use information about the patient as a guide 'diagnostic schema' that enables them to weigh possible interpretations against the pre-test probability of disease (Wood, 1999). Similarly, having access to a patient's prior images or imaging reports can significantly increase a radiologist's confidence, facilitate new observations and may result in more specific diagnosis (Aideyan et al., 1995).

Accuracy

Detection of disease

When using a diagnostic test with binary results (i.e. positive or negative), there are four possible outcomes because the patient may or may not have the disease and the test result could be positive or negative. These possibilities may be illustrated by a 2 x 2 table (Table 1).

False negative results occur when a disease or condition is present, but is not detected. In diagnostic imaging, this is liable to occur if images are obtained of the wrong body part, images are poor quality, or if the lesion is too small to be resolved. False positive results occur if a patient that does not have the disease under investigation has a test result that is interpreted as positive for that disease. In diagnostic imaging, this is liable to occur if technically poor images are obtained that mimic an abnormality, an anatomical variant is misinterpreted as abnormal or signs of an unrelated subclinical condition are misinterpreted as the cause of clinical signs (Fig. 3). Trainees in radiology are particularly prone to false positive errors, possibly because they lack sufficient knowledge of radiographic anatomy and/or have an unrealistically high expectation that the images will be abnormal (Lamb et al., 2007). Within increasing experience, radiologists become more accurate mainly because they make fewer false positive errors (Lamb et al., 2011).

The sensitivity of a test is defined as the proportion of affected patients that have a positive test result. Sensitivity = true positive (TP)/(TP + false negative [FN]). A highly sensitive test gives a positive result in nearly all diseased subjects. Specificity is defined as the proportion of unaffected patients that have a negative test result. Specificity = true negative (TN)/(TN + false positive [FP]). A test of high specificity gives a negative result in most patients without the disease.

158 Sensitivity and specificity are often calculated in papers describing the results of
159 imaging in clinical patients, but it should be emphasised that these indices do not represent
160 intrinsic properties of the test in question. Estimates of sensitivity and specificity will vary
161 because of differences in the definition of the disease, the way the imaging is performed, and
162 the characteristics of patients with and without the target disease (Whiting et al., 2004). For
163 example, patients attending primary care practices will generally have disease at an earlier
164 stage than patients at referral practices, which may mean a test is less sensitive when it is
165 used in primary care practices. Similarly, investigators sometimes collect subjects for study in
166 a way that maximises the differences between affected and unaffected groups, for example,
167 by using healthy individuals (such as dogs volunteered by their owners) as the unaffected
168 group. This could be valid for 'Phase 1' research, which aims to identify tests with potential
169 clinical utility, but the results will not be applicable to a clinical setting in which all test
170 subjects are patients (Sackett and Haynes, 2002). For 'Phase 2' studies intended to estimate
171 diagnostic test accuracy in clinical patients, the unaffected group should be subjects who are
172 similar to the affected group in all aspects except their diagnosis (Guyatt, 2006). Key point:
173 interpreting reported values for sensitivity and specificity of a diagnostic test requires
174 knowledge of the patients and methods used to derive these estimates.
175
176 Few imaging tests have both high sensitivity and high specificity. One example is
177 US for pregnancy diagnosis in farm animals (Hansen and Christiansen, 1976; Davey, 1986).
178 Knowing that a test has high sensitivity or specificity helps us to use it more effectively in
179 practice. Although it seems obvious that a highly sensitive test could be used to detect
180 disease, the most powerful way to take advantage of a test with high sensitivity is to use a
181 negative result to rule out disease. For example, bone scintigraphy is considered to be a
182 highly sensitive test for stress fracture in human athletes; this means it is positive in virtually

all affected individuals, and obtaining a normal (negative) bone scan in a lame athlete rules out the possibility of stress fracture (Kanstrup, 1997). Conversely, tests of very high specificity can be used to rule in a diagnosis. The terms SpPIn (for a snsitive test, a negative result can rule a diagnosis out) and SnNOut (for a test of high specificity, a positive result can rule a diagnosis in) were designed to help practitioners memorise these principles.

There are no published examples of veterinary imaging tests that are convincing SnNOuts. On the contrary, there are numerous well-documented examples of insensitive imaging studies, including – all in dogs – radiography for pulmonary nodules (Nemanic et al., 2006), extended ventrodorsal radiographs for hip dysplasia (Lust et al., 2001), radiography for fragmented medial coronoid process (Snaps et al., 1997), US for inflammatory bowel disease (Rudorf et al., 2005), US for gastrointestinal ulceration (Pastore et al., 2007) and MRI for meningoencephalitis (Lamb et al., 2005). Examples of veterinary imaging tests that may be considered SpPIns are tibial compression radiography for cranial cruciate ligament injuries in dogs (de Rooster et al., 1998) and US for congenital portosystemic shunts in dogs (Lamb, 1996).

The problem of the 'Rule out'

A diagnosis that has been ruled out has a probability that is not significantly different from zero. Clinicians frequently speak of the need to rule out a diagnosis in their patients and differential diagnoses are sometimes labelled 'rule-outs'. This terminology implies that the process of diagnosis depends on testing to prove that certain conditions are not present and that when a condition cannot be ruled out, it may be the diagnosis. Although this seems like a logical process, it is not suitable for medical diagnosis, for several reasons: first, most diagnostic tests are not sufficiently sensitive that a negative result produces a post-

test probability approaching zero; second, if the pre-test probability is very high, it will remain high even if after a sensitive diagnostic test has produced a negative result; third, sequential testing to rule out a series of conditions will inevitably be inefficient compared to testing to rule in the condition considered most likely based on consideration of the patient's history and signs. Following a process of sequential rule outs has been criticised as a defensive-medicine-minded approach adopted by clinicians relatively unconcerned about burdening their patients with the wrong diagnosis (Jha, 2014). In contrast, patients (and their owners and health insurance companies) expect and deserve a more selective approach by a clinician exercising their clinical judgment and seeking to rule in the diagnosis they consider most likely.

Predictive value of a test

Knowing the sensitivity and specificity of a test is of limited value in clinical practice because these indices have no direct diagnostic meaning (Moons and Harrell, 2003). Sensitivity is the probability of a patient having a positive test result if they have a disease; however, clinicians usually want to know the probability of their patient having disease if the test result is positive or negative. The likelihood that the result of a diagnostic test is a true reflection of the disease status of the patient is known as the predictive value: positive predictive value = TP/TP+FP; negative predictive value = TN/TN+FN.

Predictive value is markedly affected by the prevalence (pre-test probability) of disease. Intuitively, one might expect that a negative test result always makes the diagnosis unlikely, but this is not possible if the pre-test probability is high. Similarly, positive predictive value is low when the prevalence is low, even for tests of high specificity (Fig. 4). Unless the prevalence of disease is relatively high, a positive test result is likely to be a false

positive. The positive predictive value of a test can be maximised by using the test selectively in those patients considered most likely to have the target condition. A well-known, non-imaging example of this principle is use of blood tests for hyperadrenocorticism in dogs (Kaplan et al., 1995). If a blood test for hyperadrenocorticism (such as ACTH-stimulation test) is used in all dogs presented with polydipsia, this will include many dogs with conditions other than hyperadrenocorticism, such as renal insufficiency and diabetes, hence the pre-test probability of hyperadrenocorticism will be low and a large proportion of positive test results will be false positives; however, if testing for hyperadrenocorticism is reserved for dogs that have polydipsia and other signs of hyperadrenocorticism (e.g. hepatomegaly, pendulous abdomen, alopecia), the pre-test probability of hyperadrenocorticism will be higher and a larger proportion of positive test results will be true positives. Key point: selective use of diagnostic testing in patients produces results with higher predictive value than non-selective testing

Likelihood ratios

In clinical practice, it is useful to be able to estimate how much a test result affects the probability of disease. Sensitivity and specificity do not provide this information and although predictive values do enable estimates of the probability of a disease, they depend greatly on pre-test probability, which cannot be known precisely. Likelihood ratios represent a useful alternative index for summarising the accuracy of diagnostic tests. Likelihood ratio is the ratio between pre- and post-test odds of disease: pre-test odds of disease x likelihood ratio = post-test odds of disease.

When the likelihood ratio associated with a positive test result (PLR) is high (>10), a positive result greatly increases the probability of the target condition. Conversely, when the

likelihood ratios associated with negative test results (NLR) are low (<0.1) a negative result markedly decreases the probability of the target condition.

The strength of the association between an imaging sign and pathology can be usefully expressed using likelihood ratios. For example, in dogs with chronic nasal signs, one of the main aims of radiography is to distinguish the two principal differential diagnoses: rhinitis and neoplasia. Based on data in a case-control study of dogs with nasal disease (Russo et al., 2000), the radiographic signs most strongly associated with rhinitis are nasal structures that look normal (LR 3.3, 95% confidence interval 1.4–7.7) and intranasal lucent foci (LR 3.3, 95% confidence interval 1.7–6.4) whereas the radiographic signs most strongly associated with nasal neoplasia are lysis of bone around margins of nasal cavity (LR 10.3, 95% confidence interval 3.4–31.2) and soft tissue/fluid opacity in the ipsilateral frontal sinus (LR 4.9, 95% confidence interval 2.3–10.7). Of these signs, lysis of bone around margins of the nasal cavity has the highest likelihood ratio and, therefore, may be considered the most accurate sign for distinguishing rhinitis and nasal neoplasia.

What is the accuracy of veterinary imaging studies?

In a systematic review of 5936 articles published in the period 1976-2006, only 88 contained sufficient data to assess the diagnostic performance of imaging studies (Lamb, 2008a). These 88 articles described 103 studies involving a range of imaging modalities and target conditions, with widely varying sensitivities and specificities. Excluding studies of pregnancy diagnosis, the median sensitivity was 78% (range 0-100%) and specificity 92% (range 33-100%). PLR was >10 in 21 (27%) studies and NLR was <0.1 in 13 (17%), and only 8 (10%) diagnostic imaging tests had both high PLR and low NLR. For most imaging

tests for which performance data are available, sensitivity and specificity are only moderate, hence it appears that few imaging tests could be used to rule in or rule out a diagnosis.

What is the accuracy of veterinary imaging studies that employ measurements?

In a recent systematic review of veterinary imaging tests that employ measurements, the median sensitivity was 77% (range 38-99%), specificity was 82% (range 50-99%), PLR was 4.1 (1-103) and NLR was 0.29 (0.01-1) (Lamb and Nelson, 2015). These moderate values for sensitivity and specificity primarily reflect the fact that the normal size ranges for many anatomical structures are very wide, hence there is marked overlap between normal and pathologic ranges. This overlap is particularly marked in dogs, which exhibit exceptionally wide phenotypic variation compared to other animals. Even for anatomical structures that would not be expected to vary greatly with conformation, wide normal size ranges may be observed. For example, abdominal lymph nodes in dogs are variable in size and number in CT images (Beukers et al., 2013), which complicates interpretation of size in clinical patients. Furthermore, the association between lymph node size and presence of nodal metastasis is relatively weak, hence assessment of lymph node size alone is insufficient for accurate clinical staging of neoplasia. When a significant risk of lymphatic metastasis exists in a patient, cytologic or histologic examination of regional lymph nodes is indicated regardless of the size of those nodes (Williams and Packer, 2003).

There is a tendency among clinicians to assume that making measurements of structures in diagnostic images will increase diagnostic accuracy, particularly for inexperienced observers; however, there is no evidence that this is true. For example, two studies found that observers making radiologic measurements of the heart in dogs with suspected cardiac disease and the small intestinal diameter in dogs with suspected intestinal

obstruction were no more accurate than when they relied on subjective assessment alone (Lamb et al., 2000; Ciasca et al., 2013). These findings applied equally to experienced and inexperienced observers (Lamb et al., 2000; Ciasca et al., 2013). In general, emphasis on measurements is unwarranted because the pathologic effects of disease are invariably multiple and optimal radiographic interpretation depends on assessment of all the possible ways in which the image may be abnormal.

Strength of imaging-pathological correlations

A judgement that diagnostic images are abnormal constitutes a positive test result, but that represents only a superficial summary of the meaning of the images, which invariably show morphological features representing the abnormality. Reports of imaging studies always include a description of abnormalities according to six possible morphological 'Roentgen' signs: number, size, shape, position and margination. The remaining sign is signal amplitude, which is depicted as the grey level in the image. This sign is modality-specific: we speak about opacity for radiography, echogenicity for US, density or attenuation for CT and intensity for MRI.

One of the goals of diagnostic imaging is to enable specific diagnosis based on correctly deducing the pathological nature of a lesion from its imaging signs. This works quite well at the macroscopic level, where imaging signs frequently correspond closely to the changes found at surgery or necropsy. If a radiologist reports a fracture, a mass, pulmonary consolidation, pleural or peritoneal fluid or presence of calculi, the surgeon or pathologist will frequently find that abnormality on gross inspection. Particularly with cross-sectional imaging there is the potential for relatively detailed imaging-pathological correlations. For example, a recent study found that features of CT images of canine adrenal neoplasms

correlated well with pathological features including vascular invasion, pseudoencapsulation, haemorrhage and necrosis (Gregori et al., 2015).

Less good correlations may be expected when attempting to deduce microscopic features of lesions, such as the type of cells in a mass, from the imaging signs. This problem is illustrated by recent studies attempting to correlate patterns of contrast accumulation in CT images of hepatic masses with their histological diagnosis (Fukushima et al., 2012; Kutara et al., 2014; Jones et al., 2016). The rationale for this approach is that benign hepatic masses containing relatively well-differentiated hepatocytes will tend to enhance most strongly in early post-contrast images because of their relatively abundant arterial blood flow and lack of necrotic or haemorrhagic components, whereas malignant hepatic masses will tend to out-grow their blood supply and have a significant necrotic component, so enhance less. However, marked enhancement in early post-contrast images was found to occur both with malignant neoplasms, such as hepatocellular carcinoma, and with non-malignant lesions, such as hepatic adenoma and nodular hyperplasia (Fukushima et al., 2012; Kutara et al., 2014; Jones et al., 2016). Fundamentally, the histologic diagnosis of these hepatic lesions is based on cellular architectural features that occur on a scale far below that depicted in CT images. Furthermore, the histologic features used by pathologists for diagnosis of hepatic masses exist in a spectrum of severity in which the boundaries between well-differentiated hepatocellular carcinoma and adenoma, and between adenoma and nodular hyperplasia, are not always clearly defined. Consequently, links between imaging signs, which primarily represent non-specific macroscopic features, and histologic diagnoses will be tenuous (Fig. 5). To date, no consistent differences in quantitative or categorical CT data between malignant and non-malignant hepatic masses have been identified, hence diagnosis still relies on histology.

Impact

Clinical studies often focus on the accuracy of diagnostic imaging; however, the ultimate standard of the usefulness of a diagnostic test is not its accuracy, but whether it improves patient outcomes (Guyatt et al., 2006; Sistrom, 2009). Tests with the greatest diagnostic impact are available for all patients that need testing, inexpensive, sufficiently accurate that other tests become unnecessary and lead to improved patient outcomes.

Although it may be assumed that newer, more advanced imaging techniques are better than radiography because they are more sensitive, this does not mean that patients automatically benefit from the introduction of new technology. For example, in veterinary practices with CT, few dogs or cats have survey radiography to investigate nasal signs because they have CT instead. A CT scan of the head may be done more quickly, provides a more detailed depiction of most lesions and may be interpreted with more confidence than a series of radiographs; however, differentiating rhinitis from nasal neoplasia is based on the same criteria as for survey radiography, hence the diagnostic accuracy of CT is similar (Saunders and van Bree, 2003; Saunders et al., 2003; Tromblee et al., 2006; Karnik et al., 2009). Furthermore, imaging of the nasal cavity in a referral setting is invariably followed by endoscopy, nasal flushing or biopsy for definitive diagnosis, and this is true for patients having radiography or CT. The additional benefit of CT for dogs or cats with chronic nasal signs may be negligible if the remainder of the diagnostic work-up is unchanged.

Few veterinary studies provide good evidence of benefits to patients occurring as a result of diagnostic imaging. We looked for evidence of improved outcomes for canine spinal patients having MRI, which has largely replaced myelography in small animal practice

(Naude et al., 2008; Robertson and Thrall, 2011). A retrospective cross-sectional study was done of 107 dogs with non-ambulatory thoracolumbar spinal disease that had myelography or MRI during a period when MRI was available only 2 days per week, hence choice of imaging was primarily determined by day of admission rather than patient factors or clinician preference. Outcome variables included length of hospitalisation, change in neurological grade, total cost of hospitalisation and mortality. No significant association was found between type of imaging and any outcome variables except cost of hospitalisation, which was £670 higher on average for dogs that had MRI (Parry et al., 2010). Hence, although MRI may be considered advantageous compared to myelography because it is non-invasive and provides superior anatomical detail, no beneficial effect on outcome of dogs with non-ambulatory thoracolumbar spinal disease was found.

In these examples, introduction of CT or MR has no apparent impact. It is also possible to identify clinical scenarios in which imaging applied with good intentions has a negative impact on patients.

Screening

Diagnostic testing is done because of clinical suspicion of disease in an individual patient (or group of patients) whereas screening implies using a test in individuals considered at risk for disease, but not showing any clinical signs (Brawley and Kramer, 2005). The aim of screening is generally to identify affected individuals before they develop clinical signs, and the potential benefit is easier and/or more effective treatment of the disease, which has been detected at an earlier stage than it would otherwise have been. This, in turn, may lead to reduced morbidity and mortality. The best documented example of screening based on imaging is mammography to detect breast cancer in women (Welch and Frankel, 2011;

Gotzsche and Jorgensen, 2013). Although there are relatively few screening programmes for companion animals (e.g. radiography for hip dysplasia), screening for subclinical disease occurs in health programmes for healthy geriatric animals and in comprehensive work-ups for sick animals.

Screening is usually done when the prevalence of disease is low in the population and the pre-test probability of diagnosis is low in each individual being tested. For this reason, a positive result is likely to be a false positive unless the specificity of the screening test is unusually high (Lamb, 2008b). Screening tests have great potential for harm because of the morbidity that follows unnecessary further testing or treatment of individuals with false positive results (Gotzsche and Jorgensen, 2013). Key point: the benefit of screening can be determined only by a randomised clinical trial.

For example, the finding of neoplasia at necropsy in 23% old dogs with primary brain tumours prompted a recommendation that screening tests (to look for additional tumours) should be performed before imaging the brain (of dogs with suspected intracranial neoplasia) (Snyder et al., 2006). Clinicians should be cautious about routinely following this recommendation. In a dog presenting only with neurological signs referable to the brain, logic dictates that the most likely outcomes of screening the rest of the body will be a negative result or a positive result that represents an unrelated, clinically silent lesion. Despite the obvious possibility that the clinically silent lesion may never cause clinical signs, the tendency in such cases is to investigate the new lesion and withhold or delay further work-up and/or treatment for the original condition, which risks increased mortality. The least likely outcomes of screening the rest of the body of this patient will be a distant lesion that explains the neurological signs or an unrelated lesion that is considered so serious that it contra-

indicates further work-up. Hence, it should be evident that screening a dog with suspected intracranial neoplasia is more likely to have a negative impact (because of waste of resources and increased mortality) than to benefit the patient (by improving outcome for the presenting complaint). Concentrating on the problem for which the patient presented is preferable to screening for unrelated disease.

Staging neoplasia

The results of staging in a patient with cancer should carry a prognostic meaning that helps predict the likely outcome; however, the World Health Organisation stage does not necessarily correlate with outcome measures in veterinary patients (Flory et al., 2007). Also, as more sensitive imaging modalities are used for staging neoplasia, signs of nodal or distant metastasis are identified in a larger proportion of patients than those staged previously using less sensitive imaging, such as radiography alone. This effect, known as stage inflation (Flory et al., 2007), is a problem because it confounds comparisons between results of clinical trials, which may undermine decisions by clinicians managing patients with neoplasia.

CT has higher sensitivity for pulmonary nodules than radiography (Nemanic et al., 2006), hence it is recommended for staging animals with malignant neoplasms liable to metastasise to the lung; however, caution is necessary when interpreting pulmonary CT images of such patients because lack of visible nodules does not rule out the possibility of metastasis and because a pulmonary nodule could represent a benign lesion unrelated to the primary neoplasm. There are limited veterinary data on this subject, but in children with cancer pulmonary nodules that represent benign or incidental findings cannot be reliably distinguished from malignant nodules without biopsy (Absalon et al., 2008). Finding large numbers of pulmonary nodules at CT is associated with malignancy (Absalon et al., 2008),

457 but finding a solitary nodule is problematical. For nodules that are not amenable to biopsy, it
458 is usual to repeat the CT after a period of time to look for changes (Libby et al., 2004). Lack
459 of enlargement of a nodule supports a diagnosis of 'non-malignant'.
460
461 A similar problem occurs when examining the liver for signs of metastasis in dogs
462 with abdominal neoplasia. The high prevalence of benign hepatic lesions in older dogs means
463 that a hepatic nodule could easily represent a benign, incidental finding rather than a
464 metastasis (Clendaniel et al., 2014). Similarly, although multiple hepatic lesions might be
465 assumed to be more likely to represent metastasis than a solitary lesion (Cuccovillo and
466 Lamb, 2002; Clendaniel et al., 2014), this is not a safe assumption (Levinson et al., 2009).
467
468 Staging of patients with malignant neoplasms should be done based on knowledge of
469 the usual biological behaviour of the neoplasm. For example, the large majority of canine
470 long-bone osteosarcomas metastasise to the lung and a small proportion metastasise to the
471 regional lymph nodes, so the lungs and lymph nodes should be examined in affected dogs. In
472 contrast, metastasis to abdominal organs, such as the liver or kidneys, is rare (pre-test
473 probability is very low), so there is no more than a weak indication to examine the abdominal
474 organs (Wallace et al., 2013). Pursuing a weak indication can be counter-productive. Use of
475 abdominal imaging in dogs with long-bone osteosarcomas is far more likely to produce
476 incidental findings than signs of metastasis, which leads to unnecessary additional work-up,
477 over-diagnosis and reduced survival time (Sacornrattana et al., 2013).
478
479 *The problem of the 'incidentaloma'*
480 One of the drawbacks of using imaging for screening or staging patients is the
481 occurrence of incidental findings, i.e. abnormalities without associated clinical signs (Fig. 6).

It can be difficult to decide if a finding is likely to be incidental or relevant, particularly in patients with non-specific or vague clinical signs, and whether or not to pursue it with further diagnostic tests, such as biopsy (Aspinall et al., 2013). Liaison between the primary clinician and the radiologist is essential when considering what to do next. Incidental findings complicate a diagnostic work-up, can confuse the clinician and/or animal owner and can contribute to increased morbidity and costs without any corresponding benefit to the patient.

In a recent study, potentially incidental findings were reported in 77% cats without respiratory signs that had thoracic CT, for example, to look for metastasis or as part of a comprehensive medical work-up (Lamb and Jones, 2016). The most prevalent finding was pulmonary collapse, which was likely exacerbated by sedation or anaesthesia for CT, but clinically silent bronchial lesions and space-occupying lesions were also observed frequently. Another well-recognised example of incidentaloma is the occurrence of hyperplastic nodules of the liver, spleen or adrenal glands in dogs (Stowater et al., 1990; Myers, 1997; Warren-Smith et al., 2012; Cook et al., 2014). The prevalence of both hyperplastic nodules and neoplasia increases with age (Myers, 1997), hence distinguishing these conditions is most often a problem encountered when managing older dogs.

Overdiagnosis

Overdiagnosis refers to diagnosis of disease that may never cause clinical signs during a patient's lifetime. The diagnosis may be correct but, if a lesion never causes any clinical signs, it is irrelevant. For example, diagnosis of malignancy is sometimes based on subtle histological abnormalities, such as capsular invasion. In tumours that metastasise infrequently, the rationale for labelling such tumours as ‘cancer’ on this basis is questionable (Williams, 2000). Using the term cancer for a tumour unlikely to cause significant harm to

the patient widens the definition of cancer and is one type of overdiagnosis (Moynihan et al., 2012).

Overdiagnosis leads to reclassification of normal individuals as diseased and reclassification of patients presenting with one condition as patients with multiple conditions. Use of advanced imaging contributes to overdiagnosis by detecting ever smaller abnormalities. Prevention of overdiagnosis requires mature judgement by clinicians and specific measures, such as raising thresholds for disease (Moynihan et al., 2012). Overdiagnosis is recognised as a growing problem in medicine, but there are currently no veterinary studies of this subject.

Conclusions

It is important that clinicians are selective in their use of imaging studies for their patients, that existing clinical information is used when interpreting images and that the results of imaging are applied sensibly in the context of the needs of individual patients. There is a need for more clinical research to assess the impact of diagnostic imaging studies for veterinary patients with common conditions to help clinicians make decisions conducive to optimal patient care.

Conflict of interest statement

The author has no financial or personal relationship with other people or organisations that could inappropriately influence or bias the content of the paper.

References

Absalon, M.J., McCarville, M.B., Liu, T.B., Santana, V.A., Daw, N.C., Navid, F., 2008. Pulmonary nodules discovered during the initial evaluation of pediatric patients with bone and soft-tissue sarcoma. Pediatric Blood and Cancer 50, 1147-1153.

Agoritsas, T., Courvoisier, D.S., Combescure, C., Deom, M., Perneger, T.V., 2011. Does prevalence matter to physicians in estimating post-test probability of disease? A randomized trial. Journal of General Internal Medicine 26, 373-378.

Aideyan, U.O., Berbaum, K., Smith, W.L., 1995. Influence of prior radiologic information on the interpretation of radiographic examinations. Academic Radiology 2, 205-208.

Aspinall, S.R., Ong, S.G.S., Wilson, M.S.J., Lennard, T.W.J., 2013. How shall we manage the incidentally found thyroid nodule? Surgeon-Journal of the Royal Colleges of Surgeons of Edinburgh and Ireland 11, 96-104.

Attia, J.R., Nair, B.R., Sibbritt, D.W., Ewald, B.D., Paget, N.S., Wellard, R.F., Patterson, L., Heller, R.F., 2004. Generating pre-test probabilities: a neglected area in clinical decision making. Medical Journal of Australia 180, 449-454.

Berbaum, K.S., Franken, E.A., Jr., Anderson, K.L., Dorfman, D.D., Erkonen, W.E., Farrar, G.P., Geraghty, J.J., Gleason, T.J., MacNaughton, M.E., Phillips, M.E., et al., 1993. The influence of clinical history on visual search with single and multiple abnormalities. Investigative Radiology 28, 191-201.

Berbaum, K.S., Franken, E.A., Jr., Dorfman, D.D., Barloon, T., Ell, S.R., Lu, C.H., Smith, W., Abu-Yousef, M.M., 1986. Tentative diagnoses facilitate the detection of diverse lesions in chest radiographs. Investigative Radiology 21, 532-539.

Beukers, M., Grosso, F.V., Voorhout, G., 2013. Computed tomographic characteristics of presumed normal canine abdominal lymph nodes. Veterinary Radiology and Ultrasound 54, 610-617.

Brawley, O.W., Kramer, B.S., 2005. Cancer screening in theory and in practice. Journal of Clinical Oncology 23, 293-300.

Ciasca, T.C., David, F.H., Lamb, C.R., 2013. Does measurement of small intestinal diameter increase diagnostic accuracy of radiography in dogs with suspected intestinal obstruction? Veterinary Radiology and Ultrasound 54, 207-211.

Clendaniel, D.C., Sivacolundhu, R.K., Sorenmo, K.U., Donovan, T.A., Turner, A., Arteaga, T., Bergman, P.J., 2014. Association between macroscopic appearance of liver lesions and liver histology in dogs with splenic hemangiosarcoma: 79 Cases (2004-2009). Journal of the American Animal Hospital Association 50, E6-E10.

Cook, A.K., Spaulding, K.A., Edwards, J.F., 2014. Clinical findings in dogs with incidental adrenal gland lesions determined by ultrasonography: 151 cases (2007-2010). Journal of the American Veterinary Medical Association 244, 1181-1185.

Cuccovillo, A., Lamb, C.R., 2002. Cellular features of sonographic target lesions of the liver and spleen in 21 dogs and a cat. Veterinary Radiology and Ultrasound 43, 275-278.

Davey, C.G., 1986. An evaluation of pregnancy testing in sheep using a real-time ultrasound scanner. Australian Veterinary Journal 63, 347-348.

de Rooster, H., Van Ryssen, B., van Bree, H., 1998. Diagnosis of cranial cruciate ligament injury in dogs by tibial compression radiography. Veterinary Record 142, 366-368.

Flory, A.B., Rassnick, K.M., Stokol, T., Scrivani, P.V., Erb, H.N., 2007. Stage migration in dogs with lymphoma. Journal of Veterinary Internal Medicine 21, 1041-1047.

Fukushima, K., Kanemoto, H., Ohno, K., Takahashi, M., Nakashima, K., Fujino, Y., Uchida, K., Fujiwara, R., Nishimura, R., Tsujimoto, H., 2012. CT characteristics of primary hepatic mass lesions in dogs. Veterinary Radiology and Ultrasound 53, 252-257.

Gotzsche, P.C., Jorgensen, K.J., 2013. Screening for breast cancer with mammography. Cochrane Database Syst Rev 6, CD001877.

Gregori, T., Mantis, P., Benigni, L., Priestnall, S.L., Lamb, C.R., 2015. Comparison of computed tomographic and pathologic findings in 17 dogs with primary adrenal neoplasia. Veterinary Radiology and Ultrasound 56, 153-159.

Guyatt, G., Sackett, D., Haynes, B., 2006. Evaluating diagnostic tests. In: Sackett, D.L., Haynes, R.B., Tugwell, P., Guyatt, G. (Eds.), Clinical Epidemiology: How to Do Clinical Practice Research. Lippincott Williams and Wilkins, Philadelphia, pp. 273-322.

Hansen, L.H., Christiansen, I.J., 1976. The accuracy of porcine pregnancy diagnosis by a newly developed ultrasonic A-scan tester. British Veterinary Journal 132, 66-67.

Jha, S., 2014. Stop hunting for zebras in Texas: end the diagnostic culture of "rule-out". BMJ 348, g2625.

Jones, I.D., Lamb, C.R., Drees, R., Priestnall, S.L., Mantis, P., 2016. Associations between dual-phase CT features and histopathologic diagnoses in 52 dogs with hepatic or splenic masses. Veterinary Radiology and Ultrasound 57, 144-153.

Kanstrup, I.L., 1997. Bone scintigraphy in sports medicine: a review. Scandinavian Journal of Medicine and Science in Sports 7, 322-330.

Kaplan, A.J., Peterson, M.E., Kemppainen, R.J., 1995. Effects of disease on the results of diagnostic-tests for use in detecting hyperadrenocorticism in dogs. Journal of the American Veterinary Medical Association 207, 445-451.

Kutara, K., Seki, M., Ishikawa, C., Sakai, M., Kagawa, Y., Iida, G., Ishigaki, K., Teshima, K., Edamura, K., Nakayama, T., Asano, K., 2014. Triple-phase helical computed tomography in dogs with hepatic masses. Veterinary Radiology and Ultrasound 55, 7-15.

Lamb, C.R., 1996. Ultrasonographic diagnosis of congenital portosystemic shunts in dogs: results of a prospective study. Veterinary Radiology and Ultrasound 37, 281-288.

Lamb, C.R., 2008a. Accuracy of veterinary diagnostic imaging: a systematic review (abstract) Veterinary Radiology and Ultrasound 49, 96.

Lamb, C.R., 2008b. Statistical briefing: screening for disease. Veterinary Radiology and Ultrasound 49, 109-110.

Lamb, C.R., David, F.H., 2012. Advanced Imaging Use and Misuse. Journal of Feline Medicine and Surgery 14, 483-497.

Lamb, C.R., Jones, I.D., 2016. Associations between clinical signs and abnormalities reported in thoracic CT scans of cats. Journal of Small Animal Practice, accepted.

Lamb, C.R., Nelson, J.R., 2015. Diagnostic accuracy of tests based on radiologic measurements of dogs and cats: a systematic review. Veterinary Radiology and Ultrasound 56, 231-244.

Lamb, C.R., Parry, A.T., Baines, E.A., Chang, Y.M., 2011. Does changing the orientation of a thoracic radiograph aid diagnosis of rib fractures? Veterinary Radiology and Ultrasound 52, 75-78.

Lamb, C.R., Pfeiffer, D.U., Mantis, P., 2007. Errors in radiographic interpretation made by veterinary students. Journal of Veterinary Medical Education 34, 157-159.

Lamb, C.R., Tyler, M., Boswood, A., Skelly, B.J., Cain, M., 2000. Assessment of the value of the vertebral heart scale in the radiographic diagnosis of cardiac disease in dogs. Veterinary Record 146, 687-690.

Leib, M.S., Larson, M.M., Grant, D.C., Monroe, W.E., Troy, G.C., Panciera, D.L., Rossmeisl, J.H., Werre, S.R., 2012. Diagnostic utility of abdominal ultrasonography in dogs with chronic diarrhea. Journal of Veterinary Internal Medicine 26, 1288-1294.

Levinson, J.G., Bouma, J.L., Althouse, G.C., Rieser, T.M., 2009. Prevalence of malignancy when solitary versus multiple lesions are detected during abdominal ultrasonographic examination of dogs with spontaneous hemoperitoneum: 31 cases (2003-2008). Journal of Veterinary Emergency and Critical Care 19, 496-500.

Libby, D.M., Smith, J.P., Altorki, N.K., Pasmantier, M.W., Yankelevitz, D., Henschke, C.I., 2004. Managing the small pulmonary nodule discovered by CT. Chest 125, 1522-1529.

Loy, C.T., Irwig, L., 2004. Accuracy of diagnostic tests read with and without clinical information: a systematic review. JAMA 292, 1602-1609.

Moons, K.G., Harrell, F.E., 2003. Sensitivity and specificity should be de-emphasized in diagnostic accuracy studies. Academic Radiology 10, 670-672.

Moynihan, R., Doust, J., Henry, D., 2012. Preventing overdiagnosis: how to stop harming the healthy. BMJ 344.

Myers, N.C., 1997. Adrenal incidentalomas - Diagnostic workup of the incidentally discovered adrenal mass. Veterinary Clinics of North America-Small Animal Practice 27, 381-399.

Nemanic, S., London, C.A., Wisner, E.R., 2006. Comparison of thoracic radiographs and single breath-hold helical CT for detection of pulmonary nodules in dogs with metastatic neoplasia. Journal of Veterinary Internal Medicine 20, 508-515.

O'Neill, D.G., Church, D.B., McGreevy, P.D., Thomson, P.C., Brodbelt, D.C., 2014a. Prevalence of disorders recorded in cats attending primary-care veterinary practices in England. Veterinary Journal 202, 286-291.

O'Neill, D.G., Church, D.B., McGreevy, P.D., Thomson, P.C., Brodbelt, D.C., 2014b. Prevalence of disorders recorded in dogs attending primary-care veterinary practices in England. Plos One 9, e90501.

Peterson, C., 1999. Factors associated with success or failure in radiological interpretation: diagnostic thinking approaches. Medical Education 33, 251-259.

Russo, M., Lamb, C.R., Jakovljevic, S., 2000. Distinguishing rhinitis and nasal neoplasia by radiography. Veterinary Radiology and Ultrasound 41, 118-124.

Sackett, D.L., Haynes, R.B., 2002. The architecture of diagnostic research. BMJ 324, 539-541.

Sacornrattana, O., Dervisis, N.G., McNiel, E.A., 2013. Abdominal ultrasonographic findings at diagnosis of osteosarcoma in dogs and association with treatment outcome. Veterinary and Comparative Oncology 11, 199-207.

Sistrom, C.L., 2009. The appropriateness of imaging: a comprehensive conceptual framework. Radiology 251, 637-649.

Snyder, J.M., Shofer, F.S., Van Winkle, T.J., Massicotte, C., 2006. Canine intracranial primary neoplasia: 173 cases (1986-2003). Journal of Veterinary Internal Medicine 20, 669-675.

Stowater, J.L., Lamb, C.R., Schelling, S.H., 1990. Ultrasonographic features of canine hepatic nodular hyperplasia. Veterinary Radiology 31, 268-272.

Wallace, M., Selmic, L., Withrow, S.J., 2013. Diagnostic utility of abdominal ultrasonography for routine staging at diagnosis of skeletal OSA in dogs. Journal of the American Animal Hospital Association 49, 243-245.

Warren-Smith, C.M.R., Andrew, S., Mantis, P., Lamb, C.R., 2012. Lack of associations between ultrasonographic appearance of parenchymal lesions of the canine liver and histological diagnosis. Journal of Small Animal Practice 53, 168-173.

Weinstein, S., Obuchowski, N.A., Lieber, M.L., 2005. Clinical evaluation of diagnostic tests. American Journal of Roentgenology 184, 14-19.

731 Welch, H.G., Frankel, B.A., 2011. Likelihood that a woman with screen-detected breast
732 cancer has had her "life saved" by that screening. Archives of Internal Medicine 171,
733 2043-2046.
734
735 Whiting, P., Rutjes, A.W., Reitsma, J.B., Glas, A.S., Bossuyt, P.M., Kleijnen, J., 2004.
736 Sources of variation and bias in studies of diagnostic accuracy: a systematic review.
737 Annals of Internal Medicine 140, 189-202.
738
739 Williams, E.D., 2000. Guest Editorial: Two proposals regarding the terminology of thyroid
740 tumors. Int Journal of Surgical Pathology 8, 181-183.
741
742 Williams, L.E., Packer, R.A., 2003. Association between lymph node size and metastasis in
743 dogs with oral malignant melanoma: 100 cases (1987-2001). Journal of the American
744 Veterinary Medical Association 222, 1234-1236.

745

746 **Table 1**. Possible results of binary tests

	Test result	
Patients	+	-
Disease present	TP	FN
Disease absent	FP	TN

747 TP, true positive; FN, false negative; FP, false positive; TN, true negative.

748

Figure legends

Fig. 1. Schematics illustrating the effect of positive and negative test results on the probability of disease. (A) For an accurate diagnostic test (sensitivity = 90% and specificity = 90%), a pre-test probability of 50% (circle) is increased to 90% by a positive test result (triangle) and decreased to 10% by a negative test result (square). (B) The most marked increase in post-test probability occurs with a positive result for a test of high specificity whereas the most marked decrease in post-test probability occurs with a negative result for a test of high sensitivity.

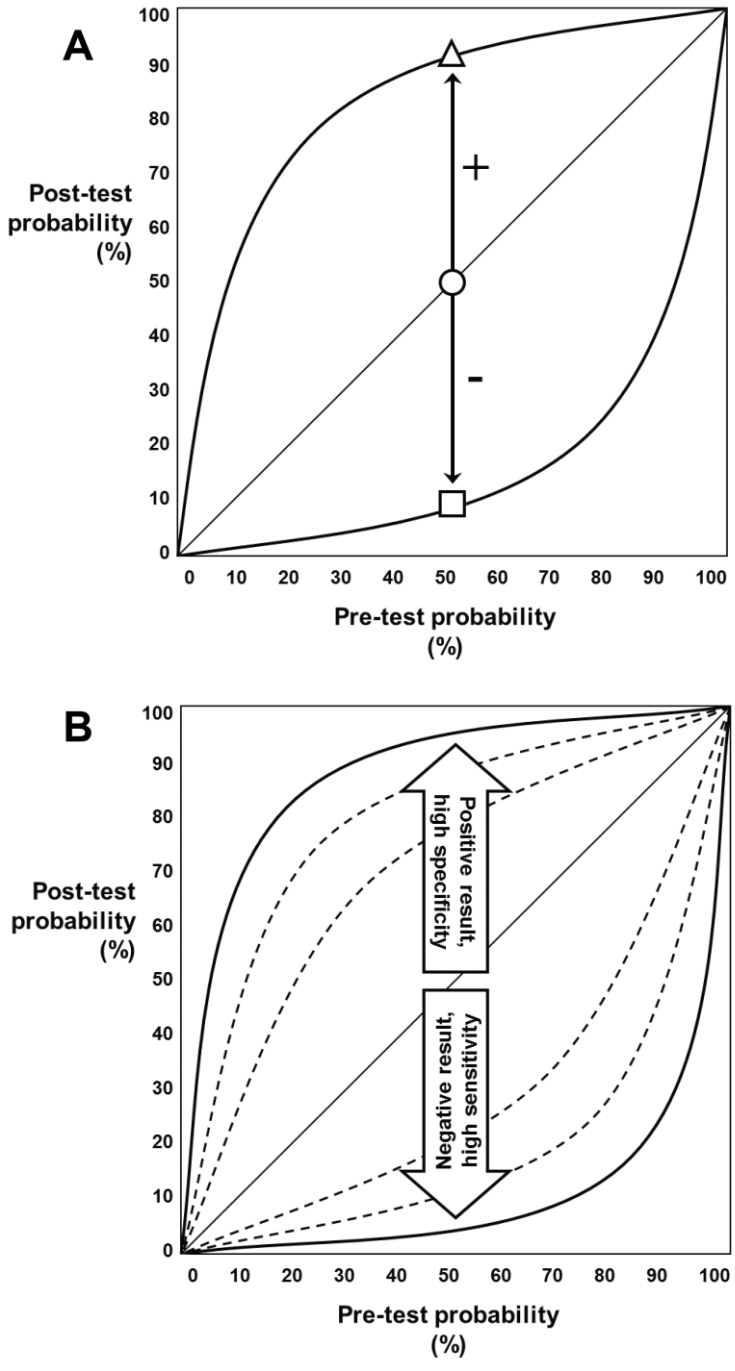


Fig. 2. A negative CT scan does not rule out the possibility of pulmonary metastasis. Tranverse CT images of a St. Bernard dog with osteosarcoma of the right distal radius. Initial scan (A) appears normal, but repeat scan (B) only 3 months later shows multiple pulmonary metastases.

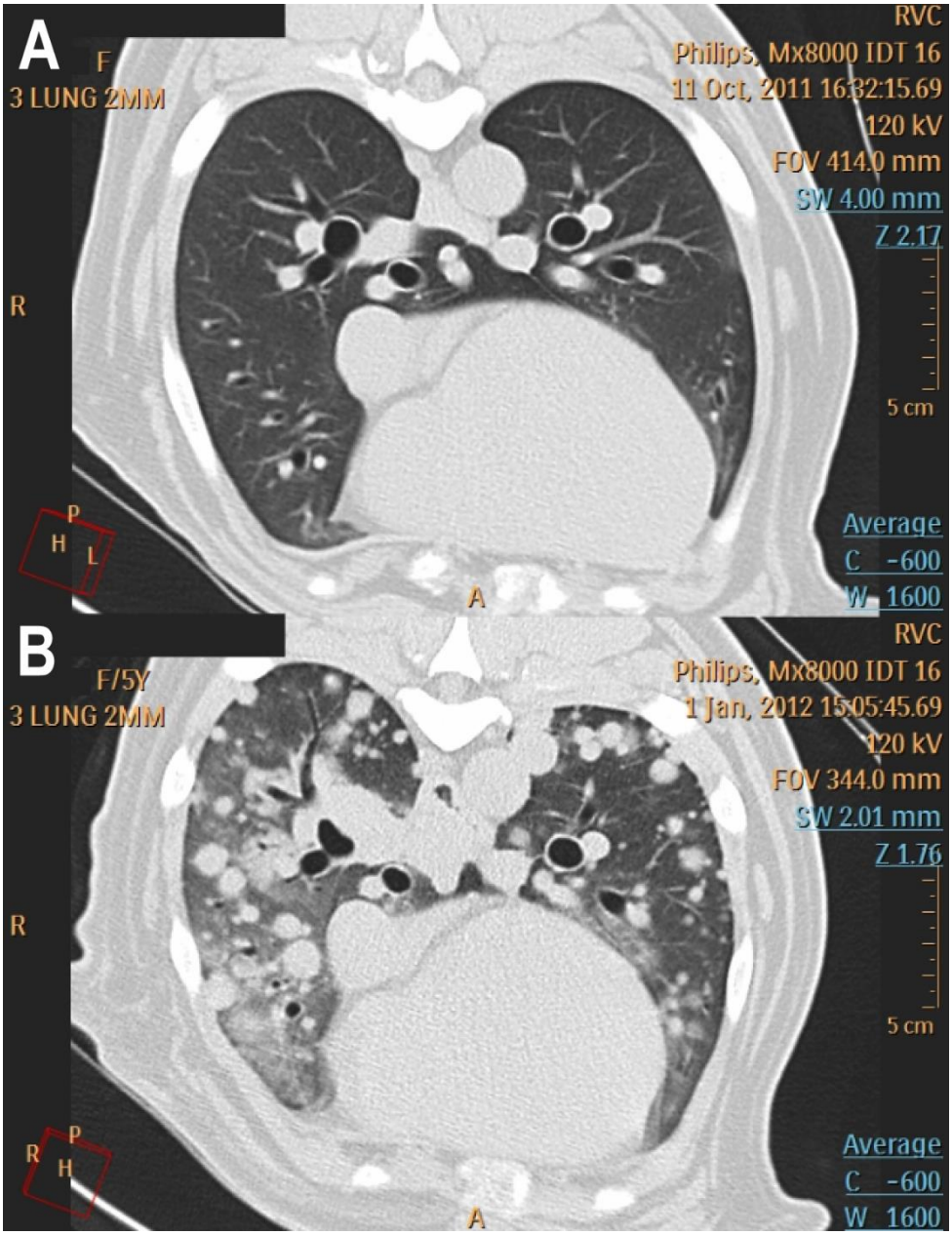
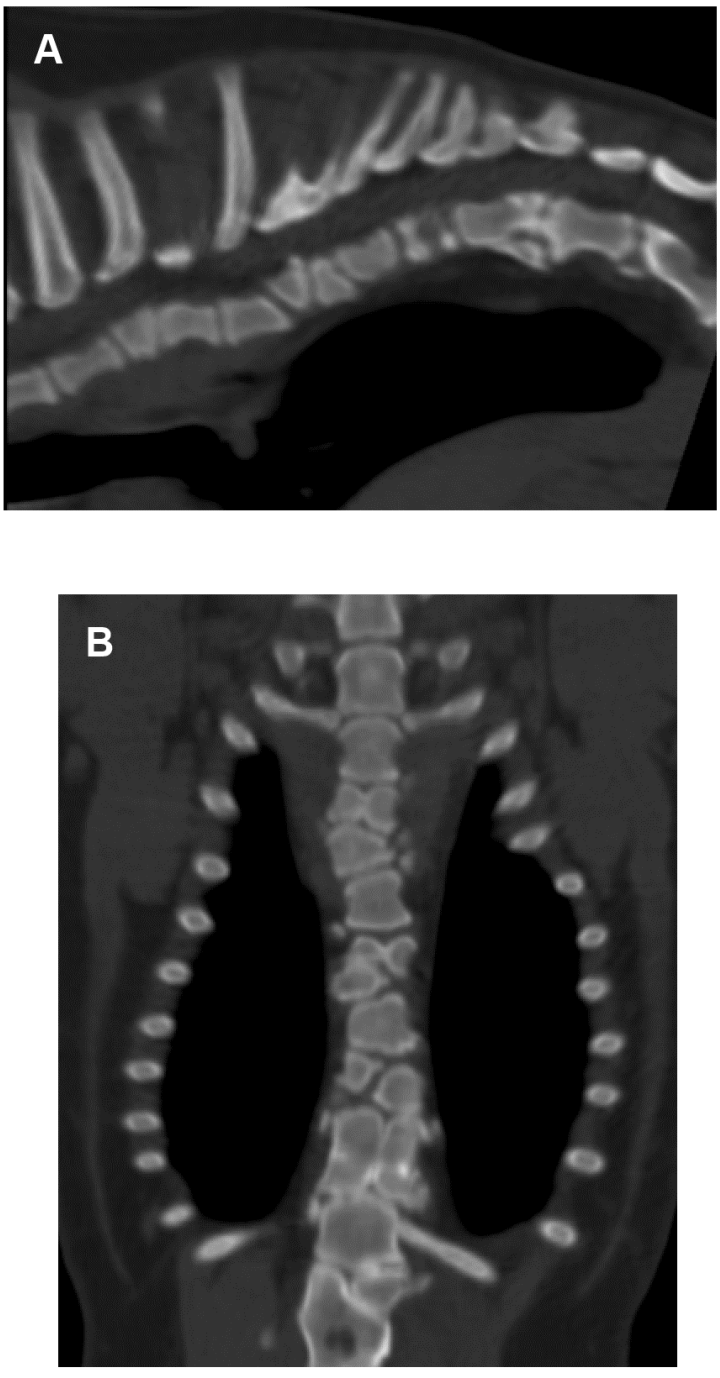
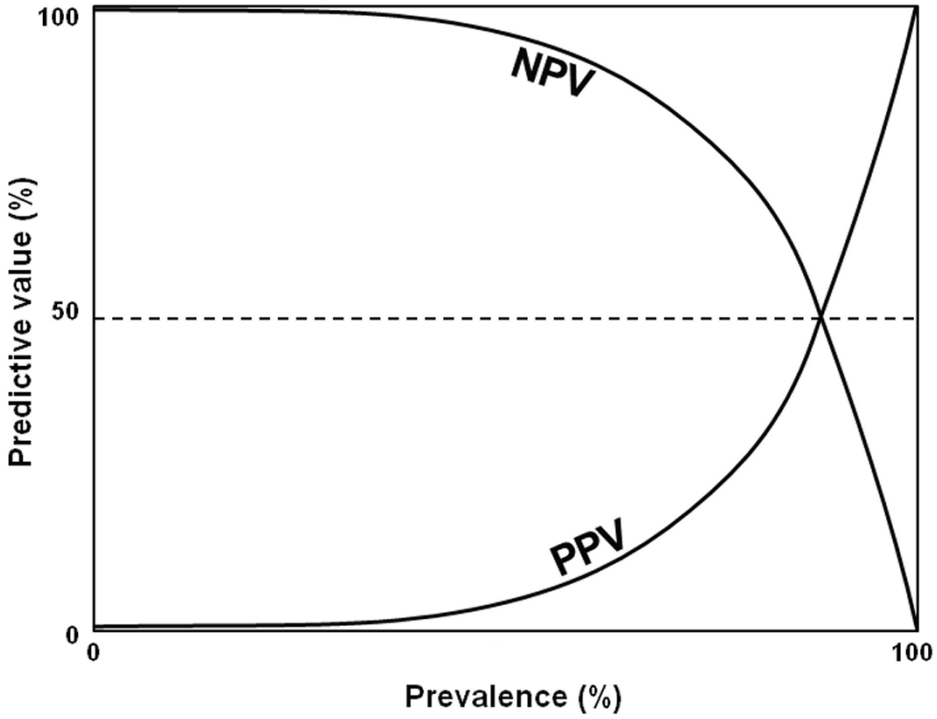


Fig. 3. Example of a subclinical condition that could be misinterpreted as the cause of clinical signs. Sagittal (A) and dorsal (B) CT images of the thoracic spine of a French bulldog with signs of spinal pain. Multiple hemivertebrae are present, but these usually represent a subclinical finding in this breed. In this instance, the clinical signs were related to a cervical disc extrusion.



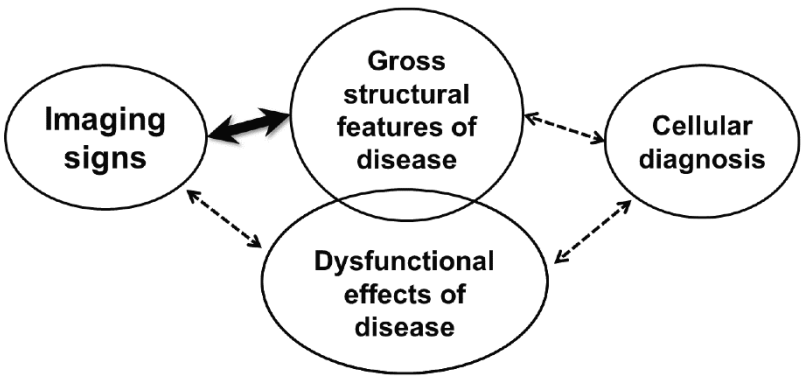
768 **Fig. 4.** Schematic illustrating the relationship between the predictive value of a test result and prevalence of
769 disease. Positive predictive value (PPV) is low when prevalence is low. Unless the prevalence of disease is
770 relatively high, a positive test result is likely to be a false positive (i.e. predictive value <50%). The opposite is
771 true for negative predictive (NPV) value.



772

773

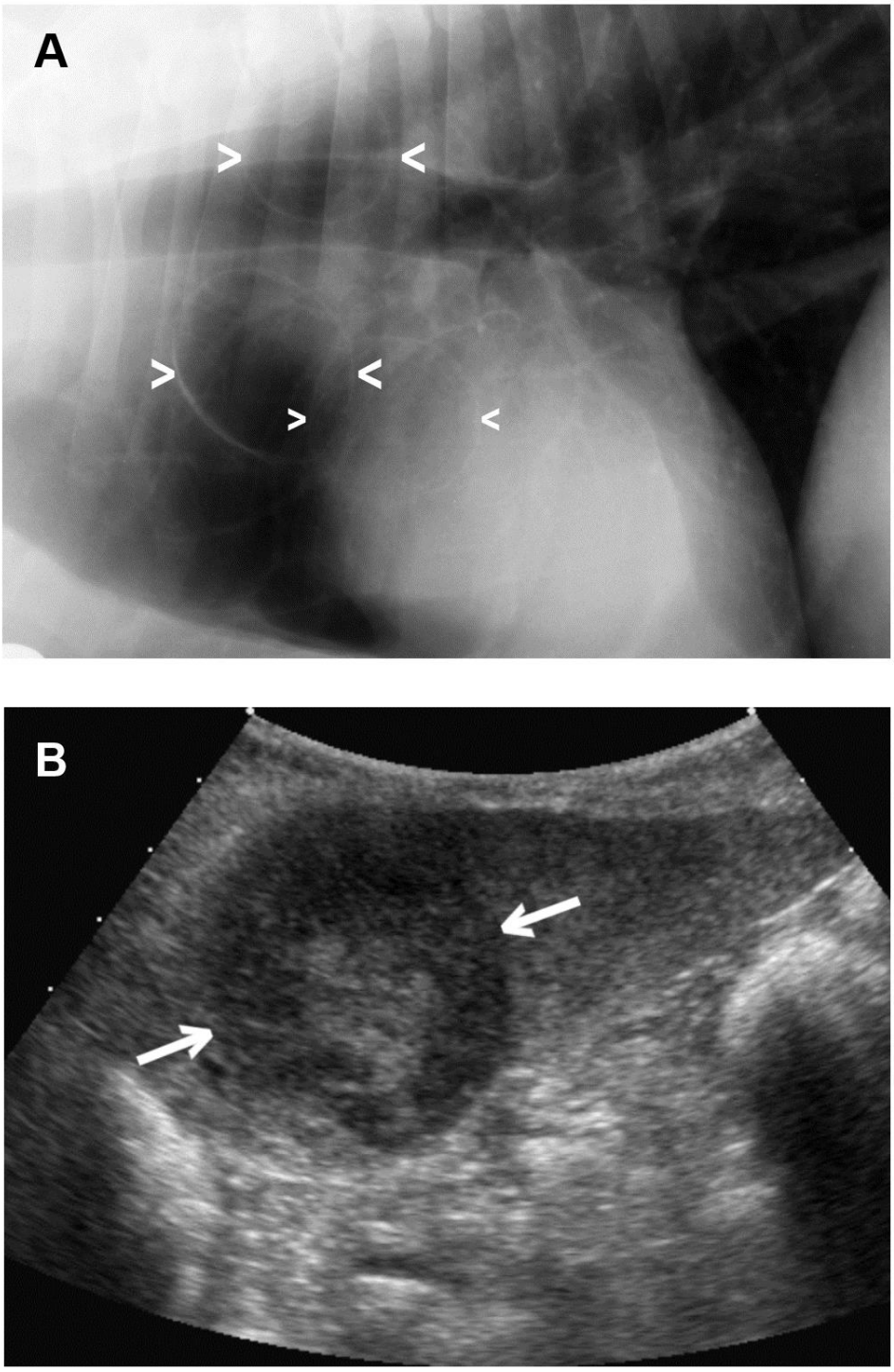
774 **Fig. 5.** The imaging signs associated with a specific disease will be most closely related to its gross
775 (macroscopic) structural features, less closely related to its dysfunctional effects and indirectly related to the
776 cellular features that are the basis for the pathological diagnosis.



777

778

779 **Fig. 6.** Examples of incidental findings. (A) Multiple pulmonary bullae (arrowheads) in a thoracic radiograph of
780 a dog with a cough that resolved with conservative treatment. (B) Splenic mass (arrows) discovered during
781 comprehensive work-up of a dog with dysrhythmia. The dysrhythmia resolved spontaneously and the splenic
782 mass was subsequently proved to be a haematoma.



783