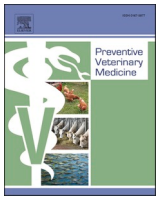
Preventive Veterinary Medicine 217 (2023) 105970

Contents lists available at ScienceDirect

Preventive Veterinary Medicine

journal homepage: www.elsevier.com/locate/prevetmed



Relative diagnostic accuracy of point-of-care tests to rule-in *Dirofilaria immitis* infection in clinically suspect dogs: A systematic review and meta-analysis

Peter J. Atkinson [*], Ryan O'Handley, Torben Nielsen [1], Charles GB Caraguel [1]

University of Adelaide, Roseworthy Campus, Mudla Wirra Rd, Roseworthy, SA 5371, Australia

ARTICLE INFO

Keywords:
Canine heartworm
Diagnostic test evaluation
Dirofilaria immitis
Meta-analysis
Point-of-care

ABSTRACT

Canine heartworm, *Dirofilaria immitis*, can cause severe disease and sometimes death of the host. Associated clinical signs, lack of preventative usage and regional endemicity are unlikely sufficient by themselves to reach a definitive diagnosis. Several point-of-care (POC) diagnostic tests are commercially available to aid in-clinic diagnosis, however, there is variable diagnostic accuracy reported and no synthesis of published evidence. This systematic review aims at meta-analysing the likelihood ratio of a positive result (LR^+) to inform the selection and interpretation of POC tests in practice to rule-in heartworm infection when there is clinical suspicion. Three literature index interfaces (Web of Science, PubMed, Scopus) were searched on November 11th, 2022, for diagnostic test evaluation (DTE) articles assessing at least one currently commercialised POC test. Risk of bias was assessed adapting the QUADAS-2 protocol and articles with no evidence of high risk of bias were meta-analysed if deemed applicable to our review objective. Substantial between DTE heterogeneity was investigated including potential threshold or covariate effects. A total of 324 primary articles were sourced and 18 were retained for full text review of which only three had low risk of bias in all four QUADAS-2 domains. Of the nine heartworm POC tests evaluated, only three, IDEXX SNAP (n DTEs = 6), Zoetis WITNESS (n DTEs = 3) and Zoetis VETSCAN (n DTEs = 5) could be analysed. Both WITNESS and VETSCAN DTEs showed substantial heterogeneity due to a putative threshold effect and no summary point estimates could be reported. SNAP DTEs showed acceptable heterogeneity, and a summary LR^+ was estimated at 559.0 (95%CI: 24.3–12,847.4). The quality and heterogeneity of heartworm POC test DTEs was highly variable which restricted our summary of the diagnostic accuracy to only the SNAP test. A positive result from the SNAP test provides strong evidence of the presence of an infection with adult heartworm(s) in a dog patient and this test is warranted to rule-in clinical suspicion(s) in clinics. However, our review did not appraise the literature to assess the fitness of SNAP test, or any other POC tests, to rule-out heartworm infection in dogs without clinical suspicion or following heartworm therapy.

1. Introduction

The canine heartworm, *Dirofilaria immitis*, is globally distributed with reports in all continents apart from Antarctica (Simón et al., 2012), and report frequency increases towards warmer subtropical and tropical regions (Wang et al., 2014; Bowman et al., 2016). *D. immitis* is a parasite of both wild and domestic canids, causing a progressive illness leading to pulmonary hypertension and *cor-pulmonale* (right-sided congestive heart failure), which can be fatal if left untreated (Bowman and Atkins, 2009).

Dirofilaria immitis colonises the pulmonary arteries and right atrium at the sub-adult and adult stages (McCall et al., 2008). Once mature, adults sexually reproduce, and females release microfilaria into the blood stream which later can be ingested by feeding mosquitoes. Microfilaria must go through two development stages within a mosquito before being deposited close to the vector feeding site on the host's skin (Knight and Lok, 1998). They then migrate into the skin puncture and

Abbreviations: POC, point-of-care; DSe, diagnostic sensitivity; DSp, diagnostic specificity; LR, diagnostic likelihood ratio; DTE, diagnostic test evaluation; COI, conflict of interest; HSROC, hierarchical summary receiver operator characteristic, PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; QUADAS-2, Quality Assessment of Diagnostic Accuracy Studies 2.

* Corresponding author.
E-mail address: peter.atkinson@adelaide.edu.au (P.J. Atkinson).

[1] Equal contribution to the supervision of this work

https://doi.org/10.1016/j.prevetmed.2023.105970
Received 7 February 2023; Received in revised form 20 June 2023; Accepted 23 June 2023
Available online 29 June 2023

undergo maturation and moults in the host muscle tissues (Lichtenfels et al., 1985), finally maturing and living in the pulmonary vasculature, a process that takes at least six months (McCall et al., 2008; Bowman and Atkins, 2009). Heartworms originate exclusively exogenously from mosquito-based transmission and cannot replicate endogenously (i.e. within dog replication is not possible without passage through a mosquito).

Point-of-care (POC) tests may be used to either rule-in or rule-out heartworm infection and Fig. 1 contextualises the various clinical scenarios where those tests may be used. We considered the three ruling-in scenarios (red) to be the most common occurrence in practice compared to the two ruling-out scenarios (yellow), and therefore we aimed this review on the fitness of the tests to rule-in infection. Although the history (e.g. staying in an endemic area without preventative coverage) or the clinical presentation (e.g. *cor pulmonale*) of a dog may be suggestive of *D. immitis* infection, they are unlikely sufficient on their own to rule-in the condition. POC tests are commonly used in veterinary practice as complementary diagnostic evidence to support the aetiology (*D. immitis* infection) and reach a final diagnosis (dirofilariosis). To rule-in infection, we want to maximise our confidence in a positive result, and therefore minimise the chance of a false positive result, i.e. favouring diagnostic specificity (DSp, accuracy of a test among non-infected) over diagnostic sensitivity (DSe, accuracy of a test among infected). To assess the fitness-for-purpose of the test, it is further recommended to use diagnostic likelihood ratio (LR) to balance both DSp and DSe (Caraguel and Colling, 2021). Indeed, any test with excellent DSp would only be suitable disease ruling-in if it also provides an acceptable DSe and a test with a slightly lower DSp can still provide stronger evidence on a positive result if its DSe is much superior. Therefore, we focused on reviewing and comparing the LR of a positive test result (LR^+) when assessing the fitness of POC tests to rule-in heartworm infection. When assessing tests to rule-out infection, the focus should be on the LR of a negative test result (LR^-).

Until the late 20th century, antemortem diagnosis of canine dirofilariosis relied on microscopic visualisation of circulating microfilaria using non-concentrating direct blood smear, concentrating modified Knott's test, or microscopic screening of buffy coat from a microhematocrit tube (Knott, 1939; Martini et al., 1991). The DSe, as well as the reproducibility, of these methods are accepted as imperfect due to natural variability in microfilaraemia in infected hosts, including within-day variation (Evans et al., 2017; Ionică et al., 2017). Alternative POC tests have been considered to detect the presence of other heartworm stages, in particular adults that cause clinical disease. In the late 1980 s, antibody-based tests were developed to detect uterine antigens released by adult female worms into the host's bloodstream (Weil et al., 1985; Goodwin, 1998). Diagnostic companies have commercialised those antigen tests either as enzyme-linked immunosorbent assay (ELISA), lateral flow immunoassay (LFI) or indirect fluorescent antibody test (IFAT) kits. Whilst several test variants have come and gone from the market since 1986 (Rohrbach and Patton, 2013), the detection of female uterine antigen has been the mainstay POC testing method since 1998, with diagnostic companies continuing today to offer LFI testing kits. The detection of the antigen suggests the presence of at least one sexually mature female worm, however, it will miss immature females or male-only infections (Goodwin, 1998). High filarial burden is necessary to induce clinical disease (Polizopoulou et al., 2000; McCall et al., 2008), and those burdens require repeated infection events with multiple larvae that most likely would include at least one female. The detection of adult female heartworm associated antigen is therefore expected in a diseased host, however, adult female worms may be present without clinical disease.

Several diagnostic test evaluation (DTE) studies have reported variable diagnostic accuracy estimates within and between POC tests targeting the same antigen (Starkey et al., 2017; Henry et al., 2018). Reasons for the variability in accuracy may include the selection of study subjects, inherent threshold effect, analytical changes to testing kits over time or choice of reference standard. For instance, a reference standard commonly used in early DTEs was necropsy, considered by some as a perfect standard (Atkins, 2003; Ranjbar-Bahadori et al., 2007; Genchi et al., 2012; Gruntmeir et al., 2021). Recently, non-lethal plate-based ELISAs were also used as reference standards. The two plate-based ELISA assays currently available are PetCHEK® (IDEXX Laboratories) and DiroCHEK® (Zoetis Inc), although these too rely on detection of the uterine antigen and likely have similar limitations to POC tests in detection of immature female, or male only infections (IDEXX Laboratories; Zoetis Inc) as well as some degree of conditional dependence with the POC kits. Due to the variability of accuracy estimates reported, and the current lack of synthesis, the aim of this report was to systematically review and meta-analyse accessible DTE articles of canine heartworm POC tests and assess their fitness for accurate confirmation of heartworm infection in clinically suspect dogs.

2. Methods

This report complies with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses of Diagnostic Test Accuracy

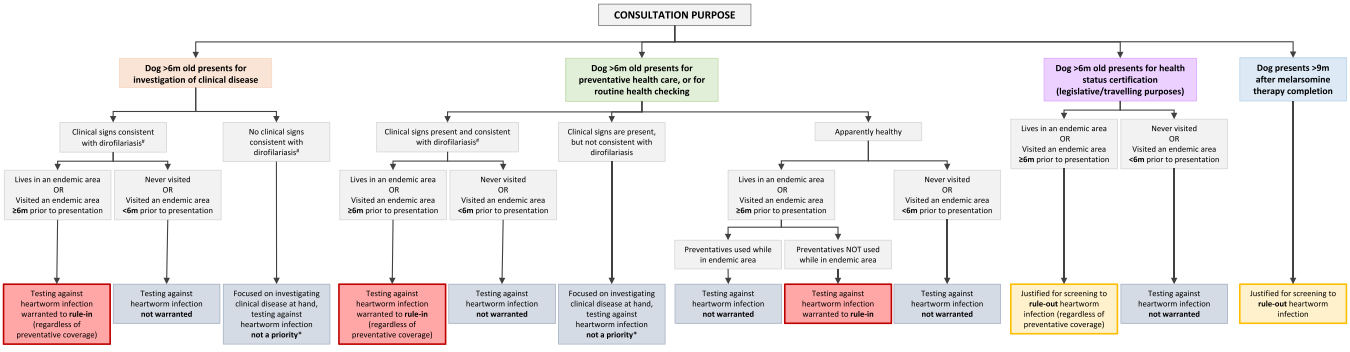


Fig. 1. Scenario tree representing the four consultation scenarios leading to testing a dog against canine heartworm (*Dirofilaria immitis*) infection in practice. Our meta-analysis focused on summarising the accuracy of current commercialised point-of-care tests to rule-in (confirm) heartworm infection in clinically suspected patients (three branches ending with a red box). It did not focus on ruling-out (screening) the infection (two branches ending with a yellow box). Branches ending with a blue box did not warrant heartworm POC testing. Those targeted clinical scenarios to rule-in infection were compared to the clinical profile of study subjects to assess the applicability of diagnostic test evaluation articles to the meta-analysis.# The following clinical presentations are consistent with dirofilariosis in dogs, with at least one sign required to be shown for inclusion in that category: (i) mild to moderate form - cough, exercise intolerance, abnormal lung sounds; (ii) moderate to severe form - dyspnoea, abnormal heart and lung sounds, cor-pulmonale, hepatomegaly, syncope, ascites; and (iii) caval syndrome: acute onset severe weakness with associated haemoglobinaemia and haemoglobinuria.*Whilst heartworm infection is possible, resources (time, labour, finances) may restrict the investigation to disorder(s) related to presenting clinical signs.

(PRISMA-DTA) (McInnes et al., 2018). The protocol of this review was not previously registered.

2.1. Information sources and search strategy

A systematic search of three literature search platforms, Web of Science, PubMed and Scopus was conducted. The search strategy was adapted from the Cochrane Handbook for DTA Reviews guidelines (de Vet et al., 2008). The following search strings were used:

- Web of Science, all databases selected: *(TS=(dog) AND TS= (heartworm OR "Dirofilaria immitis" AND TS=(test*) AND TS= (accuracy OR sensitivity OR specificity))*
- PubMed: *((dog) AND (heartworm OR "Dirofilaria immitis") AND (test*) AND (accuracy OR sensitivity OR specificity))*
- Scopus: *(TITLE-ABS-KEY (dog) AND TITLE-ABS-KEY (heartworm OR "Dirofilaria immitis") AND TITLE-ABS-KEY (test*) AND TITLE-ABS-KEY (accuracy OR sensitivity OR specificity))*

No Subject Headings were used. We also screened reference lists of relevant articles for additional articles that were not captured in the initial search, by checking any reference to previous diagnostic accuracy measures mentioned. The search strategy included any article up to the search date of November 11th, 2022.

2.2. Article selection

The articles yielded from all databases were imported into the systematic review web-based tool Covidence® (Covidence systematic review software, 2022) for ease of duplicate removal and collaborative screening. Articles were first screened for topic relevance based on title and abstract and were excluded if they clearly did not assess a diagnostic test or were not addressing *D. immitis* diagnosis. If deemed relevant, full text appraisal was undertaken for inclusion into the systematic review. We differentiate here an individual 'diagnostic test evaluation' (DTE) from a 'DTE study' and from a 'DTE article'. A DTE is defined as the evaluation of the diagnostic accuracy of a single index test against a single reference standard using a single set of study subjects. A DTE study is defined as the evaluation of one or more index tests against one single reference standard using a single set of study subjects. A DTE article is defined as the report of one or more DTE studies. Therefore, a given DTE article may include more than one DTE study and multiple DTEs. DTE articles not available in English, reporting secondary research (reviews, book chapters), with purpose irrelevant to assess a diagnostic test accuracy, such as prevalence surveys and cohort studies on treatment efficacy, or published prior the commercialisation of the antigen detection technology in 1998 (Goodwin, 1998) were excluded. Additional exclusion criteria were DTE studies investigating POC tests in host species other than domestic dogs (wild canids, felids) or not comparing a POC test to a reference standard. Finally, DTEs assessing index tests that were laboratory-based (non POC), not currently available or deviating from manufacturer instructions were also excluded. A POC test is defined here as a self-contained assay, providing an ante-mortem result within 30 min (Busin et al., 2016).

2.3. Article appraisal

As guided by the Cochrane Collaboration guidelines (Whiting et al., 2011), the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) protocol was adapted to appraise each eligible article for risk of bias across the four following domains: subject selection (named patient selection in QUADAS-2), index test, reference standard and flow and timing. For each domain, the QUADAS-2 assessment criteria were refined and extended to fit into a classification tree with each scenario branch leading to classification into 'low', 'possible' or 'high' risk of study bias. DTEs with high risk of bias were excluded from consideration for the meta-analysis while low risk DTEs were assessed for applicability to our review before consideration for the meta-analysis. DTEs with possible risk of bias were also assessed for applicability and a possible effect of bias was explored using covariate analysis in the meta-analysis.

2.3.1. Subject selection domain

This domain was assessed at the 'DTE study' level because it corresponds to one set of study subjects, regardless of the number of index tests used on those subjects (see DTE study definition above). The selection bias of concern for this domain is called *spectrum bias* (Ransohoff and Feinstein, 1978). This bias occurs when the disease profile, including alternative diseases or co-morbidities, of the study subjects is not matching the profile of individuals in the target population. Spectrum bias is expected when there is a misrepresentation of the i) range of severity of the target condition or ii) type and severity of plausible co-morbidities among the case subjects, or iii) alternative (or similar) diseases among the non-case subjects (Rutjes et al., 2005). For instance, DTEs with experimentally challenged subjects are unlikely to be representative of the natural disease spectrum and constitute an unacceptably high risk of spectrum bias (Greiner and Gardner, 2000). Spectrum bias is also likely to occur when case and non-case study subjects are recruited separately from source populations with distinct clinical profiles (e.g. cases selected from clinic patients and non-cases from the general apparently healthy population). Rutjes et al. (2005) referred to this discriminative selection approach as a 'two-gate' subject selection which we considered associated with an unacceptably high risk of spectrum bias. In contrast, DTEs using a 'one-gate' subject selection (i.e. case and non-cases selected from the same or similar clinical populations) are more likely associated with a lower risk of bias (Greiner and Gardner, 2000). Whilst some authors consider any 'case-control' subject selection (i.e. study subjects selected disproportionally from the source population based on their index test or reference standard results) to represent high risk of spectrum bias (Lijmer et al., 1999), we only considered this to be an unacceptably high when the subjects were selected disproportionally based on their index test results. We acknowledge the resource saving capabilities of selection based on reference standard results, especially for a condition where common reference standard is consequential like necropsy. We further investigated the contribution of subject selection as a source of heterogeneity when found. The classification tree which guided our DTE study's classification, either as 'high', 'possible' or 'low' risk of spectrum bias, is accessible in the Supplementary Material Fig. S1.

2.3.2. Index test domain

This domain was assessed at the 'DTE' level where one index test is evaluated against one reference standard (see DTE definition above). Threshold selection bias (Whiting et al., 2013) was not applicable here as the test operator cannot knowingly select a threshold, given all commercialised heartworm POC tests provide binary test results. We focused our appraisal on the possible presence of test review bias (Begg, 1987) resulting from lack of, or insufficient blinding of the index test's operator(s) to the reference standard results or, to the source or any other information about the study subjects. Prior knowledge about the reference standard result or any other subject's information may change how an operator runs a test, how much scrutiny they assess the absence (or presence) of development of a positive result or whether an index test is repeated if there is a discordant result. LFIs are expected to be highly robust, i.e. operator or conditions independent (Crowther et al., 2006; Waugh and Clark, 2021) and we considered lack of blinding to only cast a 'possible' risk of test review bias. The potential impact of no or possible blinding was subsequently investigated in the meta-analysis when substantial heterogeneity was found. The classification tree used to guide our DTE's classification, either as 'possible' or 'low' risk of test review bias, is accessible in the Supplementary Material Fig. S2.

2.3.3. Reference standard domain

This domain was assessed at the 'DTE' level. Reference standards that do not classify subjects perfectly will introduce some information bias. A reference standard can rely on one or more detection methods (mixture or combination). We deemed appropriate any reference standard that used the following single methods because they detect an analytical target consistent with the diagnostic target of interest (i.e. infection with adult heartworm, or not): necropsy macroscopic examination of the pulmonary vascularisation (direct visual detection), plate-based ELISA (uterine antigen detection) and PCRs that detect nucleic acids from either *D. immitis* or its obligate endosymbiont bacteria *Wolbachia spp*. When more than one method is used, we deemed not acceptable to include an index test in the composition of the reference standard. Those DTEs were classified as at high risk of *incorporation bias* because they provide a partisan evaluation of the index test(s). Although commonly used in combination with other methods, microfilarial detection methods, were not considered accurate enough (Mylonakis et al., 2004) and DTEs using this method were classified at high risk of *inappropriate reference standard bias* (Whiting et al., 2004). DTEs that changed detection methods according to the subject's background were also classified as at high risk of inappropriate reference standard bias because study subjects were not classified consistently and independently. As for the index test, a *diagnostic review bias* may have occurred when the operator(s) of a reference standard were not blinded to the index results or to other information about the study subjects (Ransohoff and Feinstein, 1978). Those DTEs were classified as 'possible' risk of bias and their accuracy estimates were compared to blinded DTE in the meta-analysis if substantial heterogeneity was detected. The classification tree used to guide our DTE's classification, either as 'high', 'possible' or 'low' risk of reference standard related information bias, is accessible in the Supplementary Material Fig. S3.

2.3.4. Flow and timing domain

This domain was assessed at the 'DTE' level. Several types of selection or information bias may occur in this domain. First, *disease progression* and *treatment paradox biases* can occur when the health status of interest has changed between running the index test and running the reference standard (Whiting et al., 2004). This is of no concern when the same specimen is used for all tests or when all specimens are collected at the same time. However, if the time lag between runs is not null, common knowledge of the canine heartworm occurrence, progression and eventual regression was used to appraise the risk of the subject status to change. Next, if the fraction of subjects verified by the reference standard depended on their index test results, the risk of *partial verification bias* (Whiting et al., 2004) was classified as high, or low otherwise. When distinct reference standards were used to verify study subject index test results, we classified the DTE at high risk of differential verification bias (Whiting et al., 2004). This bias may lead to an inappropriate exclusion of subjects negative to the index test by one reference standard, or positive to the index test by another reference standard, and tends to select extreme or unambiguous cases and exaggerates test accuracy estimates (Kohn et al., 2013). Finally, *withdrawal* or *uninterpretable test results biases* can occur when subjects' results are unproportionally excluded from the data analysis (Whiting et al., 2013). DTEs fitting these criteria were classified at high risk of bias and were excluded from the meta-analysis. For instance, DTE where some subjects were purposively excluded from the analysis based on their heartworm burden found at necropsy were deemed at high risk of withdrawal bias. The classification tree which guided our DTE's classification, either as 'high' or 'low' risk of bias, is accessible in the Supplementary Material Fig. S4.

2.4. Applicability assessment

This domain was assessed at the 'DTE study' level. Our review aimed at summarising POC test diagnostic accuracy when confirming of heartworm infection in clinically suspected dog subjects and, to be classified as low review applicability concerns, a DTE study needed to select subjects that would have raised clinical suspicion in practice. We used the scenario tree in Fig. 1 to contextualise the three scenarios where a consulting dog should raise clinical suspicion of heartworm infection and would warrant rule-in testing with a POC test. DTE studies with missing or insufficient detail on subject selection were classified as having 'possible' applicability concerns. However, studies with dogs sourced from non-endemic areas, receiving heartworm preventatives, or screened to rule-out the infection were classified as having high review applicability concerns and were not included in the meta-analysis. For the WITNESS POC kit, only studies dated after 2015 were included in the meta-analysis because this kit was substantially modified to improve its performance at that time (Henry et al., 2018).

2.5. Data gathering

Raw or cross-tabulated (2 ×2 contingency table) count data corresponding to true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) were extracted for each pair of index test/reference standard reported in applicable DTEs. DTEs which reported only diagnostic sensitivity (DSe = TP/TP+FN) and diagnostic specificity (DSp = TN/TN+FP) estimates required back calculation to obtain count data. Fraction values were rounded to the nearest whole integer. POC test under evaluation, subject selection method, reference standard used, blinding methods, author conflict of interest (COI), location of study and female *D. immitis* burden (when available). Author COI was recorded as author affiliated (AA), funding associated (FA) or not detected (ND). The Review Manager (RevMan) software (Review Manager 5.4, Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2020) was used to centralise bias and applicability appraisal following a standard QUADAS-2 template as well as raw data entry. RevMan generated summary figures to compare the DTEs' appraisal across the QUADAS-2 domains. For each reviewed DTE (index test/reference standard pair), RevMan reproduced DSe and DSp calculations from the count data and estimated the corresponding 95% CIs. RevMan also generated separate forest plots for DSe and DSp. Diagnostic likelihood ratios of a positive and negative test result (LR^+ = DSe/(1-DSp) and LR^- = (1-DSe)/DSp, respectively) and their 95%CI were computed separately from the count data using the command *diagt* in the statistical package Stata 17.1 (StataCorp 17.1, TX: StataCorp LLC, 2021). For DTEs with either no false positives (100% DSp), or no false negatives (100% DSe), the lower or upper bounds of the 95% CI of LR^+ and LR^- were calculated using a stochastic approach from DSe and DSp 95% confidence intervals (see Supplementary Materials SM2. for details).

2.6. Meta-analysis of tests' accuracy

Accuracy estimates were synthesised using Stata 17.1 for a given index test when a minimum three DTEs were deemed applicable to our review. The Stata code and dataset are accessible here -https://doi.org/10.25909/21767390. Between DTE heterogeneity was first assessed using Zhou and Dendukuri's I^2 statistic (fraction of total variance attributable to the between DTEs variance) estimated separately for the DSe and DSp estimates (I^2_{DSe} and I^2_{DSp}, respectively) (Zhou and Dendukuri, 2014). According to Harbord et al. (2008), the hierarchical summary receiver operating characteristic (HSROC) modelling approach is the most suitable for conducting meta-analyses of diagnostic test accuracy. This approach accounts for two unique features of diagnostic test meta-analysis; i) subjects within each DTE are likely to have a degree of similarity that is greater than the similarity of subjects between DTEs and ii) the DSe and DSp demonstrated in each DTE are possibly correlated. The HSROC model allows us analytical control of these two factors, and was used through the Stata command *metadta* (Nyaga and

Arbyn, 2022) after installation. A cut-off value < 25% for both I^2_{DSe} and I^2_{DSp} indicated low heterogeneity (Higgins et al., 2003) and summary point estimates of the index test accuracy were reported after modelling diagnostic check-up. Model diagnostics were run using the Stata command *midas* (Dwamena, 2007) which uses the bivariate modelling approach, requiring at least four DTEs. Potential outliers and residual deviation were explored using residual goodness-of-fit and assessing Cook's distance (Viechtbauer and Cheung, 2010), which were available when at least four DTEs were present. Publication bias was assessed using Deek's funnel plot asymmetry (Deeks et al., 2005; Van Enst et al., 2014). When evidence of substantial heterogeneity was found from either I^2_{DSe} or I^2_{DSp}, a potential threshold effect was first investigated. A correlation coefficient (ρ) < −0.5 between DSe and DSp estimates suggested strong evidence of threshold effect and a summary receiver operator characteristic (SROC) curve was generated and reported instead of summary points. In the absence of evidence of threshold effect (ρ > −0.5), between DTE heterogeneity was further investigated by building a multifactorial meta-regression to explore the potential contribution of DTE covariates of concern including author COI (detected or not), subject selection approach, reference standard choice and reported blinding. If a set of covariates significantly explained the observed heterogeneity, the report was restricted to summary point estimates of the desirable covariate pattern (combination of covariate levels supporting low risk of study bias) if it included sufficient DTE count (at least three) and residual heterogeneity was negligible (I^2 < 25%). When residual heterogeneity could not be controlled, the range of accuracy estimates across DTEs was reported.

3. Results

3.1. Article selection for review

The steps and counts of article selection are summarised in Fig. 2. Our original search returned 324 unique articles of which 205 were excluded based on title and abstract screening, due to topic irrelevance. The full text of 119 articles were assessed, and a further 103 were excluded based eligibility criteria (see Fig. 2 for details). Sixteen articles were selected for the systematic review and two additional articles were found from reference lists resulting in a total of 18 DTE articles reviewed, including 20 DTE studies (two articles reported on two sperate study populations) and 41 DTEs (up to five index test vs. reference standard pairwise comparisons per DTE study). The diagnostic accuracy of a total of nine commercial POC tests were reported involving necropsy, plate-based ELISAs (PetCHEK or DiroCHEK), PCR or experimental infection as reference standards. The characteristics of the 18 articles under review are summarised in Table 1. All but five articles were published from the USA. Publication year ranged from 2001 to 2022 with 12 articles published in the last 10 years. Only seven articles had no evidence of COI.

3.2. Article bias appraisal

3.2.1. Subject selection

We report here our appraisal of subject selection at the DTE study level (n = 20). Most studies sourced dog subjects from the USA (n = 14) and the rest from Europe (n = 3), Middle East (n = 2) or South America (n = 1). Approximately half of the studies (n = 8) used a cross-sectional selection of subjects, while another ten used case-control selection (n = 7 one-gate recruitment, n = 1 two-gate recruitment, n = 2 possible two-gate recruitment). One study was suspected to have used case-control selection (one-gate) but there was insufficient information to classify with certainty, and one study experimentally infected subjects, using laboratory bred controls. Most studies (n = 16) did not report the intended purpose for testing of subjects. For three studies, the testing was initiated by the subjects being clinically suspected to be infected,

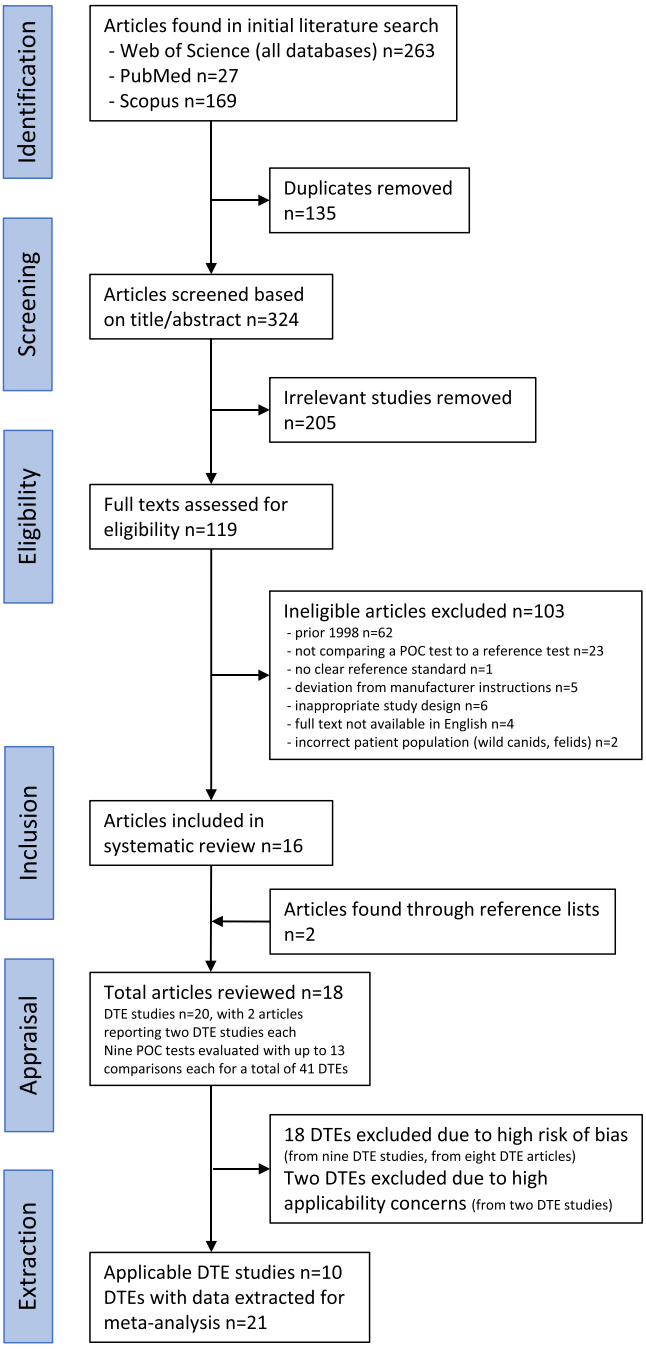


Fig. 2. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow chart to summarise the outcomes of our search and selection of articles evaluating the diagnostic accuracy of canine heartworm (*Dirofilaria immitis*) point-of-care (POC) tests to detect clinically suspect cases. Exclusion reasons and count are listed. An article may report the evaluation of more than one index tests in comparison to more than one reference standard. This resulted *in fine* into a higher count of diagnostic test evaluations (DTEs) that compares the results of one index test to one reference test than the total count of reviewed articles.

and in one study testing was due to monitoring an experimental infection (off-clinic purpose). Overall, the DTE study using experimentally challenged dogs (Genchi et al., 2018) was classified at high risk of spectrum bias as well as the study that unequivocally used a two-gate recruitment (Starkey et al., 2017). Seven DTE studies were classified with possible spectrum bias and the rest as low (Fig. 3, Supplementary Materials Table S-Table 1).

Table 1
Characteristics of diagnostic test evaluation articles selected for the systematic review of diagnostic accuracy of point-of care tests against canine heartworm.

DTE Article	*Location*	*Subject selection*	*No. of DTE studies/No. of DTEs*	*Index test(s) assessed*									*Reference standard (s)*	*Possible conflict of interest*[†]
				SNAP® Heartworm RT	*WITNESS® Heartworm Rapid*	*VETSCAN® Heartworm Rapid*	*Solo Step® CH*	*Anigen Rapid®*	*ALERE™ Dirofilriose*	*Speed Diro™*	*Accuplex®*	*AbboScreen*		
Courtney & Zeng, 2001	USA	Case-control* (one-gate)	1/4	√	√		√					√	Necropsy	ND
Atkins, 2003	USA	Case-control (one-gate)	1/3	√		√	√						Necropsy	ND
Ranjbar-Bahadori et al., 2007	Iran	Cross-sectional	1/1		√								Necropsy	ND
Chandrashekar et al., 2010	USA	Case-control (two-gate*)	1/1	√									Necropsy (cases) & PetCHEK (controls)	AA (IDEXX)
Lee et al., 2011	USA	Case-control (one-gate)	2/4	√/√		√/√							Necropsy (a) or PetCHEK (b)[#]	AA (IDEXX)
Aron et al., 2012	USA	Case-control (one-gate)	1/4	√	√	√	√						Necropsy	AA (Virbac)
Eberts, 2013	USA	Case-control (one-gate)	1/2	√							√		Necropsy (cases) & PetCHEK (controls)	ND
Stillman et al., 2014	USA	Case-control (two-gate*)	1/1	√									Necropsy (cases) & PetCHEK (controls)	AA (IDEXX)
Rojas et al., 2015	Costa Rica /Israel	Cross-sectional	1/1			√							PCR	ND
Starkey et al., 2017	USA	Case-control (two-gate)	1/2	√	√								DiroCHEK	AA (IDEXX & Zoetis)
Genchi et al., 2018	Italy	Cross-sectional	2/4	√	√					√/√			PetCHEK (a) or experimental infection (b)^ [#]	AA (Virbac)
Henry et al., 2018	USA	Case-control (one-gate)	1/5	√	√	√	√	√					Necropsy	FA (Zoetis)
Liu et al., 2018	USA	Case-control (one-gate)	1/2	√		√							Necropsy (cases) & PetCHEK (controls)	AA (IDEXX)
Burton et al., 2020	USA	Cross-sectional	1/2	√		√							PetCHEK	AA, FA (IDEXX)
Lane et al., 2021	USA	Cross-sectional	1/2	√	√								DiroCHEK	AA, FA (Zoetis)
Becker et al., 2022	Hungary	Cross-sectional	1/1			√							PCR	ND
Soares et al., 2022	Brazil	Cross-sectional	1/1						√				PCR	ND
Beall et al., 2022	USA	Cross-sectional	1/1	√									PetCHEK	AA, FA (IDEXX)

* Presumed based on study settings, but not definitive due to insufficient information reported
^ Speed Diro was the only index test compared to experimental infection as a reference standard
[#] Notations *a* and *b* refer to diagnostic test evaluation studies extracted from the same article, indicating separate reference tests (analysed as separate DTE studies). These refer to the distinctions made in Fig. 3 for risk of bias and applicability concerns
[†] As reported in article text. Potential conflict of interest: not detected (ND), author affiliated (AA), funding associated (FA)

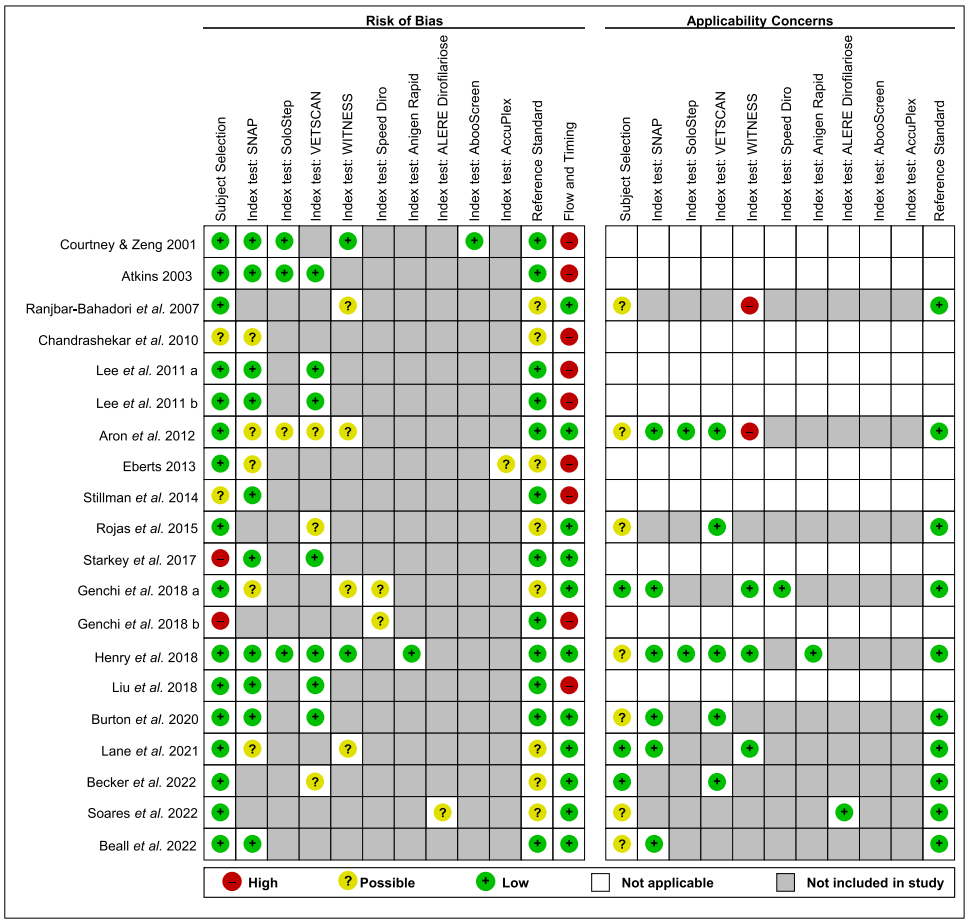


Fig. 3. Methodological quality summary for each diagnostic test evaluation (DTE) study for canine heartworm (*Dirofilaria immitis*) point-of-care test. The index test domain was assessed at the DTE level, and all others at the DTE study level. Risk of bias and applicability concerns summary assessed by authors as low, possible, or high. Applicability concerns are not shown for DTE studies with high risk of bias in any domain. Empty white boxes indicate assessment was not performed due to high risk of bias, and empty grey boxes indicate that index test was not included in the DTE stud. Lines with the same author and notated with *a* or *b* refer to DTE studies from the same DTE article comparing index tests to separate reference standards within the same article. The reference standards can be found in Table 1. Detailed classification information for each domain is provided in Supplementary Materials Table S-Table 1.

3.2.2. *Index tests*

The diagnostic accuracy of nine POC tests, currently commercialised by eight diagnostic companies, were evaluated with between one and 13 DTEs each. Those tests included Accuplex® (Antech Diagnostics), Anigen Rapid® (Bionote Co), ALERE™ Dirofilriose (Bionote Co), AbboScreen (Abbott Laboratories), SNAP® Heartworm RT Test (IDEXX Laboratories), WITNESS® Heartworm Rapid Test (Zoetis Inc), VETSCAN® Heartworm Rapid Test (Zoetis Inc, formerly Abaxis), Solo Step® CH (Heska) and Speed Diro™ (Virbac). For the remainder of this review, tests will be called Accuplex, Anigen Rapid, ALERE, AbboScreen, SNAP, WITNESS, VETSCAN, Solo Step and Speed Diro respectively. Because this was a selection criterion, all articles included in our review reported complying with the manufacturer instructions including nature of the specimen (blood, serum or plasma), kit storage and testing protocol. We report here the risk of bias in the index test domain at the DTE study level (n = 20) because all DTEs within a study had the same index test bias classification. Overall, 12 DTE studies were classified as possible risk of bias in the index test domain due to unreported blinding, and the rest as low risk of bias (Fig. 3 and Supplementary Materials Table S-Table 1).

3.2.3. *Reference standard*

Due to the range of reference standards found both between and occasionally within articles, we classified the risk of bias in the reference standard domain at the DTE study level. Five different reference standards, used solely or in combination, were reported across the 20 reviewed DTE studies – necropsy (n = 6), plate-based ELISA (PetCHEK®, n = 4; DiroCHEK®, n = 2), PCR (n = 3), experimental infection (n = 1), and a combination of necropsy for cases and PetCHEK for non-cases (n = 4), and we considered none of these approaches as a more or less appropriate reference standard for our review diagnostic target (infection with adult heartworm). Two articles reported two DTE studies each, using separate reference standards as well as study populations. Eight DTE studies (contributing to nine DTEs) had unreported blinding and were classified as of possible risk of diagnostic review bias, and 11 DTE studies (contributing 25 DTEs) were classified as of low risk of reference standard related bias because they reported blinding, or because the reference standard was run before the index test, so they could not be influenced by the index test results (Fig. 3, Supplementary Materials Table S-Table 1).

3.2.4. *Flow and timing*

We also report the risk of bias in the flow and timing domain at the DTE study level because all DTEs within a study received the same classification. Four DTE studies used a combination of post-mortem to select infected cases and PetCHEK® to select non-cases (controls), suggesting a high risk of differential verification bias. The one experimental infection study introduced high risk of disease progression bias, as the temporal lag between the experimental challenge and subsequent index testing was substantial (>6 months). Another four DTE studies excluded subjects from their analysis disproportionally based on high female *D. immitis* burden and, therefore, were considered to introduce high risk of withdrawal bias. Overall, nine DTE studies (contributing to 18 DTEs) were classified at high risk of bias and were excluded from further review consideration (Fig. 3, Supplementary Materials Table S-Table 1).

Following bias appraisal, 10 DTE studies including 20 DTEs and reported across eight articles, were excluded from further review because they were classified at high risk of bias in at least one of the four domains. This resulted into the loss of evidence for the diagnostic accuracy for Accuplex and AbboScreen POC tests (one article each).

3.3. Article applicability appraisal

Of the ten DTE studies assessed for applicability, most (n = 7/10) did not provide sufficient information outlining subject selection, and were classified as having 'possible' applicability concerns. The three remaining studies clearly described their subject selection, and matched our review objective, and were classified as having low applicability concerns in the subject selection domain. Two DTEs assessing the WITNESS POC test were classified as high applicability concerns because they assessed this test prior to kit improvements in 2015. Applicability concerns for the index test and reference standard domains were classified as low in all other DTE studies (Fig. 3, Supplementary Materials Table S-Table 1). Applicability appraisal resulted in the exclusion of two DTEs, with 19 DTEs (from 10 DTE articles and 10 DTE studies) considered in meta-analysis. A forest plot summarising each test DTE outcome is shown in Fig. 4.

3.4. Synthesis of results

There was only sufficient DTEs (at least three) to meta-analyse the diagnostic accuracy of SNAP (six DTEs), WITNESS (three DTEs) and VETSCAN (five DTEs).

3.4.1. SNAP

Minimal heterogeneity between the SNAP's DTEs ($I^2_{DSe} = 16\%$, $I^2_{DSp} = 4\%$) and limited evidence of threshold effect ($\rho = -0.20$) was found. Model diagnostics did not indicate any influential or outlier DTEs. There was no evidence of publication bias, as Deek's funnel plot was approximately symmetrical (*P*-value = 0.33). Summary estimates of DSe, DSp and LR^+ are 97.28% (95%CI: 94.06–98.78%), 99.83% (95% CI: 96.05–99.99%) and 559.0 (95% CI: 24.3–12,847.4), respectively.

3.4.2. WITNESS

Some evidence of mild heterogeneity between WITNESS's DSe estimates was found ($I^2_{DSe} = 30\%$ and $I^2_{DSp} = 17\%$). This could be explained by a threshold effect with a strong negative correlation found between DSe and DSp ($\rho = -0.73$). Model diagnostics were not possible to calculate (three DTEs present). Consequently, we report here the hierarchical summary receiver operator characteristic (HSROC) curve instead of point estimates (Fig. 5 A) for this test, however, the limited number of DTEs (n = 3) provide little certainty on the shape of the curve (i.e. non informative 95% prediction region).

3.4.3. VETSCAN

Moderate heterogeneity between VETSCAN's DTEs was found ($I^2_{DSe} = 53\%$ and $I^2_{DSp} = 41\%$). Model diagnostics did not show outlier DTEs influencing the analysis. There was strong negative correlation between DSe and DSp estimates, $\rho = -1.00$, precluding further covariate analysis or presentation of a HSROC curve for this test (Fig. 5B).

The full range of accuracy estimates for WITNESS and VETSCAN is reported in the forest plot (Fig. 4).

4. Discussion

The aim of this review was to summarise evidence on the diagnostic accuracy of canine heartworm POC tests used to rule-in infection with adult worms in clinically suspected dogs (i.e. confirmatory testing). The diagnostic accuracy of nine POC tests was reported in the accessed literature (41 DTEs from 18 articles) but it could only be summarised for the SNAP test. Although some of the reported evidence was not applicable (evaluated prior a major re-designing of WITNESS), too few (Anigen Rapid, ALERE, Solo Step and Speed Diro) or too variable (WITNESS and VETSCAN), the primary reason for not summarising test

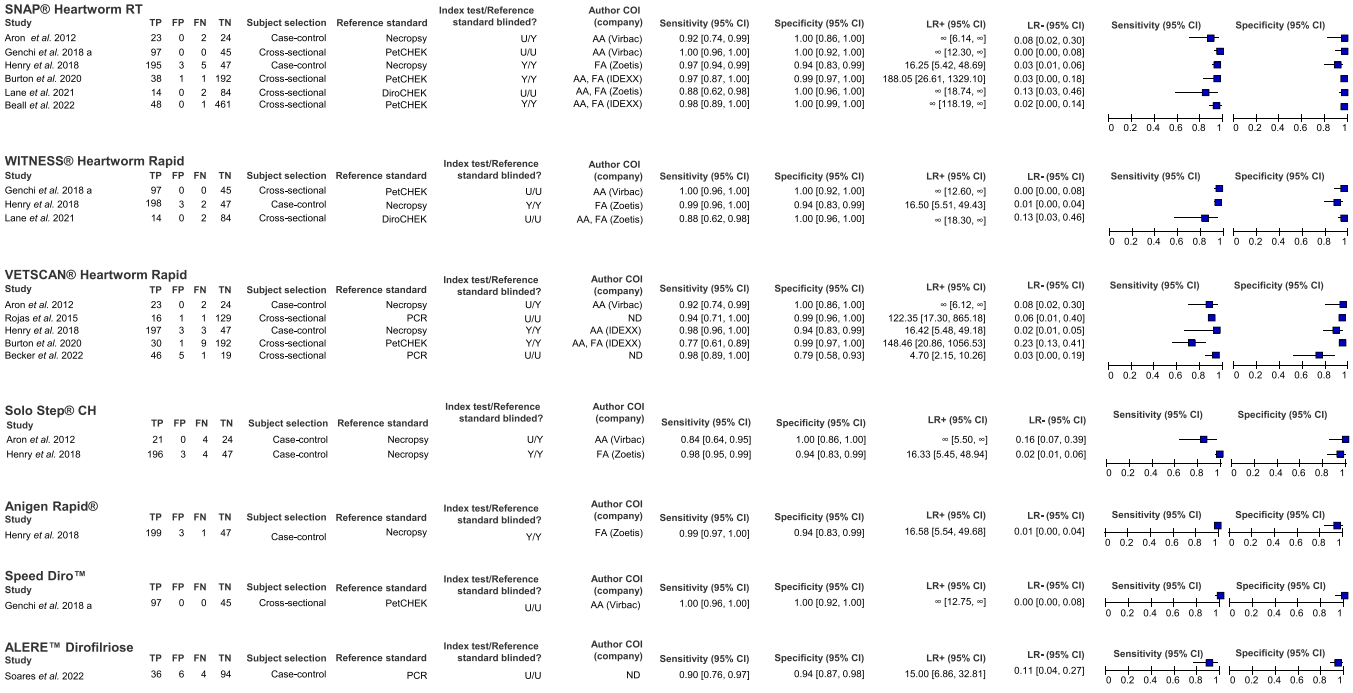


Fig. 4. Forest plot of raw data from diagnostic test evaluations (DTEs) assessing the diagnostic sensitivity and specificity of SNAP® Heartworm RT Test, WITNESS® Heartworm Rapid Test, VETSCAN® Heartworm Rapid Test, Solo Step® CH, Anigen Rapid®, Speed Diro™, ALERE™ Dirofilriose and AbboScreen® to test against canine heartworm (*Dirofilaria immitis*) with 95% confidence intervals (CI) in brackets. Likelihood ratio of a positive result and negative result (LR^+, LR^-), subject selection method, reference standard, blinding status of the index test and reference standard and conflict of interest (COI) are also included. Each DTE is named with the authors' last name and year of publication. Lines with the same author and notated with *a* or *b* refer to DTE studies comparing index tests to separate reference standards within the same article. The blue square indicates the sensitivity and specificity, and the horizontal black line depicts the 95% confidence interval. The DTEs are ordered by publication year. Plot adapted from Review Manager 5.4. Blinding status of yes = Y and unclear = U. COI was judged as either not detected = ND, author affiliated = AA or funding association = FA, and the company to which there was COI is reported in parentheses.

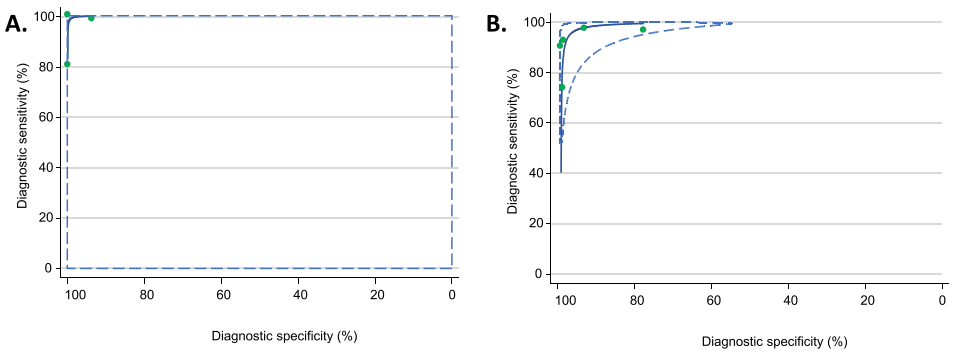


Fig. 5. Hierarchical summary receiver operator characteristic (HSROC) curves depicting the diagnostic accuracy of WITNESS® Heartworm Rapid Test **(A)** and VETSCAN® Heartworm Rapid **(B)** point-of-care tests over a range of positivity thresholds for the detection of circulating canine heartworm (*Dirofilaria immitis*) uterine antigen in canine serum. Presentation of the summary point and its associated 95% confidence region has been suppressed. The SROC line is shown as the solid blue line and the dashed blue line represents the 95% prediction region. Each green disk represents one diagnostic test evaluation (DTE).

accuracy evidence was the high risk of study bias in many reports (9/20 studies, 20/41 DTEs). Only three studies (eight DTEs) showed low risk of bias in all QUADAS domains. We excluded two DTE studies because of a high risk of bias in their subject selection (experimental challenge and two-gate sourcing). The majority of the other accessed DTE studies (11/20) used a one-gate case-control selection of subjects based on their reference standard (Rutjes et al., 2005) results which we deemed of acceptable risk of spectrum bias. Conventionally, any type of case-control selection of subjects in DTEs is regarded at high risk of bias (Lijmer et al., 1999). We justify this risk classification because when cases and non-cases are based on only the reference standard classification (i.e. true status), we expected that each subject groups represents truly infected and non-infected subjects respectively from the source population, although not overall. Therefore, this should not affect the spectrum of both groups' profile used to estimate DSe and DSp, respectively. This was further supported by the absence of evidence of case-control selection influencing the observed heterogeneity among the SNAP test DTEs, including systematically shifting DSe and DSp estimates.

Another potential source of bias we rejected was using separate reference standards for cases and non-cases (four articles), or the exclusion of subjects based on *D. immitis* burden (three articles). Differential verification could introduce information bias in an unpredictable direction when the reference standards are not perfect. Inclusion of the DTEs excluding subjects with high worm burden is likely to restrict the spectrum of infection to borderline detectable cases and therefore only provide DSe estimates for the most challenging cases. Overall, our review revealed that most DTEs of canine heartworm POC tests had several design deficiencies and in higher frequency than other recent systematic reviews of veterinary diagnostic tests (Buczinski et al., 2018; Roy et al., 2020; Fairley et al., 2021).

Blinding of test operators can have substantive effects on test performance (Cohen et al., 2016). Blinding of test operators was reported in half of the articles we reviewed (9/18). Recent veterinary DTE systematic reviews report highly variable blinding rates ranging from 17% to 82% (Buczinski et al., 2018; Li et al., 2022). Lack of reporting prevented us from determining whether other reviewed articles did not use blinding. In 2003, the Standards for Reporting of Diagnostic Accuracy Studies (STARD) guidelines were released and updated in 2015 (Bossuyt et al., 2003; Bossuyt et al., 2015). These outline essential DTE study information to report on and include a specific item to report whether the test operator(s) were blinded to the results of the other test(s) or any other clinical information about the subjects. Off the reviewed 16 articles published after the release of STARD first version in 2003, more than half (nine) did not report on blinding. Those articles were published in 11 journals of which only three currently required to comply with the STARD guidelines (*Veterinary Parasitology, Journal of the American Veterinary Medical Association* and *American Journal of Veterinary Research*). We wish to emphasise the importance of reporting blinding in DTE study (whether implemented or not) to allow a transparent appraisal of the evidence by the end POC users. It is likely that those studies did not use blinding, however, we could not appraise them beyond doubt and could only classify them with a 'possible' risk of bias.

We observed a chronological trend where non-lethal detection methods (ELISA and PCR) appeared to progressively replace necropsy as the reference standard in recent DTEs. We considered all reference tests used in the accessed DTEs were acceptable, however, we recognise that none are likely perfect (Wilks, 2001). For instance, PetCHEK has reported DSe estimates ranging from 85.2% (McCall et al., 2000) to 98.0% (IDEXX Laboratories) when compared to necropsy. Tests that target the same analyte are highly likely to be conditionally dependent and, therefore, to agree and possibly over-estimate POC accuracy (Vacek, 1985). It is therefore not surprising to observe SNAP with high DSe and DSp estimates when compared to the two plate-based ELISA methods. However, we also observe low heterogeneity between all SNAP DTEs, despite necropsy use as reference standard in two of the six DTEs meta-analysed, suggesting similar agreement between SNAP and a reference standard with an alternative analytical target.

Although two POC tests (WITNESS and VETSCAN) had sufficient applicable DTEs to synthesise, evidence of a threshold effect was found, disallowing presentation of single summary of LR^+ estimates. Presence of threshold effect for diagnostic tests is expected with assays providing a continuous output, and the test reader need to use a threshold value to binarise the test results. It is concerning to find evidence of threshold effect for binary tests, as selection of a known threshold (and thus knowledge of DSe and DSp at that threshold) is not possible for the test operator. As shown in Fig. 4, individual LR^+ estimates ranged from 16.50 to 4.70 (for WITNESS and VETSCAN respectively), with some DTEs yielding no false positive subjects (100% DSp), and, therefore, their LR^+ reaching infinity. However, we cannot exclude the possibility that LR^+ varies beyond those ranges due to some observed threshold effect. Causes of the observed threshold effects are possibly related to manufacturing or consumable inconsistencies as well as variable test performances based on individual DTE study conditions. We cannot recommend the use of WITNESS and VETSCAN tests at this stage based on the absence of consistent diagnostic accuracy evidence. We suggest investigating and correcting for the potential source(s) of those threshold effects before using those tests.

Ultimately, SNAP was the only POC test for which we could summarise the diagnostic accuracy. The SNAP LR^+ summary estimate ($LR^+ = 559.0$, 95% CI: 24.3–12,847.4) indicates that a positive SNAP test strongly supports the presence of an adult heartworm infection and is suitable to rule in dirofilariosis in dogs with clinical suspicion. This finding indicates a positive test result is 559 times more likely to occur in an infected than a non-infected subject (Hayden and Brown, 1999). The LR^+ value can be used to compare the added value of a SNAP positive result to other heartworm tests when ruling-in heartworm infection, however, we could not obtain those summary values for the other reviewed commercial POC tests.

One may consider using a POC heartworm test to rule-out infection either to certify heartworm-free status of a dog for legislative or travelling reasons, or to assure the parasite clearance post therapeutic

treatment (Fig. 1). In this case, a test with a low (closer to zero) diagnostic likelihood ratio for a negative result (LR^-) (suggestive of high DSe, and competitive DSp) would be suitable (Caraguel and Colling, 2021). Although the summary estimate of the LR^- for SNAP from our meta-analysis appeared strong (data not shown), it would be inappropriate to use this estimate for this purpose. Indeed, the SNAP meta-analysis did not consider DTEs with subjects that would have been a candidate for rule-out testing. To design a DTE study for the context of ruling-out infection, subjects would be recruited in a different manner. The spectrum of infection of those subjects is therefore expected to differ from dogs suspected to have infection and likely have different test accuracy. This appears to hold true in DTEs that excluded dogs with high adult female heartworm burden from their analysis and reported substantially lower DSe estimates (data not shown).

Diagnostic accuracy estimates can be combined to the pre-test probability to calculate the post-test probability of infection, also known as predictive value. Several tools are available to facilitate this process including user-friendly smartphone applications. As an example, the free iOS application DocNomo (iOS App Store) provides three graphical approaches to interpret a test result – the two-step Fagan's nomogram (Caraguel and Vanderstichel, 2013), a probability-modifying axis (Hayden and Brown, 1999), and a probability-modifying plot (Whiting et al., 2008). The summary diagnostic accuracy estimates of the SNAP test can be directly entered into DocNomo using the following hyperlink.

We developed here a transparent, objective and reproducible framework based on scenario trees to contextualise POC test use for heartworm infection in dogs and to classify study bias in reported DTEs. This approach should be considered to ease the review process as well as reporting of future diagnostic test accuracy systematic reviews and meta-analyses. We provide summary point estimates of diagnostic accuracy for the SNAP test, one of the most common POC heartworm test used in practice. The summary DSe, DSp and LR^+ point estimates, and their associated 95% confidence intervals were 97.28% (94.06–98.78%), 99.83% (96.05–99.99%) and 559.0 (24.3–12,847.4) respectively. Clinicians may use this information to select or interpret this test results when confirm suspected cases of canine heartworm disease. However, the reported estimates should not be used to make decisions on test selection or interpretation when ruling-out heartworm infection in dogs.

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

PJ Atkinson is the recipient of an Australian Research Training Program (RTP) scholarship funded by the Australian Government.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.prevetmed.2023.105970.

References

Aron, K., Mehra, R., Tockman, C., DeJong, K., 2012. Canine heartworm testing – Understanding low worm burden testing and a comparison of tests on the market. Vetcom Publication.
Atkins, C.E., 2003. Comparison of results of three commercial heartworm antigen test kits in dogs with low heartworm burdens. J. Am. Med. Assoc. 222, 1221–1223.
Beall, M.J., Mainville, C.A., Arguello-Marin, A., Clark, G., Lemieux, C., Saucier, J., Thatcher, B., Breitschwerdt, E.B., Cohn, L.A., Qurollo, B.A., Chandrashekar, R., 2022. An improved point-of-care ELISA for the diagnosis of Anaplasmosis and Ehrlichiosis during the acute phase of tick-borne infections in dogs. Top. Companion Anim. Med., 100735
Becker, Z., Holló, N., Farkas, R., Gyurkovszky, M., Reiczigel, J., Olaszy, K., Vári, Z., Vörös, K., 2022. Serodiagnostic difficulties and possibilities of heartworm disease in regions where both *Dirofilaria immitis* and *Dirofilaria repens* infections occur. Acta Vet. Hung. 92–99.
Begg, C.B., 1987. Biases in the assessment of diagnostic tests. Stat. Med. 6, 411–423.
Bossuyt, P.M., Reitsma, J.B., Bruns, D.E., Gatsonis, C.A., Glasziou, P.P., Irwig, L.M., Lijmer, J.G., Moher, D., Rennie, D., de Vet, H.C., 2003. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. BMJ 326, 41–44.
Bossuyt, P.M., Reitsma, J.B., Bruns, D.E., Gatsonis, C.A., Glasziou, P.P., Irwig, L., Lijmer, J.G., Moher, D., Rennie, D., De Vet, H.C.W., Kressel, H.Y., Rifai, N., Golub, R. M., Altman, D.G., Hooft, L., Korevaar, D.A., Cohen, J.F., 2015. STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. BMJ h5527.
Bowman, D.D., Atkins, C.E., 2009. Heartworm biology, treatment, and control. Vet. Clin. North Am. Small Anim. Prac. 39, 1127–1158.
Bowman, D.D., Liu, Y., McMahan, C.S., Nordone, S.K., Yabsley, M.J., Lund, R.B., 2016. Forecasting United States heartworm *Dirofilaria immitis* prevalence in dogs. Parasites Vectors 9, 540.
Buczinski, S., Gicquel, E., Fecteau, G., Takwoingi, Y., Chigerwe, M., Vandeweerd, J.M., 2018. Systematic review and meta-analysis of diagnostic accuracy of serum refractometry and brix refractometry for the diagnosis of inadequate transfer of passive immunity in calves. J. Vet. Intern. Med. 32, 474–483.
Burton, K.W., Liu, J., Drexel, J.P., Pulaski, C., Heartsill, S., Dupuy, C., 2020. Comparative evaluation of field samples using 2 in-clinic assays for heartworm antigen detection in dogs. Vet. Parasitol. 283, 109163.
Busin, V., Wells, B., Kersaudy-Kerhoas, M., Shu, W., Burgess, S.T.G., 2016. Opportunities and challenges for the application of microfluidic technologies in point-of-care veterinary diagnostics. Mol. Cell. Probes 30, 331–341.
Caraguel, C.G., Vanderstichel, R., 2013. The two-step Fagan's nomogram: ad hoc interpretation of a diagnostic test result without calculation. Evid. Based Med. 18, 125–128.
Caraguel, C.G.B., Colling, A., 2021. Diagnostic likelihood ratio - the next-generation of diagnostic test accuracy measurement. Rev. Sci. Tech. 40, 299–309.
Chandrashekar, R., Mainville, C.A., Beall, M.J., O'Connor, T., Eberts, M.D., Alleman, A. R., Gaunt, S.D., Breitschwerdt, E.B., 2010. Performance of a commercially available in-clinic ELISA for the detection of antibodies against *Anaplasma phagocytophilum*, *Ehrlichia canis*, and *Borrelia burgdorferi* and *Dirofilaria immitis* antigen in dogs. Am. J. Vet. Res 71, 1443–1450.
Cohen, J.F., Korevaar, D.A., Altman, D.G., Bruns, D.E., Gatsonis, C.A., Hooft, L., Irwig, L., Levine, D., Reitsma, J.B., De Vet, H.C.W., Bossuyt, P.M.M., 2016. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ 6, e012799.
Courtney, C.H., Zeng, Q.Y., 2001. Comparison of heartworm antigen test kit performance in dogs having low heartworm burdens. Vet. Parasitol. 96, 317–322.
Covidence systematic review software, 2022. Covidence systematic review software. Veritas Health Innovation, Melbourne, Australia.
Crowther, J.R., Unger, H., Viljoen, G.J., 2006. Aspects of kit validation for tests used for the diagnosis and surveillance of livestock diseases: producer and end-user responsibilities. Rev. Sci. Tech. 25, 913–935.
Deeks, J.J., Macaskill, P., Irwig, L., 2005. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J. Clin. Epidemiol. 58, 882–893.
Dwamena, B., 2007. MIDAS: Stata module for meta-analytical integration of diagnostic test accuracy studies. Stat. Softw. Compon.
Eberts, M., 2013. Performance of SNAP® 4Dx® plus and AccuPlex™4 in dogs with different heartworm burdens. Int. J. Appl. Res. Vet. Med. 11.
Evans, C.C., Burkman, E.J., Dzimianski, M.T., Moorhead, A.R., Savadelis, M.D., Angenendt, C., Zymny, S., Kulke, D., 2017. Periodicity of *Dirofilaria immitis* in longterm infections. Parasitol. Res. 116, 75–80.
Fairley, L., Smith, S., Maisrikrod, S., Henning, L., 2021. Systematic review and meta-analysis of diagnostic tests for diagnosis of melioidosis. Acta Trop. 214, 105784.
Genchi, C., Ferrari, N., Venco, L., Billi, M., Bataille, K.S., Genchi, M., 2012. Evaluation of the performance of heartworm antigen test kits and heartworm antibody test kits for the diagnosis of the infection in dogs and cats. Veterinaria 26, 35–40.
Genchi, M., Mangia, C., Ferrari, N., Loukeri, S., 2018. Evaluation of a rapid immunochromatographic test for the detection of low burden *Dirofilaria immitis* (heartworm) in dogs and cats. Parasitol. Res. 117, 31–34.
Goodwin, J.-K., 1998. The serologic diagnosis of heartworm infection in dogs and cats. Clin. Tech. Small Anim. Pr. 13, 83–87.
Greiner, M., Gardner, I.A., 2000. Epidemiologic issues in the validation of veterinary diagnostic tests. Prev. Vet. Med. 45, 3–22.
Gruntmeir, J.M., Thompson, N.M., Long, M.T., Blagburn, B.L., Walden, H.D.S., 2021. Detection of heartworm antigen without cross-reactivity to helminths and protozoa following heat treatment of canine serum. Parasites Vectors 14.

Harbord, R.M., Whiting, P., Sterne, J.A., Egger, M., Deeks, J.J., Shang, A., Bachmann, L. M., 2008. An empirical comparison of methods for meta-analysis of diagnostic accuracy showed hierarchical models are necessary. J. Clin. Epidemiol. 61, 1095–1103.

Hayden, S.R., Brown, M.D., 1999. Likelihood ratio: a powerful tool for incorporating the results of a diagnostic test into clinical decisionmaking. Ann. Emerg. Med. 33, 575–580.

Henry, L.G., Brunson, K.J., Walden, H.S., Wenzlow, N., Beachboard, S.E., L Barr, K., Long, M.T., 2018. Comparison of six commercial antigen kits for detection of *Dirofilaria immitis* infections in canines with necropsy-confirmed heartworm status. Vet. Parasitol. 254, 178–182.

Higgins, J.P., Thompson, S.G., Deeks, J.J., Altman, D.G., 2003. Measuring inconsistency in meta-analyses. BMJ 327, 557–560.

IDEXX Laboratories, PetChek® Canine Heartworm Antigen Test Kit.

Ionică, A.M., Matei, I.A., D'Amico, G., Bel, L.V., Dumitrache, M.O., Modry, D., Mihalca, A.D., 2017. *Dirofilaria immitis* and *D. repens* show circadian co-periodicity in naturally co-infected dogs. Parasites Vectors 10.

Knight, D.H., Lok, J.B., 1998. Seasonality of heartworm infection and implications for chemoprophylaxis. Clin. Tech. Small Anim. Pract. 13, 77–82.

Knott, J., 1939. A method for making microfilarial surveys on day blood. Trans. R. Soc. Trop. Med. Hyg. 33, 191–196.

Kohn, M.A., Carpenter, C.R., Newman, T.B., 2013. Understanding the direction of bias in studies of diagnostic test accuracy. Acad. Emerg. Med. 20, 1194–1206.

Lane, J.N., Litster, A., Little, S.E., Rodriguez, J.Y., Mwacalimba, K.K., Sundstrom, K.D., Amirian, E.S., Guerios, S.D., Serrano, M.A., Hays, K.M., Levy, J.K., 2021. Optimizing heartworm diagnosis in dogs using multiple test combinations. Parasites Vectors 14.

Lee, A.C., Bowman, D.D., Lucio-Forster, A., Beall, M.J., Liotta, J.L., Dillon, R., 2011. Evaluation of a new in-clinic method for the detection of canine heartworm antigen. Vet. Parasitol. 177, 387–391.

Li, R., Tian, X., Pang, J., Li, L., Yuan, J., Tian, Z., Wang, Z., 2022. Point-of-care tests for rapid detection of porcine epidemic diarrhea virus: a systematic review and meta-analysis. Viruses 14, 1355.

Lichtenfels, J.R., Pilitt, P.A., Kotani, T., Powers, K.G., 1985. Morphogenesi s of Developmental Stages of Dirofilaria immitis (Nematoda) in the Dog. In, Proceedings of the Helminthological Society of Washington, 98–113.

Lijmer, J.G., Mol, B.W., Heisterkamp, S., Bonsel, G.J., Prins, M.H., van der Meulen, J.H., Bossuyt, P.M., 1999. Empirical evidence of design-related bias in studies of diagnostic tests. J. Am. Med. Assoc. 282, 1061–1066.

Liu, J., Drexel, J., Andrews, B., Eberts, M., Breitschwerdt, E., Chandrashekar, R., 2018. Comparative evaluation of 2 In-clinic assays for vector-borne disease testing in dogs. Top. Companion Anim. Med. 33, 114–118.

Martini, M., Poglayen, G., Capelli, G., Roda, R., 1991. Diagnosis of canine filariosis: relative sensitivity and specificity of some haematological techniques. Angew. Parasitol. 32, 133–136.

McCall, J., Supakorndej, N., Donoghue, A.R., Turnbull, R.K., 2000. Study to determine the performance of commercially available canine heartworm antigen test kits. In, American Association of Veterinary Parasitologists 45th Annual Meeting, Salt Lake City, Utah.

McCall, J.W., Genchi, C., Kramer, L.H., Guerrero, J., Venco, L., 2008. Chapter 4: Heartworm Disease in Animals and Humans. Adv. Parasitol. Academic Press,, pp. 193–285.

McInnes, M.D.F., Moher, D., Thombs, B.D., McGrath, T.A., Bossuyt, P.M., Clifford, T., Cohen, J.F., Deeks, J.J., Gatsonis, C., Hooft, L., Hunt, H.A., Hyde, C.J., Korevaar, D. A., Leeflang, M.M.G., Macaskill, P., Reitsma, J.B., Rodin, R., Rutjes, A.W.S., Salameh, J.P., Stevens, A., Takwoingi, Y., Tonelli, M., Weeks, L., Whiting, P., Willis, B.H., 2018. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. J. Am. Med. Assoc. 319, 388–396.

Mylonakis, M.E., Papadopoulos, E., Koutinas, A.F., Paitaki, C., Leontides, L., 2004. Comparative methodology for the detection and differentiation of circulating microfilariae of *Dirofilaria immitis* in the dog. J. Helminthol. 78, 137–140.

Nyaga, V.N., Arbyn, M., 2022. Metadta: a Stata command for meta-analysis and meta-regression of diagnostic test accuracy data - a tutorial. Arch. Public. Health 80, 95.

Polizopoulou, Z.S., Koutinas, A.F., Saridomichelakis, M.N., Patsikas, M.N., Desiris, A.K., Roubies, N.A., Leontidis, L.S., 2000. Clinical and laboratory observations in 91 dogs infected with *Dirofilaria immitis* in northern Greece. Vet. Rec. 146, 466–469.

Ranjbar-Bahadori, S., Eslami, A., Bokaic, S., 2007. Evaluation of different methods for diagnosis of *Dirofilaria immitis*. Pak. J. Bio. Sc. 10, 1938–1940.

Ransohoff, D.F., Feinstein, A.R., 1978. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. N. Engl. J. Med. 299, 926–930.

Rohrbach, B.W., Patton, S., 2013. Effects of diagnostic test accuracy and treatment efficacy on the occurrence of suspected failure of heartworm prophylaxis in dogs. J. Vet. Intern. Med 27, 791–797.

Rojas, A., Rojas, D., Montenegro, V.M., Baneth, G., 2015. Detection of *Dirofilaria immitis* and other arthropod-borne filarioids by an HRM real-time qPCR, blood-concentrating techniques and a serological assay in dogs from Costa Rica. Parasites Vectors 8, 170.

Roy, A., Infantes-Lorenzo, J.A., de la Cruz, M.L., Domínguez, L., Álvarez, J., Bezos, J., 2020. Accuracy of tuberculosis diagnostic tests in small ruminants: A systematic review and meta-analysis. Prev. Vet. Med. 182, 105102.

Rutjes, A.W., Reitsma, J.B., Vandenbroucke, J.P., Glas, A.S., Bossuyt, P.M., 2005. Case–control and two-gate designs in diagnostic accuracy studies. Clin. Chem. 51, 1335–1341.

Simón, F., Siles-Lucas, M., Morchón, R., González-Miguel, J., Mellado, I., Carretón, E., Montoya-Alonso, J.A., 2012. Human and animal dirofilariasis: The emergence of a zoonotic mosaic. Clin. Microbiol. Rev. 25, 507–544.

Soares, L.A., Matias, I.C., Silva, S.S., Ramos, M.E.O., Silva, A.P., Barretto, M.L.M., Brasil, A.W.L., Silva, M.L.C.R., Galiza, G.J.N., Maia, L.A., 2022. Parasitological, serological and molecular diagnosis of *Dirofilaria immitis* in dogs in Northeastern Brazil. Exp. Parasitol. 236–237, 108233.

Starkey, L.A., Bowles, J.V., Payton, M.E., Blagburn, B.L., 2017. Comparative evaluation of commercially available point-of-care heartworm antigen tests using well-characterized canine plasma samples. Parasites Vectors 10, 475.

Stillman, B.A., Monn, M., Liu, J., Thatcher, B., Foster, P., Andrews, B., Little, S., Eberts, M., Breitschwerdt, E.B., Beall, M.J., Chandrashekar, R., 2014. Performance of a commercially available in-clinic ELISA for detection of antibodies against *Anaplasma phagocytophilum, Anaplasma platys, Borrelia burgdorferi, Ehrlichia canis,* and *Ehrlichia ewingii* and *Dirofilaria immitis* antigen in dogs. J. Am. Vet. Med. Assoc. 245, 80–86.

Vacek, P.M., 1985. The effect of conditional dependence on the evaluation of diagnostic tests. Biom 41, 959–968.

Van Enst, W.A., Ochodo, E., Scholten, R.J., Hooft, L., Leeflang, M.M., 2014. Investigation of publication bias in meta-analyses of diagnostic test accuracy: a meta-epidemiological study. BMC Med. Res. Methodol. 14, 70.

de Vet, H., Eisinga, A., Riphagen, I., Aertgeerts, B., Pewsner, D., 2008. Chapter 7: Searching for Studies. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy.

Viechtbauer, W., Cheung, M.W., 2010. Outlier and influence diagnostics for meta-analysis. Res. Synth. Methods 1, 112–125.

Wang, D., Bowman, D.D., Brown, H.E., Harrington, L.C., Kaufman, P.E., McKay, T., Nelson, C.T., Sharp, J.L., Lund, R., 2014. Factors influencing U.S. canine heartworm (*Dirofilaria immitis*) prevalence. Parasites Vectors 7, 264.

Waugh, C., Clark, G., 2021. Factors affecting test reproducibility among laboratories. Rev. Sci. Tech. 40, 131–143.

Weil, G.J., Malane, M.S., Powers, K.G., Blair, L.S., 1985. Monoclonal antibodies to parasite antigens found in the serum of *Dirofilaria immitis*-infected dogs. J. Immunol. 134, 1185–1191.

Whiting, P., Rutjes, A.W., Reitsma, J.B., Glas, A.S., Bossuyt, P.M., Kleijnen, J., 2004. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. Ann. Intern. Med. 140, 189–202.

Whiting, P.F., Rutjes, A.W., Westwood, M.E., Mallett, S., 2013. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. J. Clin. Epidemiol. 66, 1093–1104.

Whiting, P.F., Sterne, J.A., Westwood, M.E., Bachmann, L.M., Harbord, R., Egger, M., Deeks, J.J., 2008. Graphical presentation of diagnostic information. BMC Med. Res. Methodol. 8, 20.

Whiting, P.F., Rutjes, A.W.S., Westwood, M.E., Mallett, S., Deeks, J.J., Reitsma, J.B., Leeflang, M.M.G., Sterne, J.A.C., Bossuyt, P.M.M., 2011. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann. Intern. Med. 155, 529–536.

Wilks, C., 2001. Gold standards as fool's gold. Aust. Vet. J. 79, 115.

Zhou, Y., Dendukuri, N., 2014. Statistics for quantifying heterogeneity in univariate and bivariate meta-analyses of binary data: The case of meta-analyses of diagnostic accuracy. Stat. Med 33, 2701–2717.

Zoetis Inc, DiroCHEK® Heartworm Antigen Test Kit.