Diagnostic performance of ocular thermography for estimating body temperature in dogs

Sahar Rostami
Mississippi State University https://orcid.org/0009-0005-1023-0840
Kimberly A. Woodruff
Mississippi State University
Abigail McBride
Mississippi State University
David R. Smith
DSmith@cvm.msstate.edu

Mississippi State University

Research Article

Keywords: Canine, infrared thermography, rectal temperature, diagnostic accuracy, Bayesian latent class analysis

Posted Date: April 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9520207/v1

Additional Declarations: The authors declare no competing interests.

Abstract

Objective.

To estimate the sensitivity (Se) and specificity (Sp) of eye temperature (ET) and rectal temperature (RT) for detecting abnormal temperature in kennel-housed dogs without a gold standard.

Methods.

This observational study was conducted at a single shelter, and included 238 visit-level observations from 121 dogs (89 puppy-visits; 149 adult-visits). Tests were dichotomized using pre-specified reference intervals. A hierarchical Bayesian latent-class model (BLCA) was fit with two subpopulations (puppy and adult), a dog-level random intercept, and a dependence term between tests. Posterior draws were used to estimate predictive values across pre-test probabilities and for combination rules (OR: any positive; AND: both positive). Prior sensitivity analysis was conducted to evaluate the robustness of the results.

Results.

Posterior medians (95% credible interval) were: ET-Se 0.40 (0.10–0.73), ET-Sp 0.94 (0.89–0.99); RT-Se 0.32 (0.09–0.68), RT-Sp 0.99 (0.95–1.00).

Conclusions.

Ocular thermography and rectal thermometry demonstrated similar diagnostic performance. For both tests, given the dog has abnormal core body temperature, the reading is more likely to be wrong. Given the dog has a normal core body temperature, the reading is likely to be correct. Abnormal temperature readings are informative, over wide range of clinical judgement on the likelihood that the dog's core temperature is abnormal. However normal readings are not informative over what was already known based on clinical judgement.

Clinical Relevance.

Interpretation of readings of either test depend on an accurate assessment of the likelihood of abnormal core body temperature based on clinical judgement.

Introduction

Clinical and field studies often must judge new screening tests without the benefit of a perfect reference standard. In such settings, comparing a new assay to an imperfect test can affect accuracy, and mislead decisions based on the test results. Bayesian latent class analysis (BLCA) provides a principled alternative: it treats the true disease state as unobserved and infers test sensitivity (Se), specificity (Sp), and prevalence jointly from the observed data, using prior knowledge to guide estimates. BLCA has been applied in veterinary epidemiology where gold standards are often infeasible, unsafe, or ethically

impractical.[1,2] It has been used to estimate diagnostic test accuracy for a diverse range of pathogens, including parasites, bacteria, and viruses, in various host species, often when a perfect reference test is unavailable.[3–6]

Hui–Walter framework is the simplest form of BLCA, two tests applied in two populations, it yields identifiable estimates when three assumptions hold: (1) the populations differ in prevalence; (2) each test's sensitivity and specificity are constant across populations; and (3) the tests are conditionally independent given the (latent) disease state.[7,8] As later authors emphasize, those assumptions are often inconvenient in practice, especially the independence condition, and even the "two tests, two populations" design sits at the boundary of identifiability where the number of free parameters matches the degrees of freedom in the data.[8] If two tests share biological signal (for example, both share the same pathophysiology), their errors tend to co-vary even after conditioning on disease status. Vacek showed that ignoring such conditional dependence can substantially bias Se and Sp.[9] Subsequent methodological work provide ways to parameterize dependence (e.g., covariance terms within diseased and non-diseased),[9] but also made clear that some dependence models are not identifiable from two tests alone without further structure or prior information.[8] Bayesian modeling allows us to incorporate prior knowledge about accuracy and prevalence, and to pass on that information with the likelihood into posterior inferences.[10,11] In practice, posterior sampling is performed with Markov chain Monte Carlo (MCMC) using flexible softwares such as WinBUGS or JAGS; convergence is assessed by diagnostics applied to multiple chains.[3,12,13]

Previously we validated the thermal camera's temporal stability and determined the reference range for normal eye temperature (ET) readings in canines.[14,15] The objective of this study is to estimate the Se and Sp of ET and rectal temperature (RT).

Material and Methods

All animal procedures were carried out following approval from the Institutional Animal Care and Use Committee at Mississippi State University, with shelter management granting authorization for the enrollment of dogs housed in their facility. The study was conducted at a single animal shelter and to simulate a realistic, low-stress screening scenario, all measurements were performed while dogs remained in their kennels. Individual animals were identified by a shelter ID, which allowed for the linkage of repeated observations across multiple visit days. The dogs in this study didn't have any evidence of ocular disease.

Thermal imaging

Prior to each imaging session, ambient temperature and humidity were recorded using a digital thermometer-hygrometer (INKPET, accuracy ± 1.0°C and ± 5% RH). These environmental values were recorded in the dataset and used to set the environmental parameters on the thermal camera.

Ocular temperatures were measured using a Fotric 347A infrared camera. For all measurements, emissivity was set to 0.98, while standing in the aisle at a distance of approximately 60 cm from the dog. Thermal images were captured from the kennel aisle, with the camera aimed at the dog's eyes. If animal movement prevented a clear still image, a short video was recorded to ensure a usable frame could be obtained.

Rectal thermometry

At the end of the imaging process for a given visit, RT was obtained while the dog remained in the kennel. While the animals were not moved from their kennel environment, adigital probe thermometer was properly lubricated and then inserted to a uniform depth in the rectum. To prevent sudden movement and ensure safety, the animal was held in a natural standing position by a member of the research team. The thermometer remained in place until a final, stable value was indicated by an audible signal.

Image processing

All images and videos were processed in AnalyzIR Mercury. For images, we extracted the maximum ET detected in either eye and used that value for analysis. For videos, the first clear, in-focus frame of an eye was identified, and the maximum temperature from that single frame was recorded.

Bayesian latent class model

We used a hierarchical BLCA to estimate the diagnostic Se and Sp of ET (Test 1) and RT (Test 2) for detecting abnormal body temperature in the absence of a perfect gold standard. The latent outcome was the presence of abnormal body temperature (outside reference interval) at the time of sampling. Tests were recorded as binary outcomes (abnormal temperature = test positive, normal temperature = test negative). A positive ET test was defined as an ET outside the normal range, with hypothermia defined as < 35.0°C and hyperthermia defined as > 38.5°C.[15] A positive RT test was defined as a RT outside the normal range, with hypothermia defined as < 37.7°C and hyperthermia defined as > 39.5°C.[16] Each observation is a dog-visit (a dog measured on a specific sampling day). Some dogs were seen on multiple days; therefore, we modeled outcomes at the visit level and included a dog random intercept to capture within-dog correlation. We used the classic two-test/latent-class framework (Hui–Walter)[7] with two subpopulations (puppy vs adults), and allowed conditional dependence via disease-specific (DevD = deviation from independence in diseased group) and healthy-specific (DevH = deviation from independence in healthy group) covariance terms, as both ET and RT can be influenced by shared physiological pathways.[9,17] Puppies and adults were modeled as two subpopulations following the Hui–Walter framework. Although their historical abnormal body temperature prevalence estimates were similar, age group was retained as a grouping factor to allow for potential biological and measurement-

related differences in temperature regulation, and to permit group-specific prior information on prevalence. This two-population formulation also facilitated the hierarchical structure of the Bayesian model, allowing prevalence and test accuracy parameters to be estimated with appropriate uncertainty for each age group.

Priors

Due to a lack of prior information in the literature for this specific application, weakly informative (Beta (2, 2)) priors were initially assigned to the Se and Sp parameters for both tests. After prior predictive checks, we planned sensitivity runs with alternative weakly informative priors.[18] Group-specific prevalence priors were centered on historical estimates from the same shelter (puppies $\pi_1 \approx 0.125$; adults $\pi_2 \approx 0.134$).

Likelihood

We modeled each shelter visit as a four-level outcome representing the joint ET and RT test pattern: (ET−/RT−), (ET−/RT+), (ET+/RT−), and (ET+/RT+). For visit i, the latent probability of abnormal temperature was modeled on the logit scale as logit(p[i]) = μ[group[i]] + u[dog[i]], where μ[group] is the group-specific baseline prevalence for puppies versus adults, and u[dog] is a dog-level random effect to account for repeated measures and unobserved heterogeneity. Given disease status, the 2×2 cell probabilities were parameterized by test sensitivities and specificities, with conditional dependence incorporated via DevD and DevH states. The observed four-pattern outcome was then linked to the latent disease state using a collapsed-mixture categorical likelihood, in which each pattern probability is a weighted sum of the diseased and non-diseased cell probabilities, with weights p[i] and 1 − p[i].

Computation and diagnostics

Models were implemented in JAGS 4.3.1 via rjags[19], with three chains, 2,000 adaptation iterations, 20,000 burn-in iterations, and 80,000 sampling iterations thinned by 5 (per chain). Convergence was assessed using trace plots, and Gelman–Rubin R-hat (< 1.1 considered acceptable).

We also converted the posterior distributions of Se and Sp from the BLCA into post-test probabilities. For each MCMC draw (by age group), we computed likelihood ratios LR + = Se/(1 − Sp) and LR−=(1 − Se)/Sp, and then obtained positive predictive value (PPV) and negative predictive value (NPV); we then plotted the posterior medians and 95% credible intervals of PPV/NPV against pre-test probability for each test/rule.

Sensitivity analyses

We conducted targeted prior sensitivity by shifting the means of the Beta priors Se shifts for both RT and ET: ±10 and ± 20, percentage points (pp). Sp_{ET} shifts: −10, + 5 pp and Sp_{RT} shifts: -10, + 5 pp. Prevalence shifts: −5, + 10, +20 pp.

Results

In this study 238 visit-level observations from 121 dogs (puppy-visits = 89 from 58 dogs; adult-visits = 149 from 64 dogs; one dog contributed to both categories) were analyzed. Observed test results were: ET + 20/238 (8.4%), RT + 8/238 (3.4%). Joint patterns were: ET−/RT− 210/238 (88.2%), ET−/RT + 8/238 (3.4%), ET+/RT− 20/238 (8.4%), ET+/RT + 0/238 (0%). Among the abnormal ET values, 2/20 (10%) were hypothermia and 18/20 (90%) were hyperthermia. Regarding the abnormal RT values, 5/8 (62.5%) were hypothermia and 3/8 (37.5%) were hyperthermia.

Prior-predictive simulations indicated a mismatch for flat priors Beta (2,2). Therefore, we adopted weakly informative priors for Se ($Se_{ET,RT}$ ~ Beta (3.5, 3.5)) and mildly informative priors for Sp (Sp_{ET} ~ Beta (10.8, 1.2); Sp_{RT} ~ Beta (13.95, 1.05)) which yielded prior predictions covering the observed margins and patterns. Regarding prevalence prior, to avoid an overly informative prior dominating a BLCA and to reflect possible non-exchangeability with the current cohort, we reduced the effective sample size to ≈ 20 after prior predictive checks (π_1 ~ Beta(2.50, 17.50); π_2 ~ Beta(2.68, 17.32)), retaining the same means but widening uncertainty relative to the original stronger priors (π_1 ~ Beta(24.78, 173.47); π_2 ~ Beta(19.82, 127.84)).

Posterior summaries

The hierarchical BLCA was fit at the visit level using a four-category joint-test likelihood with dog-specific random effects on prevalence. Convergence was met for all monitored parameters (R-hat ≈ 1.00; large effective sample sizes), and trace/density plots showed stable mixing.

Regarding ET, posterior median sensitivity was 0.4, while specificity was 0.94. Thus, ET tended to be highly specific but relatively low-sensitivity at the chosen threshold. For RT, posterior median sensitivity was 0.32 and specificity was 0.99. RT was very specific but, like ET, showed modest sensitivity given the mentioned reference interval (Table 1). Using a hierarchical model with a dog-level random effect, we report two prevalence summaries. The population-average prevalence (for a randomly selected dog) was 10.2% in puppies and 10.8% in adults. The corresponding typical-dog prevalence (model parameter: the prevalence if every dog were exactly average), was 8.7% in puppies and 9.3% in adults (Table 2). The slightly higher population averages reflect between-dog heterogeneity in underlying risk (sd ≈ 0.58).

Table 1
Posterior summaries for sensitivity (sn), specificity (sp), prevalence, random effects, and dependence parameters

Parameter	Mean	Median	l95	u95	ESS	rhat
Se1	0.402	0.399	0.1	0.728	26376	1
Sp1	0.942	0.941	0.887	0.996	20895	1
Se2	0.34	0.321	0.0895	0.682	24761	1
Sp2	0.985	0.989	0.953	1	27534	1
pi[1]	0.0873	0.0806	0.0221	0.191	40717	1
pi[2]	0.093	0.0856	0.0224	0.206	26056	1
sd_u	0.58	0.492	0.023	1.64	29208	1
DevD	-0.0687	-0.0639	-0.194	0.0447	37348	1
DevH	0.00266	0.00169	-0.00069	0.0112	47404	1

This table presents posterior summaries from the model, including estimates of sensitivity (Se), specificity (Sp), baseline prevalence (π[1], π[2]) for puppies and adults, dog-level random effect standard deviation (sd_u), and deviation-from-independence terms (DevD for disease-positive, DevH for disease-negative). For each parameter, the mean, median, 95% credible interval (l95, u95), effective sample size (ESS), and R-hat ($\hat{R}$) convergence statistic are reported. These values provide a summary of model estimates and their associated uncertainty.

Table 2
Prevalence summaries by age group under the hierarchical latent-class model

Estimand	Group	Mean	l95	u95
baseline_u_0	Puppy	0.0873	0.0221	0.191
baseline_u_0	Adult	0.093	0.0224	0.206
new_dog	Puppy	0.102	0.0271	0.219
new_dog	Adult	0.108	0.0268	0.234

This table presents posterior prevalence estimates for puppies and adult dogs under the hierarchical latent-class model. The baseline_u_0 values represent the prevalence for a “typical” dog within each group (π_g), accounting for the dog-level random effect. The new_dog values represent the population-average prevalence, integrating over the between-dog variability. For each estimate, the posterior mean and 95% credible interval (l95, u95) are reported.

Post-test probabilities

The posterior probability of abnormal temperature in this population given each test pattern was: both tests negative 0.03; only RT+, 0.8; only ET + 0.4, and both positive, 0.7 (Tables 3 and 4). Across the full range of pre-test probabilities, rectal testing and the AND (both+) rule provided the strongest rule-in

evidence (higher PPV), with uncertainty bands reflecting posterior credible intervals. NPV changed little as the pre-test probability changed, which means a negative result adds little extra rule-out value at the observed baseline risk (Fig. 1). Likelihood-ratio analysis showed Rectal provided the strongest rule-in evidence (largest median LR+), with AND (both +) second and OR/Eye providing only moderate LR+. For ruling out, OR (any +) had the smallest LR−, rectal and eye were intermediate, and AND did not meaningfully reduce LR− (median near 1) (Fig. 2).

Table 3
Posterior probability of abnormal body temperature by test pattern.

Pattern description	Group	Mean	Median	l95	u95
ET -, RT -	Puppy	0.0405	0.0298	0.00222	0.141
ET -, RT -	Adult	0.0414	0.031	0.00229	0.141
ET -, RT -	overall_visit	0.0409	0.032	0.00266	0.13
ET -, RT +	Puppy	0.713	0.811	0.0926	0.997
ET -, RT +	Adult	0.723	0.805	0.135	0.997
ET -, RT +	overall_visit	0.728	0.811	0.141	0.997
ET +, RT -	Puppy	0.422	0.374	0.0391	0.978
ET +, RT -	Adult	0.431	0.383	0.0274	0.984
ET +, RT -	overall_visit	0.433	0.384	0.0372	0.982
ET +, RT +	Puppy	0.616	0.66	0.0772	0.986
ET +, RT +	Adult	0.623	0.665	0.0894	0.985
ET +, RT +	overall_visit	0.629	0.672	0.0989	0.985

This table presents the posterior probability of abnormal body temperature for each combination of ET (eye temperature) and RT (rectal temperature) results under the hierarchical Bayesian latent-class model. Probabilities are reported separately for puppies, adults, and pooled across all visits (overall_visit). For each group and test pattern, the posterior mean, median, and 95% credible interval (l95, u95) are shown. Values indicate that an RT+ result alone produces the largest increase in post-test probability, whereas two negative test results (ET−, RT−) maintain a low probability of abnormal temperature.

Table 4
Positive and negative predictive values for simple decision rules at observed prevalence.

Rule	PPV_median	PPV_l95	PPV_u95	NPV_median	NPV_l95	NPV_u95
OR (any positive)	0.509	0.189	0.899	0.968	0.87	0.997
AND (both positive)	0.672	0.0989	0.985	0.907	0.802	0.966

This table shows the posterior median and 95% credible intervals (CrI) of positive predictive value (PPV) and negative predictive value (NPV) for two simple decision rules combining ET and RT results: OR (any positive) and AND (both positive). At the observed prevalence, the OR rule maximizes NPV, indicating that a negative result on both tests reliably rules out abnormal temperature, whereas the AND rule maximizes PPV, indicating that both tests positive strongly suggest abnormal temperature.

Sensitivity analyses:

We evaluated 15 prior scenarios targeted shifts in test Se/Sp and group prevalence, under the same dependence prior as the baseline. The planned + 5 pp shift for Sp_{RT} resulted in computational instability. Therefore, we used the closest feasible shift of + 4 pp. This minor adjustment does not materially affect posterior estimates or substantive conclusions and ensured model convergence. Prior sensitivity was summarized by identifying, (for each core parameter = Se_{ET}, Sp_{ET}, Se_{RT}, Sp_{RT}, π_{puppy}, π_{adult}), the three prior scenarios that produced the largest absolute shift in the posterior median relative to the baseline fit. Results are shown in table S1 and report the scenario, the posterior median (95% CrI), the baseline median, and the signed difference (Δ). Large shifts were confined to parameters whose priors were changed directly and to prevalence. Se_{ET} moved by − 0.225 and + 0.188 under ± 20-pp Se_{ET} prior shifts, with a smaller cross-effect from Prevalence + 20pp (− 0.124). Se_{RT} increased by + 0.197 with Se_{RT} +20 point and decreased by − 0.161 with Se_{RT} -20 pp, while Prevalence + 20 pp pulled its median down (− 0.181). Specificities were robust: Sp_{ET} changed by at most + 0.030 and Sp_{RT} within − 0.019 to + 0.008 across all scenarios. Prevalence medians responded as expected to direct prevalence prior shifts (e.g., Prevalence + 20 point raised π_1 and π_2 by + 0.155 and + 0.171, respectively).

Discussion

Beyond laboratory tests, non-invasive tools such as IRT can complement decision-making by monitoring health status in dogs: for example, surface temperatures rise more after defense than obedience exercises and remain elevated in distal limbs, patterns consistent with thermoregulation after exertion.[20] Recent field work also links ocular and nasal thermal metrics to physiologic indicators, underscoring IRT's potential as an adjunct for welfare and performance monitoring.[21]

In veterinary epidemiology, diagnostic tests add value only when they change what we do for patients. Misuse and overuse of tests remain common pitfalls when Sn, Sp, and predictive values are considered in isolation or without regard to case mix and prevalence, which can inflate false positives or false

negatives and ultimately affect decision quality.[22,23] As McKenzie argues, clinicians must pair knowledge of test characteristics with the clinical context and explicitly think in terms of Bayesian updating, linking Se, Sp, likelihood ratios, prevalence, and predictive values, to avoid both misdiagnosis and overdiagnosis. Positive and negative likelihood ratios are particularly useful because they quantify how much a result shifts disease probability; by contrast, predictive values depend on the underlying prevalence and can mislead when pre-test probability is very low or very high.[22]

Taken together, rational testing in veterinary epidemiology requires aligning test choice with pre-test probability, using likelihood ratios to update that probability, and integrating emerging physiologic measures when they provide actionable, patient-centered information.

In this study, our aim was to estimate accuracy of ET/RT without a gold standard in shelter dogs using a hierarchical BLCA in JAGS.[13] The BLCA with two subpopulations (puppies vs adults) and dog-level random effects yielded convergent, well-mixed posterior distributions and a consistent qualitative picture: both tests were highly specific, whereas sensitivities were modest with wide uncertainty.

Kennel-side screening minimizes handling time and stress while enabling frequent checks. Although ET is appealing as a fully non-contact screen, its posterior sensitivity was low with wide intervals. That is consistent with the broader veterinary thermography literature, which highlights substantial biological and environmental variability and recommends cautious clinical integration of infrared measures as adjuncts rather than replacements for core vital signs.[24,25]

We adopted the classical Hui–Walter two-test/two-population framework to estimate test accuracy without a gold standard[7] and allowed for conditional dependence between ET and RT in the diseased and healthy strata.[9,17] Conditional dependence is epidemiologically plausible because both measures can respond to shared physiologic processes. When this dependence is ignored, sensitivity or specificity can be biased.[8,9]

Although the Hui–Walter model assumes that the compared populations differ in disease prevalence, the observed prevalence of abnormal temperature among puppies and adults were similar in both the historical and current datasets. This limits strict identifiability under the original Hui–Walter framework. However, the use of informative priors and a hierarchical Bayesian formulation allowed the model to borrow strength across levels and generate coherent posterior estimates. The close similarity in prevalence also suggests that, within this shelter population, age group might not be a strong determinant of temperature abnormalities, an observation that might have biological meaning as well as methodological implications.

We estimated post-test probabilities assuming both tests are done on every visit. In a serial testing ET-first workflow, RT would only follow ET+, so ET−/RT+ would not occur and its probability is not relevant for that pathway. Because ET was only moderately sensitive, an ET-first strategy could miss some RT+ dogs after an ET− screen. Our model also suggested slight negative dependence among positives (DevD

< 0), making concurrent-positive results less common. As a result, in the parallel analysis ET−/RT+ carried a higher post-test probability than ET+/RT+. These statements are based on our chosen thresholds, other cut points would shift Se and Sp.

Abnormal temperatures were rare in this cohort, with only 20 ET positives and 8 RT positives. ET detected a larger number of abnormalities than RT, and the nature of abnormalities differed: ET largely captured hyperthermia, while RT largely captured hypothermia. We speculate on several hypotheses that might explain these patterns. First, ET might be inherently more sensitive to transient or peripheral temperature changes, as the eye surface might respond more rapidly to environmental and physiological stressors, whereas RT reflects core body temperature and might be more stable. Second, hyperthermic events might manifest more readily at the periphery than hypothermic events, resulting in ET detecting more hyperthermia than RT. Third, measurement factors might play a role: ET is non-invasive and faster, potentially allowing earlier detection of elevated temperatures, whereas RT measurement might be influenced by probe placement, rectal contents, or stress-related vasoconstriction that dampens extremes. These biological and technical considerations suggest that ET and RT might provide complementary information, but also highlight limitations in estimating Sn for rare outcomes, particularly for RT. In our BLCA, RT test performance (high Sp, modest Se) and posterior positive and negative predictive values were similar to ET. So as with ET, a single negative RT should not be over-interpreted. Because test results with abnormal temperature were rare, we have limited information to estimate Se, so the credible intervals around our Se estimates were wide and should be interpreted with caution.[8] Across the range of pre-test probabilities, the NPV curve was nearly linear, indicating that a negative result provided little additional information beyond the clinician prior assessment of the likelihood of abnormal temperature. However, positive results are likely correct and more informative when the clinician has suspicions that the dog has abnormal temperature.

In summary, a hierarchical BLCA with careful prior-predictive checks and dependence modeling provides a credible, estimate of ET and RT diagnostic accuracy in a real-world condition. This study was conducted in a single shelter, which might limit the generalizability of the findings. Expanding the study to multiple settings with varying populations and environmental conditions could provide more robust estimates of test performance and allow better evaluation of age and site-specific differences in temperature abnormalities. Such studies would also allow hierarchical BLCA models to assess test accuracy across a wider range of populations, potentially improving the precision and generalizability of the estimates.

Declarations

Disclosures

The authors have nothing to disclose. No AI-assisted technologies were used in the generation of this manuscript.

Funding

This work was supported by funding from the United States Department of Agriculture Agricultural Research Service (USDA ARS) grant no. 58-6064-3-017. The opinions, findings, and conclusions expressed in this study are those of the authors and do not necessarily reflect the views of the USDA.

Acknowledgements

The authors would like to thank Dr. Jingyi Shi, Dr. W. Isaac Jumper and Dr. Guiming Wang for their valuable guidance and service as members of the doctoral committee during the development of this research. We thank Mississippi State University Population Medicine Program for generously providing the historical data from the animal shelter.

References

1. Gelman A, Hill J (2007) Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press
2. Goncalves L, Subtil A, Brites N, Oliveira MR, Margarida A, Meireles J Bayesian Latent Class Models in Veterinary and Human Epidemiology. *In 46TH SCIENTIFIC MEETING OF THE ITALIAN STATISTICAL SOCIETY*. Published online 2012
3. Branscum AJ, Gardner IA, Johnson WO (2005) Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. Prev Vet Med 68(2):145–163. 10.1016/j.prevetmed.2004.12.005
4. Bronsvoort BM, deC, Toft N, Bergmann IE et al (2006) Evaluation of three 3ABC ELISAs for foot-and-mouth disease non-structural antibodies using latent class analysis. BMC Vet Res 2(1):30. 10.1186/1746-6148-2-30
5. Benito A, Carmena D, Joseph L, Martínez J, Guisantes JA (2006) Dog echinococcosis in northern Spain: Comparison of coproantigen and serum antibody assays with coprological exam. Vet Parasitol 142(1):102–111. 10.1016/j.vetpar.2006.06.011
6. Carabin H, Balolong E, Joseph L et al (2005) Estimating sensitivity and specificity of a faecal examination method for *Schistosoma japonicum* infection in cats, dogs, water buffaloes, pigs, and rats in Western Samar and Sorsogon Provinces, The Philippines. Int J Parasitol 35(14):1517–1524. 10.1016/j.ijpara.2005.06.010
7. Hui SL, Walter SD (1980) Estimating the Error Rates of Diagnostic Tests. Biometrics 36(1):167–171. 10.2307/2530508
8. Toft N, Jørgensen E, Højsgaard S (2005) Diagnosing diagnostic tests: evaluating the assumptions underlying the estimation of sensitivity and specificity in the absence of a gold standard. Prev Vet Med 68(1):19–33. 10.1016/j.prevetmed.2005.01.006

9. Vacek PM (1985) The Effect of Conditional Dependence on the Evaluation of Diagnostic Tests. Biometrics 41(4):959–968. 10.2307/2530967
10. Gelman A (2004) Parameterization and Bayesian Modeling. J Am Stat Assoc 99(466):537–545. 10.1198/016214504000000458
11. Li Y, Lord-Bessen J, Shiyko M, Loeb R (2018) Bayesian Latent Class Analysis Tutorial. Multivar Behav Res 53(3):430–451. 10.1080/00273171.2018.1428892
12. Lunn DJ, Thomas A, Best N, Spiegelhalter D (2000) WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. Stat Comput 10(4):325–337. 10.1023/A:1008929526011
13. Plummer M JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling. Published online 2003
14. Rostami S, Woodruff KA, McBride A, Shi J, Smith DR (2026) Ocular temperature measured by infrared thermography in shelter dogs varies with biological and environmental factors. Zenodo Preprint posted online March 2. 10.5281/zenodo.18842079
15. Rostami S, Woodruff KA, McBride A, Smith DR (2026) Ocular temperature measured by infrared thermography in shelter dogs varies with biological and environmental factors. Res Square Preprint posted online April 14. 10.21203/rs.3.rs-9372827/v1
16. Dorn E, Bogedale K, Pankraz A, Neiger R (2024) Calculation of a Reference Interval for Rectal Temperature in Adult Dogs Presenting for Veterinary Care Using an Algorithm for Mixed Data. Anim (Basel) 14(13):1970. 10.3390/ani14131970
17. Dendukuri N, Joseph L (2001) Bayesian Approaches to Modeling the Conditional Dependence between Multiple Diagnostic Tests. Biometrics 57(1):158–167
18. Wesner JS, Pomeranz JPF (2021) Choosing priors in Bayesian ecological models by simulating from the prior predictive distribution. Ecosphere 12(9):1–11. 10.1002/ecs2.3739
19. Plummer M, Stukalov A, Denwood M (2025) rjags: Bayesian Graphical Models using MCMC. Published online March 24, 2025. Accessed October 14. https://cran.r-project.org/web/packages/rjags/index.html
20. Pichová K, Pavlin S, Košťál Ľ, Pintarič Š, Zupan Šemrov M (2023) Thermography as a tool to assess training effects in military working dogs. J Therm Biol 112:103441. 10.1016/j.jtherbio.2022.103441
21. Budny-Walczak A, Wilk M, Kupczyński R (2024) Infrared Thermography in Assessment of Facial Temperature of Racing Sighthound-Type Dogs in Different Environmental Conditions. Animals 14(8):1180. 10.3390/ani14081180
22. McKenzie BA (2021) Rational use of diagnostic and screening tests. J Small Anim Pract 62(11):1016–1021. 10.1111/jsap.13393
23. Smith DR (2018) I'm not positive that's a positive: Become indispensable to your clients by knowing how to choose, interpret and incorporate diagnostic testing in bovine practice. *American Association of Bovine Practitioners Conference Proceedings*. Published online September 13, :126–133. 10.21423/aabppro20183123

24. Elias B, Starling M, Wilson B, McGreevy P (2021) Influences on Infrared Thermography of the Canine Eye in Relation to the Stress and Arousal of Racing Greyhounds. Animals 11:103. 10.3390/ani11010103
25. Ring EFJ, Ammer K (2012) Infrared thermal imaging in medicine. Physiol Meas 33:R33–R46. 10.1088/0967-3334/33/3/R33

Figures

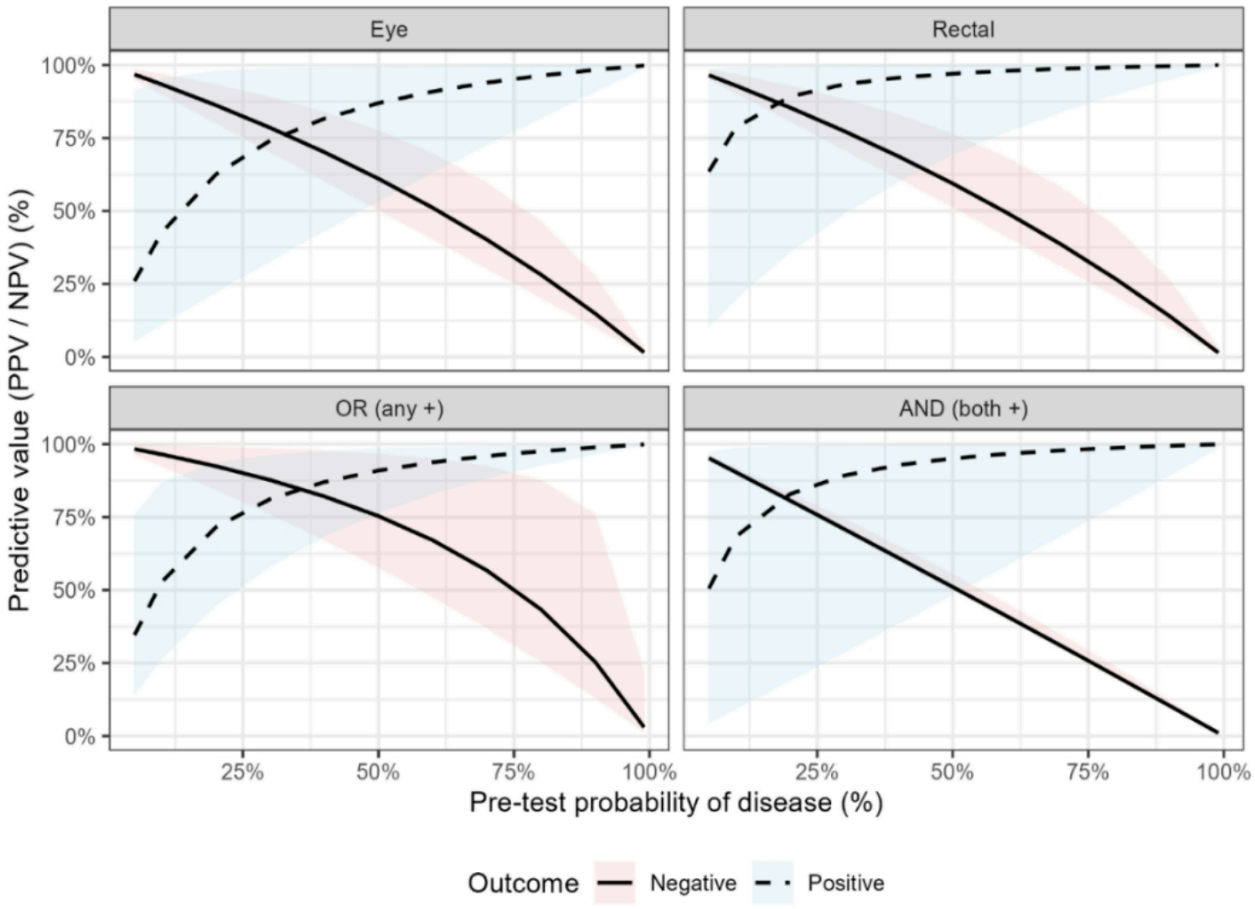

Figure 1

Posterior predictive values by pre-test probability for individual and combined tests. This figure shows the relationship between pre-test probability and posterior predictive values for detecting abnormal body temperature under the hierarchical Bayesian latent-class model. Individual tests include eye temperature (ET) and rectal temperature (RT), and combination rules include OR (any positive) and AND (both positive). Dashed lines represent the positive predictive value (PPV) after a positive test, solid lines

represent the negative predictive value (NPV) after a negative test, and shaded areas depict the 95% credible intervals. This visualization allows assessment of how test performance varies across different assumed pre-test probabilities.

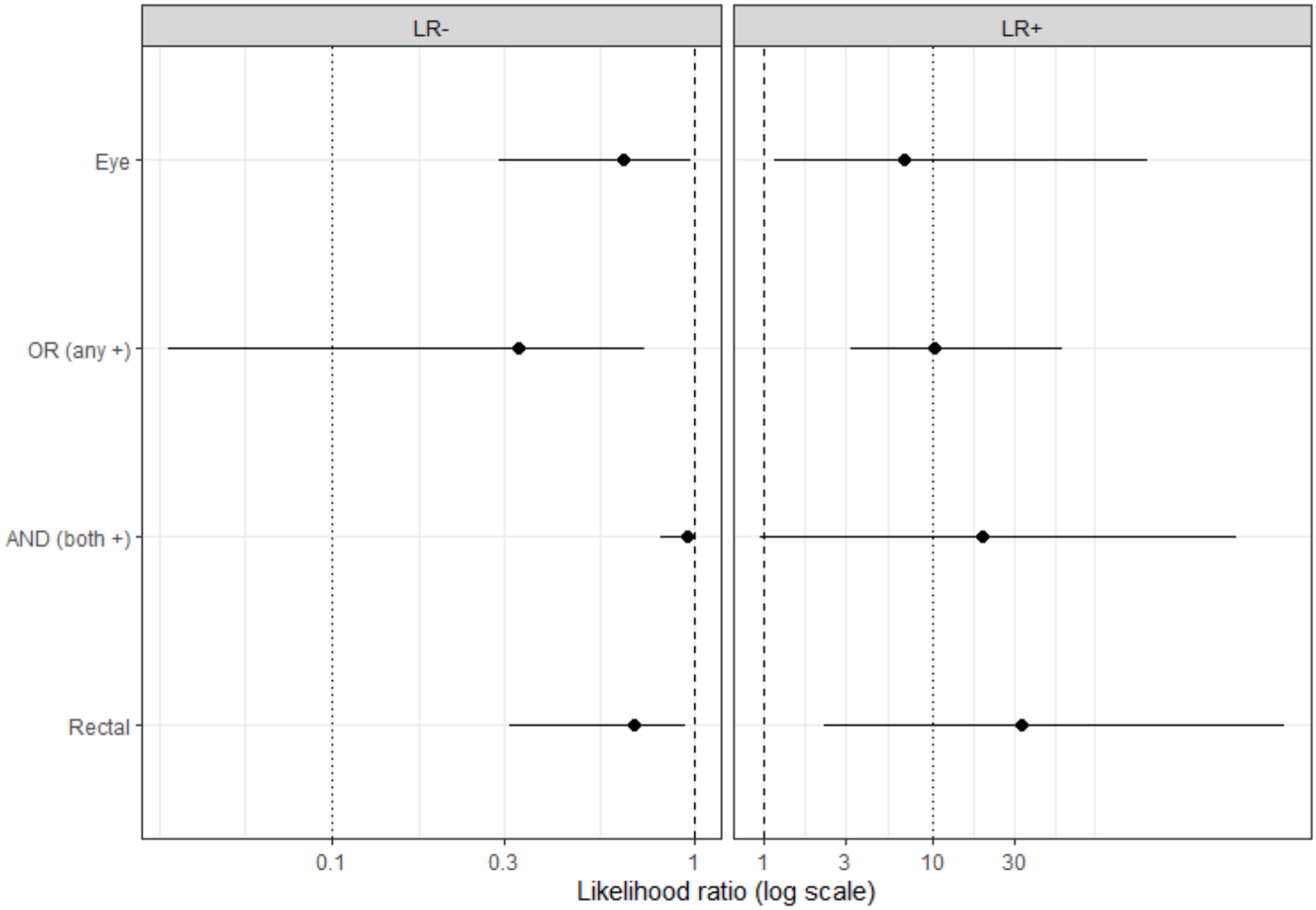

Figure 2

Posterior likelihood ratios for individual tests and combination rules. This figure presents likelihood ratios for detecting abnormal body temperature under the hierarchical Bayesian latent-class model. Points show posterior medians, and horizontal bars represent 95% credible intervals for negative likelihood ratios (LR−, left panel) and positive likelihood ratios (LR+, right panel) on a logarithmic scale. Vertical dashed lines indicate LR = 1, which represents no diagnostic information, while dotted reference lines correspond to conventional thresholds for strong evidence (LR− ≈ 0.1, LR+ ≈ 10). This visualization highlights the comparative strength of individual tests (Eye Temperature, Rectal Temperature) and decision rules (OR, AND).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.docx