



OPEN ACCESS

Key Words

bovine brucellosis, serum agglutination test, milk antibodies, Cohen's kappa, secondary data analysis

Corresponding Author

Ufuk Ülker,
Department of Preclinical Sciences,
Division of Veterinary Microbiology,
Faculty of Veterinary Medicine,
Yozgat Bozok University, Yozgat,
Türkiye

Author Designation

Received: 06 January 2026

Accepted: 20 May 2026

Published: 12 June 2026

Citation: Ufuk Ülker and Sumru ÇITAK., 2026 . Diagnostic Threshold Vulner ability and Matrix Discordance in Bovine Brucellosis Serology: An Advanced Evaluation of Granular Dilution End points and Reaction Grades. Res. J. Biol. Sci, 21:8-15,doi:10.36478/makrjbs.2026.8.15

Copy Right: MAK HILL Publication

Diagnostic Threshold Vulner ability and Matrix Discordance in Bovine Brucellosis Serology: An Advanced Evaluation of Granular Dilution End points and Reaction Grades

¹Ufuk Ülker and ²Sumru ÇITAK

¹Department of Preclinical Sciences, Division of Veterinary Microbiology, Faculty of Veterinary Medicine, Yozgat Bozok University, Yozgat, Türkiye

²Department of Biology, Division of General Biology, Faculty of Science, Gazi University, Ankara, Türkiye

ABSTRACT

Conventional binary (positive/negative) diagnostic thresholds in serological surveillance frequently overlook sub-threshold antibody reactivity, creating analytical blind spots. This retrospective study evaluated the relationship among serum agglutination test (SAT) antibody titers, visual reaction intensity grades (+1 to +4), and paired serum–milk direct enzyme-linked immunosorbent assay (I-ELISA) outcomes, utilizing archived laboratory records from 270 unvaccinated cattle investigated following abortion or clinical suspicion of brucellosis in Türkiye. Following the exclusion of ambiguous agglutination records (n = 4) and an indeterminate milk ELISA result (n = 1), 266 animals were available for titer stratification and 269 for paired matrix evaluation. Markedly high immunoassay reactivity was detected below the conventional 1:40 diagnostic cutoff: among 102 SAT-negative cattle, I-ELISA positivity was identified in 41.2% (42/102) of serum and 64.7% (66/102) of milk samples. Within the low-titer stratum (1:10–1:20), positivity reached 66.1% (37/56) in serum and 78.6% (44/56) in milk. Reaction grades within low titers demonstrated no monotonic correlation with IgG-specific ELISA positivity, reflecting functional divergence between pentameric IgM-driven lattice formation and monomeric IgG affinity. Paired matrix evaluation revealed significant directional asymmetry: 42 cattle were milk-positive yet serum-negative, compared with only six displaying the reciprocal profile (exact McNemar $p = 1.01 \times 10^{-6}$; observed agreement 82.2% [95% CI: 77.1–86.3%]; Cohen's kappa $\kappa = 0.534$ [95% CI: 0.424–0.644]). Lowering the descriptive SAT threshold from $\geq 1:40$ to $\geq 1:10$ reclassified 37 immunoassay-reactive cases, moderately increasing crude agreement from 68.8% to 75.6% (95% CI: 70.1–80.3%) and chance-adjusted concordance ($\kappa = 0.417$ to $\kappa = 0.464$ [95% CI: 0.354–0.574]); however, it concomitantly increased SAT-positive/ELISA-negative discordance from 4 to 23 animals (1.5% to 8.6%). In the absence of definitive bacteriological isolation, asymmetric milk reactivity warrants cautious interpretation, representing either localized intramammary immunoglobulin secretion or analytical matrix interference. These findings suggest that reporting discrete titration endpoints alongside reaction grades and paired matrix screening may provide a valuable methodological framework for refining bovine brucellosis surveillance and control protocols.

INTRODUCTION

Bovine brucellosis affects animal reproduction and public health. Serological testing is central to surveillance, but interpretation depends on the assay, its decision rule, and the epidemiological setting. A reactive antibody result alone does not establish viable infection or active shedding. The World Organisation for Animal Health describes the complementary roles and limitations of available tests in diagnostic and surveillance programmes^[1].

Serum agglutination testing (SAT), the Rose Bengal plate test (RBPT), complement fixation testing (CFT), and indirect enzyme-linked immunosorbent assays (I-ELISAs) interrogate different properties of the antibody response. Agglutination depends on antibody-mediated particle interactions, whereas the historical I-ELISAs used an anti-bovine IgG conjugate. Differences between their classifications can therefore arise without establishing which result correctly identifies infection^[2]. Retaining recorded dilution endpoints and reaction grades permits examination of information lost when results are reduced to a single binary classification.

The distinction between antibody detection and independently established infection is fundamental to interpreting these comparisons^[3]. Reviewed direct and indirect diagnostic approaches in livestock and emphasized the need to interpret serological findings in relation to the stage and context of infection^[4]. Compared published performance and cost data across bovine serological assays^[5]. Systematically reviewed immunological tests for cattle brucellosis and showed why interpretation must account for the epidemiological setting, including vaccination and false-positive serological reactions. More recently^[6], found that study design influenced pooled estimates of diagnostic sensitivity and specificity, with higher estimates when case-control studies were included. These findings motivate a deliberately descriptive analysis here: endpoint and grade information is examined without assigning an unverified disease status to each animal.

An archived SAT entry contains both a dilution endpoint and a visual reaction grade (+1 to +4). These are distinct laboratory observations: endpoint describes the recorded dilution, while grade describes reaction intensity at that dilution. Neither alone measures antibody affinity, avidity, or isotype composition. Examining them separately can identify heterogeneous profiles within low-dilution strata and assess how operational classification rules affect agreement with another assay.

Serum and milk also require separate interpretation because sample composition and assay decision limits differ^[7]. Evaluated serum and milk

I-ELISAs using matrix-specific cutoffs and defined herd populations. Paired serum–milk classification can reveal directional disagreement, but cannot by itself distinguish differences in antibody availability from threshold effects or nonspecific analytical signal.

The contemporary literature also identifies continuing variation in diagnostic practice^[8]. Reviewed 349 livestock studies and documented differences in test selection and adherence to diagnostic guidance. Their findings reinforce the need to specify what a recorded laboratory category represents. In the present archive, an SAT endpoint, a visual reaction grade, and an I-ELISA category are distinct recorded variables; agreement among them addresses classification consistency rather than confirmed infection. The primary objectives of this study were to characterize immuno assay reactivity patterns across discrete SAT dilution endpoints, evaluate the distribution of visual reaction intensity grades within borderline titers, and model categorical reclassification dynamics under alternative operational cutoffs. Sensitivity analyses were performed to assess the robustness of alternative classification thresholds.

Animals and diagnostic sampling: A total of 270 non-vaccinated cattle investigated for clinical suspicion of brucellosis or following abortion episodes between February 2007 and April 2008 were evaluated in this study. The diagnostic specimens originated from field submissions across seven provinces in Central and Northern Türkiye: Kastamonu, Çankırı (including Çerkeş), Çorum (including Alaca and Uğurludağ), Kırşehir (including Boztepe), Bolu (Seben), Kırıkkale (Keskin), and Ankara (Çubuk). Paired blood serum and milk samples were collected from each animal under routine veterinary diagnostic screening procedures and preserved at -20 °C until laboratory testing.

Diagnostic assays and interpretation criteria: The Rose Bengal plate test (RBPT) and milk ring test (MRT) results were evaluated qualitatively as positive or negative according to standard procedures. For the serum agglutination test (SAT), results were determined across two analytical dimensions: the terminal dilution end point (ranging from 1:10 to 1:320) and the visual tube agglutination grade (+1 to +4), scored based on the degree of supernatant clearing and sediment aggregation. Samples demonstrating no observable agglutination at the initial 1:10 dilution were categorized as non-reactive (<1:10).

Indirect ELISA (I-ELISA) testing of serum and milk samples utilized an anti-bovine IgG conjugate, with positivity determined based on percent positivity (PP) relative to the positive control. Established diagnostic

cut-offs of PP >25% for serum and PP >10% for milk were applied to categorize samples as positive or negative. Complement fixation testing (CFT) was evaluated based on serum dilution and hemolytic inhibition. Reactions demonstrating at least 50% fixation (+2 grade) at a dilution of 1:10 or higher were classified as serore active. One atypical titration recording (1:180) outside these rial two fold dilution series was excluded from quantitative CFT end point analyses.

Statistical analysis: Inter-assay agreement was evaluated using paired 2 × 2 contingency tables. Overall concordance was determined with 95% Wilson score confidence intervals (CIs). Chance-corrected agreement was calculated using Cohen's kappa (κ) coefficient with standard errors and 95% CIs derived via the multinomial delta method. Four SAT entries with atypical fraction recordings, one CFT entry outside the two fold serial dilution series (1:180), and one indeterminate milk I-ELISA record were omitted from respective quantitative models, leaving evaluable data sets of 266 for SAT and 269 for CFT and milk ELISA. All statistical analyses were performed using IBM SPSS Statistics for Windows (version 26.0; IBM Corp., Armonk, NY, USA).

RESULTS AND DISCUSSION

Overall assay reactivity across the evaluated cattle: Overall serological and milk reactivity across the 270 evaluated cattle was 68.5% (185/270) for serum I-ELISA, 61.1% (165/270) for RBPT, and 45.9% (124/270) for MRT. Following the omission of records with uninterpretable titration profiles, reactivity reached 81.8% (220/269) for milk I-ELISA and 43.9% (118/269) for CFT. Among the 266 interpretable SAT records, 102 (38.3%) showed no recorded reaction (<1:10), 56 (21.1%) had endpoints of 1:10–1:20, 22 (8.3%) had an endpoint of 1:40, and 86 (32.3%) had endpoints of at least 1:80 (Table 1).

Immuno assay reactivity within SAT dilution strata: Among 102 animals with no recorded SAT reaction (<1:10), serum I-ELISA was reactive in 42 (41.2%) and milk I-ELISA in 66 (64.7%) (Table 1). RBPT and MRT were reactive in 16 (15.7%) and 12 (11.8%), respectively. Thus, absence of observable tube agglutination did not correspond to uniformly non reactive results across the other assays.

In the borderline 1:10–1:20 stratum, 37/56 animals (66.1%) were serum I-ELISA reactive and 44/56 (78.6%) were milk I-ELISA reactive (Table 1). RBPT reactivity occurred in 41/56 (73.2%), CFT reactivity in 9/56 (16.1%), and MRT reactivity in 20/56 (35.7%). Consequently, low dilution agglutination was neither

uniformly supported nor uniformly contradicted by complementary assays.

All 22 animals at the 1:40 end point were reactive in both I-ELISAs. Among animals with end points $\geq$ 1:80, serum I-ELISA was reactive in 82/86 (95.3%) and milk I-ELISA in all 85 evaluable records. The four serum-non reactive animals in this stratum were 101 and 184 (both 1:80, +4), 216 (1:160, +4), and 232 (1:320, +4). Thus, a high recorded SAT end point corresponded closely with immuno assay reactivity without making the classifications interchangeable.

Previously published serum and milk findings: Paired serum–milk immuno assay concordance Paired evaluation between serum and milk I-ELISAs demonstrated marked directional discordance. Among the 269 evaluable paired samples, 42 cattle exhibited a milk-positive yet serum-negative profile, whereas only six animals displayed the reciprocal pattern (exact Mc Nemar test, $p = 1.01 \times 10^{-5}$; observed overall agreement: 82.2% [95% CI: 77.1–86.3%]; Cohen's $\kappa = 0.534$ [95% CI: 0.424–0.644]).

Reclassification under alternative SAT rules: Using the conventional dilution-only cut off ($\geq$ 1:40) as there ference baseline across the 266 interpretable SAT records, overall concordance with serum I-ELISA was 68.8% (95% CI: 63.0–74.1%) with moderate agreement ($\kappa = 0.417$ [95% CI: 0.328–0.506]). Reclassification dynamics and agreement trade-offs resulting from operational threshold adjustments are summarized in Table 2.

Lowering the SAT diagnostic cut off from $\geq$ 1:40 to $\geq$ 1:10 shifted 56 animals from non reactive to reactive status. Among these reclassified animals, 37 were serum I-ELISA-reactive and 19 were non reactive. Consequently, 37 discordant pairs transitioned to concordance while 19 previously concordant pairs becamed iscordant, yielding a net agreement gain of 18 pairs (+6.8 percentage points; Table 2). Overall concordance under the $\geq$ 1:10 threshold increased to 75.6% (95% CI: 70.1–80.3%) with a chance-adjusted agreement of $\kappa = 0.464$ (95% CI: 0.354–0.574).

Conversely, implementing a grade-sensitive rule (requiring a visual reaction grade of $\geq$ +2 strictly at the 1:40 dilution while maintaining reactivity or ig hertiters) reclassified two animals from reactive to non reactive. Because both animals were serum I-ELISA-reactive, two previously concordant pairs became discordant with no off setting gains (net agreement change: –2 pairs, –0.8 percentage points; Table 2). Agreement under this grade-sensitive thres hold was 68.0% (95% CI: 62.2–73.4%) with $\kappa = 0.406$ (95% CI: 0.318–0.494).

Low dilution reaction grades and individual profiles: Table 3 separates recorded visual grades within each

Table 1: Serological and immuno assay reactivity profiles stratified across serum agglutination test (SAT) dilution end points in 266 clinically suspect cattle.

SAT End point Stratum	n	Serum I-ELISA	Milk I-ELISA	RBPT	CFT	MRT
Non reactive (<1:10)	102	42/102 (41.2%)	66/102 (64.7%)	16/102 (15.7%)	7/102 (6.9%)	12/102 (11.8%)
Low endpoints (1:10–1:20)	56	37/56 (66.1%)	44/56 (78.6%)	41/56 (73.2%)	9/56 (16.1%)	20/56 (35.7%)
Intermediate (1:40)	22	22/22 (100.0%)	22/22 (100.0%)	21/22 (95.5%)	19/22 (86.4%)	15/22 (68.2%)
High titer ($\geq$ 1:80)	86	82/86 (95.3%)	85/85 (100.0%)	83/86 (96.5%)	81/85 (95.3%)	74/86 (86.0%)

Values indicate reactive/evaluable cattle (%). From 270 records, ambiguous SAT entries (n = 4; animals 86, 89, 136, 142) were excluded from dilution strata. Animals 242 (milk I-ELISA) and 96 (CFT) were excluded due to indeterminate notations. CFT positivity was defined as dilution $\geq$ 1:10 with fixation $\geq$ +2. Source I-ELISA cutoffs were retained (PP > 25% for serum, PP > 10% for milk). Proportions characterize this clinically suspect animal population and do not represent diagnostic accuracy or disease prevalence. Abbreviations: CFT, complement fixation test; I-ELISA, indirect ELISA; MRT, milk ring test; PP, percent positivity; RBPT, Rose Bengal plate test; SAT, serum agglutination test.

Table 2: Re classification dynamics and net agreement trade-off under dilution- and grade-modified SAT interpretation rules against serum I-ELISA.

Change in SAT rule	Changed animals	Serum reactive	Serum non reactive	Agreement gained	Agreement lost	Net agreement change
Dilution $\geq$ 1:40 to $\geq$ 1:10	56	37	19	37	19	+18/266 (+6.8 pp)
Dilution $\geq$ 1:40 to grade-sensitive rule	2	2	0	0	2	-2/266 (-0.8 pp)

*All transition calculations and concordance metrics are established within the 266 fully interpretable SAT records, using the conventional dilution-only threshold ($\geq$ 1:40) as there ference baseline. Changed animals denotes the subset of cattlewhose categorical SAT status (reactive vs. non reactive) shifts under the alternative rule. Serum-reactive and Serum-non reactive counts refer exclusively to the source serum I-ELISA status of these reclassified animals. Agreement gained reflects discordant-to-concordant transitions, whereas Agreement lost denotes concordant-to-discordant shifts relative to serum I-ELISA. Net agreement change represents the difference (gained minuslost) expressed as a percentage of the total evaluable animal population (n = 266). The grade-sensitive rule classifies any end point > 1:40 as reactive and requires a visual reaction grade of $\geq$ +2 strictly at the 1:40 dilution. Rules evaluated here in represent exploratory classification models rather than clinically or regulatory validated diagnostic thresholds. Abbreviations: CI, confidence interval; I-ELISA, indirect enzyme-linked immunosorbent assay; pp, percent age points; SAT, serum agglutination test.

Table 3: Distribution of paired serum and milk I-ELISA reactivity stratified by visual reaction grades (+1 to +4) within low SAT end point titers (1:10 and 1:20).

SAT Dilution End point	Visual Reaction Grade	Evaluable Animals (n)	Serum I-ELISA Reactive	Milk I-ELISA Reactive
1:10	+1	2	1/2 (50.0%)	2/2 (100.0%)
	+2	7	7/7 (100.0%)	7/7 (100.0%)
	+3	4	2/4 (50.0%)	3/4 (75.0%)
	+4	17	9/17 (52.9%)	10/17 (58.8%)
1:20	+1	1	1/1 (100.0%)	1/1 (100.0%)
	+2	3	3/3 (100.0%)	3/3 (100.0%)
	+3	6	3/6 (50.0%)	5/6 (83.3%)
	+4	16	11/16 (68.8%)	13/16 (81.2%)

*Values represent reactive records over evaluable animals within each stratum, with corresponding percentages in parentheses. Reaction grades (+1 to +4) represent recorded visual tube agglutination intensity based on supernatant clearing and bacterial sedimentation according to standard protocols. All included animals (n = 56) possess fully interpretable, concurrent results across SAT, serum I-ELISA, and milk I-ELISA. Due to limited sample sizes across individual sub-strata (n = 1 to 17), these proportions provide a descriptive profile of biological variability rather than inferential evidence of a monotonic dose–response curve, antibody affinity/avidity, or stage of infection. Abbreviations: I-ELISA, indirect enzyme-linked immunosorbent assay; SAT, serum agglutination test.

low-endpoint stratum. Serum reactivity was 7/7 at 1:10 (+2), compared with 9/17 at 1:10 (+4); at 1:20 it was 3/3 (+2) and 11/16 (+4). Milk reactivity likewise showed no consistently increasing pattern across grades. With subgroup sizes of one to 17, these observations provide a descriptive account rather than evidence of a biological dose–response or an independently predictive grade effect. No antibody affinity, avidity, or isotype distribution was measured.

Individual profiles further illustrated distinct serological discordance patterns: animal 1 presented RBPT reactivity and SAT 1:20 (+3), but was non reactive

across CFT, MRT, and both I-ELISAs; animal 175 showed SAT 1:10 (+4) with reactive serum but non reactive milk I-ELISA; and animal 91 exhibited SAT 1:10 (+1) with non reactive serum but reactive milk I-ELISA.

In the assumption-based scenario, the SAT strata contained 102, 59, 22, and 87 animals. Low-endpoint reactivity was 38/59 (64.4%) in serum and 46/59 (78.0%) in milk. Interpreting the four ambiguous SAT denominators as endpoints yielded SAT–serum agreement of 68.9% (κ = 0.419) under $\geq$ 1:40 and 75.2% (κ = 0.454) under $\geq$ 1:10. Similarity to the primary results indicates limited descriptive impact of these

particular assumptions, without resolving the original ambiguities.

Interpretation of the principal findings: The principal contribution of this study is the analytical separation of completely non reactive and low-titer SAT profiles, along side the empirical modeling of diagnostic classifications that shift under alternative operational thresholds. The low-end point stratum harbored substantial serological heterogeneity, comprising both immuno assay-reactive and non reactive individuals, where visual reaction intensity grades (+1 to +4) demonstrated no monotonic correlation with IgG-specific ELISA outcomes.

These low-dilution findings challenge the simplistic assumption that all borderline agglutination titers denote true infection. Approximately two thirds of the low-dilution animals were serum I-ELISA reactive, leaving a substantial fraction without serum immuno assay support. Milk reactivity was more frequent, but was likewise incomplete. The appropriate inference is that low-dilution agglutination reflects a heterogeneous mixture of serological profiles warranting cautious, contextual interpretation. Assigning a uniform diagnostic classification to this entire stratum would obscure the very biological diversity revealed by discretization.

Likewise, SAT non reactivity cannot be equated with the absolute absence of measurable humoral immunity. Forty-two animals without observable tube agglutination (<1:10) exhibited serum I-ELISA positivity, and 66 demonstrated milk I-ELISA reactivity. The diagnostic meaning of a non reactive result is therefore strictly assay-dependent and governed by matrix-specific decision rules. In the absence of an independent microbiological reference standard, these divergent profiles cannot be definitively categorized as SAT false negatives or immuno assay true positives.

This biological interpretation aligns with established diagnostic literature^[9]. Emphasized evaluations of serological outcomes come within a comprehensive diagnostic framework, while Ducrotoy highlighted how operational utility fluctuates across epidemiological settings. Here, the documented absence of vaccination eliminates a major confounding source of post-vaccinal antibodies, yet it does not rule out heterospecific cross-reactions or unequivocally confirm active bacteremia. The archived dataset provides a valuable basis for characterizing comparative serological behavior; however, it cannot reconstruct temporal dynamics of infection or discriminate persistent residual titers from active bacterial shedding.

Relationship to previous serum and milk assay studies: The findings are compatible with the general rationale for evaluating serum and milk as distinct

analytical matrices. In their bovine validation study, Vanzini selected different percentage positivity cut offs for the two matrices and evaluated them in defined herd populations. Their design included infected herds with prior isolation evidence and disease-free herds, whereas the current study reconstructed categorical results in a clinically selected historical sample. These differences explain why agreement in the present data set cannot be interpreted using their reported diagnostic performance estimates^[10].

Compared milk I-ELISA with MRT in bulk milk and demonstrated substantial discordance in sample classification. Greater milk I-ELISA than MRT reactivity observed across the present sampled cattle aligns with these matrix-dependent differences. However, bulk milk and individual-animal milk represent fundamentally different analytical units; sample pooling dilutes individual antibody concentrations and alters analytical sensitivity relative to herd composition. Consequently, bulk-tank performance metrics cannot be directly generalized to individual-cow screening scenarios.

Sutra demonstrated the involvement of multiple immunoglobulin classes in the milk ring reaction, supporting the premise that agglutination-based milk test and IgG-direct immuno assays capture distinct immunological properties. In the present data set, immunoglobulin sub classes were not quantitatively resolved; thus, differential sub class distribution remains a plausible mechanistic explanation rather than an established determinant of matrix discordance.

Comparative assay rankings must also be extrapolated cautiously. Gall and Nielsen synthesized test performance parameters across diverse study populations, whereas the present cattle population was sampled following abortion episodes or clinical suspicion and lacks an independently confirmed bacteriological reference standard. Furthermore, Andrade demonstrated that the inclusion of case-control designs can artificially inflate pooled sensitivity and specificity estimates. Concordance patterns within this clinically selected population should therefore not be equated with diagnostic accuracy indices derived from prospective validation studies.

Alternative explanations and evidence that limits interpretation: Greater milk reactivity may reflect differences in decision thresholds, matrix composition, antibody availability, or nonspecific signal. These explanations are not mutually exclusive. Because the historical assay used different serum and milk cutoffs and the individual optical densities were unavailable, the present study cannot distinguish a biological matrix effect from an analytical classification effect. Calling

the milk assay intrinsically more sensitive would require an independent disease reference and suitable validation data.

Cross-reactivity also limits interpretation of concordant serological results^[10]. linked false positive brucellosis reactions in cattle with responses associated with *Yersinia enterocolitica* O:9. Examined multiple assay and antigen combinations and found that no combination simultaneously resolved all sensitivity and specificity concerns. These studies caution against assuming that agreement among antigenically related tests establishes organism specific infection^[11].

Agreement between tests that share biological targets is also not equivalent to independent corroboration. Showed that conditional dependence can affect the interpretation and combined performance of veterinary diagnostic tests^[12]. SAT and the historical immuno assays may capture related antibody responses, so their agreement cannot be assumed to supply multiple independent confirmations. Conditional dependence was not estimated in this reanalysis; the limitation concerns interpretation rather than a demonstrated dependence structure in these animals. The isolated agglutination pattern in animal 1 is compatible with several explanations, including nonspecific antibodies, biological variation, or borderline analytical variability. It does not prove cross-reactivity or identify another organism. Similarly, a milk-reactive/serum-non reactive pattern does not demonstrate mammary localization, latent infection, or ongoing bacterial shedding. Serological assays detect antibody-related reactivity rather than viable organisms, and pathogen DNA detection would itself require interpretation distinct from evidence of viability^[1].

Threshold tradeoffs and the contribution of reaction grading: The increase in agreement after including reactions from 1:10 upward was modest and accompanied by more SAT reactive/serum nonreactive classifications. The reclassification table is therefore more informative than a single improved agreement percentage. It shows that a lower descriptive threshold captured additional serum reactive observations while also introducing additional disagreement against the same comparator.

This tradeoff prevents designation of 1:10 as an optimal cutoff. Its clinical value would depend on the purpose of testing, the consequences of missed infection and unnecessary follow-up, the assay standardization, and the underlying disease distribution. Kappa provides an additional summary of agreement but cannot adjudicate those consequences. Reporting kappa alongside the underlying paired counts avoids treating a qualitative agreement label as a substitute for diagnostic validation^[13-14-15].

A change in the operational SAT rule alters the marginal proportion classified as reactive as well as the concordant cells. Explained how category distributions and disagreement between classifications influence kappa, and emphasized interpretation with uncertainty and the study context^[16-17-18]. The increase from $\kappa = 0.417$ to 0.464 should therefore be read together with the 37 agreement gains and 19 agreement losses. It does not quantify an improvement in infection detection and does not justify a clinical threshold solely on the basis of a higher coefficient.

Reaction grade did not supply a simple independent ranking of immuno assay reactivity within the low dilution strata. Some smaller +2 subgroups were entirely serum reactive, whereas the larger +4 subgroups were not. The results do not establish that weaker agglutination is diagnostically superior; chance, sparse cells, and the different information conveyed by grade and endpoint are sufficient reasons to avoid that inference. The defensible contribution is to retain grade in the data record and examine it separately, rather than assume a linear relationship with antibody concentration or disease severity.

Limitations: The clinically selected animal population limits external generalize ability. Source-reported absence of vaccination was not independently verified, and individual clinical variables such as parity, lactation stage, or time elapsed since abortion were unavailable for multivariable adjustment. Furthermore, herd-level identifiers were insufficient for cluster-aware robust variance estimation; confidence intervals therefore assume independent individual observations and may under state uncertainty if substantial within-herd cluster was present. Sample reactive proportions must therefore not be interpreted as herd-orregional-level disease prevalence.

Individuals' somatic cell counts, mastitis status, immunoglobulin sub classes, and direct measures of antibody affinity or avidity were not recorded. The lack of these biological parameters prevents definitive mechanistic attribution of the observed serum–milk discordance. Additionally, the archived source data set provides categorical immuno assay classifications rather than continuous optical densities or percent positivity (PP) values, precluding a direct re-estimation of receiver operating characteristic (ROC) curves. The two matrices also relied on different historical diagnostic cut offs, which inherently influences relative activity rates.

No concurrent bacteriological culture or molecular reference standard (such as PCR) was available to independently adjudicate definitive infection status. A small subset of historical entries with ambiguous recording formats was omitted from quantitative models, and sensitivity analyses demonstrated that alternative classification

assumptions yielded comparable overall concordance patterns without resolving historical threshold ambiguities. Consequently, the threshold scenarios evaluated here in remain exploratory; observed agreement shifts do not establish a definitive regulatory or clinically validated cut off. Future prospective validation studies would require an independent microbiological reference standard, quantitative multi-isotype assay outputs, and detailed animal- and herd-level covariates.

Emphasized that diagnostic validation is fundamentally shaped by target population structure, sampling design, and underlying covariates. The non-random sampling frame, sparse subgroups amplexes across certain visual grade strata, and lack of herd-level clustering variables inherently limit direct field extrapolation of these alternative operational rules. Document edvariations in global diagnostic practices Vection further under score the necessity of context-specific empirical validation prior to policy changes. The primary contribution of this work remains the transparent characterization of dilution end point and visual grade heterogeneity within the archived diagnostic records, rather than a definitive recommendation to replace established surveillance frameworks.

CONCLUSIONS

This study demonstrates that substantial serological heterogeneity exists among cattle without detectable SAT agglutination and within the borderline 1:10–1:20 titer stratum. Within low-endpoint subgroups, visual reaction grades did not exhibit a linear relationship with serum or milk immuno assay reactivity. Lowering the descriptive SAT threshold to $\geq 1:10$ captured 37 additional serum-reactive cases, moderately improving overall agreement while simultaneously introducing 19 discordant non reactive profiles, thereby highlighting the analytical trade-offs governing threshold adjustments. Although these classification scenarios do not establish a definitive diagnostic cutoff, the systematic integration of discrete titration endpoints, visual reaction intensity, and paired milk screening is anticipated to contribute substantially to the refinement of bovine brucellosis surveillance frameworks and the diagnostic interpretation of borderline serological cases.

Acknowledgements: This study was derived from the doctoral dissertation of Ufuk Ülker, completed under the supervision of Prof. Dr. Sumru Çıtak at Gazi University (2010), preliminary descriptive findings of which were presented as an oral presentation at the International Ankara Conference on Scientific Researches and published in its proceedings book [19-20-21-22-23]. The granular dilution end points tratifications, visual tube agglutination grades, and multi-threshold

reclassification models presented here in represent novel secondary evaluations not previously published or reported else where.

Ethical considerations: No animals were handled, sampled, bled, or subjected to any clinical or experimental intervention for the purpose of the present investigation. The original biological specimens (blood and milk) were collected during 2007–2008 for routine diagnostic screening of suspected bovine brucellosis under the institutional scope of the primary doctoral research. In accordance with international ethical standards (ARRIVE and COPE) and national animal welfare regulations, formal institutional animal care and use committee approval was not required for this computational re-evaluation of pre-existing, anonymized diagnostic records.

Data availability: The data supporting the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

1. World Organisation for Animal Health., 2022. Brucellosis (infection with *B. abortus*, *B. melitensis* and *B. suis*). In Manual of diagnostic tests and vaccines for terrestrial animals Chapter 3.1.4. 0 1-10.0.
2. Nielsen, K., 2002. Diagnosis of brucellosis by serology. Vet. Microbiol., 90: 1-10.0.
3. Godfroid, J., K. Nielsen and C. Saegerman, 2010. Diagnosis of Brucellosis in Livestock and Wildlife. Croatian Med. J., 51: 1-10.0.
4. Gall, D. and K. Nielsen, 2004. Serological diagnosis of bovine brucellosis : A review of test performance and cost comparison. Revue Scientifique Tech. l'OIE, 23: 1-10.0.
5. Ducrotoy, M.J., P.M. Muñoz, R. Conde-Álvarez, J.M. Blasco and I. Moriyón, 2018. A systematic review of current immunological tests for the diagnosis of cattle brucellosis. Preventive Vet. Med., 151: 1-10.0.
6. Andrade, R.S., M.M. de Oliveira, J.S.D.B. Filho, F. Ferreira, J. Godfroid, A.P. Lage and E.M.S. Dorneles, 2024. Accuracy of serological tests for bovine brucellosis: A systematic review and meta-analysis. Preventive Vet. Med., Vol. 222 .10.1016/j.prevetmed.2023.106079 1-10.0.
7. Vanzini, V.R., N. Aguirre, C.I. Lugaresi, S.T. de Echaide and V.G. de Canavesio et al., 1998. Evaluation of an indirect ELISA for the diagnosis of bovine brucellosis in milk and serum samples in dairy cattle in Argentina. Preventive Vet. Med., 36: 1-10.0.
8. Vection, S., C.G. Laine and A.M. Arenas-Gamboa, 2025. What do we really know about brucellosis diagnosis in livestock worldwide? A systematic

- review. *PLOS Neglected Trop. Dis.*, Vol. 19 .10.1371/journal.pntd.0013185 1-10.0.
9. Vanzini, V.R., N.P. Aguirre, B.S. Valentini, S.T. de Echaide, C.I. Lugaresi, M.D. Marchesino and K. Nielsen, 2001. Comparison of an indirect ELISA with the Brucella milk ring test for detection of antibodies to *Brucella abortus* in bulk milk samples. *Vet. Microbiol.*, 82: 1-10.0.
10. Sutra, L., J.P. Caffin and G. Dubray, 1986. Role of milk immunoglobulins in the brucella milk ring test. *Vet. Microbiol.*, 12: 1-10.0.
11. Weynants, V., A. Tibor, P.A. Denoel, C. Saegerman, J. Godfroid, P. Thiange and J.J. Letesson, 1996. Infection of cattle with *Yersinia enterocolitica* O:9 a cause of the false positive serological reactions in bovine brucellosis diagnostic tests. *Vet. Microbiol.*, 48: 1-10.0.
12. Muñoz, P.M., C.M. Mariu, D. Monreal, D. González and B. Garin-Bastuji et al., 2005. Efficacy of Several Serological Tests and Antigens for Diagnosis of Bovine Brucellosis in the Presence of False-Positive Serological Results Due to *Yersinia enterocolitica* O:9. *Clin. Vaccine Immunol.*, 12: 1-10.0.
13. Gardner, I.A., H. Stryhn, P. Lind and M.T. Collins, 2000. Conditional dependence between tests affects the diagnosis and surveillance of animal diseases. *Preventive Vet. Med.*, 45: 1-10.0.
14. McHugh, M.L., 2012. Interrater reliability: The kappa statistic. *Biochemia Med.*, 22: 1-10.0.
15. Byrt, T., J. Bishop and J.B. Carlin, 1993. Bias, prevalence and kappa. *J. Clin. Epidemiol.*, 46: 1-10.0.
16. Sim, J. and C.C. Wright, 2005. The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. *Phys. Ther.*, 85: 1-10.0.
17. Greiner, M. and I.A. Gardner, 2000. Epidemiologic issues in the validation of veterinary diagnostic tests. *Preventive Vet. Med.*, 45: 1-10.0.
18. von Elm, E., D. G. Altman., M. Egger and S. J. Pocock., et al., 2007. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies., *PLOS Med.*, Vol. 4 .10.1371/journal.pmed.0040296 1-10.0.
19. Brown, L.D., T.T. Cai and A. DasGupta, 2001. Interval Estimation for a Binomial Proportion. *Stat. Sci.*, 16: 1-10.0.
20. Getachew, T., G. Getachew, G. Sintayehu, M. Getenet and A. Fasil, 2016. Bayesian Estimation of Sensitivity and Specificity of Rose Bengal, Complement Fixation, and Indirect ELISA Tests for the Diagnosis of Bovine Brucellosis in Ethiopia. *Vet. Med. Int.*, Vol. 2016 .10.1155/2016/8032753 1-10.0.
21. Cohen, J., 1960. A Coefficient of Agreement for Nominal Scales. *Educ. Psychological Meas.*, 20: 1-10.0.