Title: Diagnostic serology test comparison for Q fever and Rift Valley fever in humans and livestock from pastoral communities

Authors: Valerie Hungerbühler¹, Ranya Özcelik¹, Mahamat Fayiz Abakar², Fatima Abdelrazak Zakaria², Martin Eiden³, Sonja Hartnack⁴, Pidou Kimala², Sonja Kittl⁵, Janine Michel⁶, Franziska Suter-Riniker⁷, Salome Dürr¹

Affiliations

¹Veterinary Public Health Institute, Vetsuisse Faculty, University of Bern, Bern, Switzerland
²Institut de Recherche en Elevage pour le Développement, N'Djamena, Chad
³Friedrich-Loeffler-Institut, Federal Research Institute for Animal Health, Greifswald – Insel Riems, Germany
⁴Section of Epidemiology, Vetsuisse Faculty, University of Zurich, Zurich, Switzerland
⁵ZOBA, Center for Zoonoses, Animal Bacterial Diseases and Antimicrobial, Vetsuisse Faculty, University of Bern, Bern, Switzerland
⁶Robert Koch Institute, Centre for Biological Threats and Special Pathogens, Highly Pathogenic Viruses, Berlin, Germany
⁷IFIK, Institute for Infectious Diseases, Faculty of Medicine, University of Bern, Bern, Switzerland

Correspondence: Valerie Hungerbühler
Veterinary Public Health Institute, Vetsuisse Faculty
University of Bern
Schwarzenburgstrasse 161
3097 Liebefeld, Switzerland
M: +41 77 423 18 62
E: valerie.hungerbuehler@vetsuisse.unibe.ch

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

Q fever (QF) and Rift Valley fever (RVF) are endemic zoonotic diseases in African countries, causing significant health and economic burdens. Accurate prevalence estimates, crucial for disease control, rely on robust diagnostic tests. While enzyme-linked immunosorbent assays (ELISA) are not the gold standard, they offer rapid, cost-effective, and practical alternatives. However, varying results from different tests and laboratories can complicate comparing epidemiological studies. This study aimed to assess the agreement of test results for QF and RVF in humans and livestock across different laboratory conditions and, for humans, different types of diagnostic tests.

We measured inter-laboratory agreement using concordance, Cohen's kappa, and prevalence and bias-adjusted kappa (PABAK) on 91 human and 102 livestock samples collected from rural regions in Chad. The samples were tested using ELISA in Chad, and indirect immunofluorescence assay (IFA) (for human QF and RVF) and ELISA (for livestock QF and RVF) in Switzerland and Germany. Additionally, we examined demographic factors influencing test agreement, including district, setting (village vs. camp), sex, age, and livestock species of the sampled individuals.

The inter-laboratory agreement ranged from fair to moderate. For humans, QF concordance was 62.5%, Cohen's kappa was 0.31, RVF concordance was 81.1%, and Cohen's kappa was 0.52. For livestock, QF concordance was 92.3%, Cohen's kappa was 0.59, RVF concordance was 94.0%, and Cohen's kappa was 0.59. Multivariable analysis revealed that QF test agreement is significantly higher in younger humans and people living in villages compared to camps and tends to be higher in livestock from Danamadjí compared to Yao, and in small ruminants compared to cattle. Additionally, RVF agreement was found to be higher in younger humans.

Our findings emphasize the need to consider sample conditions, test performance, and influencing factors when conducting and interpreting epidemiological seroprevalence studies.
Q fever (QF) and Rift Valley fever (RVF) are zoonotic diseases that can be transmitted from animals to humans, causing health problems and economic losses in African countries. While various diagnostic tests for these diseases are available, they can be impractical, especially in resource-limited settings.

For this study, human and livestock samples from Chad were first tested in a local laboratory using a routine test. The same samples were then sent to laboratories in Germany or Switzerland for retesting, using the same test type for livestock and a different test type for human samples.

We analysed the agreement between the test results and investigated the influence of the demographic characteristics of the sampled individual on this agreement. Our findings are crucial as they reveal discrepancies in test results, even though the samples originated from the same individuals. Additionally, we found that factors such as the age of the sampled individual influenced test agreement.

This study underscores the importance of considering sample conditions, test performance, and influencing factors when conducting and interpreting disease prevalence studies. Enhancing diagnostic procedures will aid in more effective disease control management, benefiting local communities and global health efforts.
1. Introduction

Q Fever (QF) and Rift Valley fever (RVF) are zoonotic diseases prevalent in several African countries. Reported prevalence rates range from 7.8% to 39% for QF and 9.5% to 44.2% for RVF in livestock, and from 27% to 49.2% for QF and 13.2% to 28.4% for RVF in humans [1–4]. QF and RVF impact human health by causing a flu-like syndrome that can lead to a range of severe manifestations. QF and RVF also result in significant production losses in animals due to abortions [5,6].

High-quality samples and robust diagnostic tests are essential for obtaining accurate prevalence estimates. Epidemiological studies play a critical role in generating the necessary data, subsequently influencing government prioritization of health interventions [7,8]. This prioritization is fundamental to effective disease control.

For QF diagnostics, the indirect immunofluorescence assay (IFA) can differentiate between acute and chronic infections and is regarded as the gold standard test for humans [9,10]. Commercial kits are not available for veterinary use [11]. The enzyme-linked immunosorbent assay (ELISA) is the most widely used test and is recommended by the WOAH (World Organisation for Animal Health) for rapid routine screening and large-scale epidemiological studies in ruminant populations [8]. For RVF, the virus neutralisation test (VNT) is the most specific diagnostic serological test, but it can only be performed with live viruses and is not recommended for use in laboratories without appropriate biosecurity facilities and vaccinated personnel [6,7]. For both diseases, ELISAs offer a rapid, cost-effective, and practical alternative with less stringent biosafety requirements, making them suitable for routine use in low- and middle-income countries (LMIC) [7,8].

The use of ELISA in some studies and the gold standard test, which may differ between countries, in others can lead to discrepancies in estimated prevalence, making comparisons challenging; thus, harmonized monitoring and reporting schemes for QF and RVF have been proposed to enable consistent comparisons over time and across countries [12–14].

Several studies have assessed the inter-test agreement of ELISA for QF and RVF compared to other diagnostic tests, reporting variable agreement ranging from poor to good for QF and from good to excellent for RVF [15–21]. Diagnostic test validation can be achieved through various methods,
including assessing the agreement between different tests without assuming one as the gold standard [22]. Concordance, the proportion of test results in agreement over the number of all tests performed, is a straightforward measure but does not account for agreement beyond chance. Therefore, Cohen's kappa statistic, which adjusts for random matches, is often used to measure the agreement between two test results [22]. Cohen’s kappa values range from zero (agreement is equal to that expected by chance) to one (complete agreement beyond chance), with benchmarks between agreement categories varying among authors [23–25]. Although Cohen's kappa is a standard measure, it has limitations such as prevalence and bias effects. Prevalence effects arise when the proportion of positive results deviates significantly from 50% [26]. The effect of prevalence depends on the method of modeling agreement and can substantially reduce kappa values [26]. Bias effects occur when there is a disparity in the proportion of positive results between the two tests, which leads to reduced kappa values [26]. To address these effects, the prevalence- and bias-adjusted kappa (PABAK) can be calculated [27,28].

The reasons for disagreement between diagnostic tests have rarely been thoroughly investigated. Potential factors include poor sample quality, variability in tests used, and discrepancies arising from the same test being conducted in different laboratories [17,21,29]. Additionally, biological factors such as age, sex, other diseases, and species may influence the consistency of test results for the same sample. Previous studies have suggested associations between test performance and variables such as region, age, and livestock species [30–32]. However, these studies have not provided conclusive evidence or statistical significance.

The objective of this study was to assess the inter-laboratory agreement, measured by the concordance, Cohen's kappa, and PABAK, of results obtained from commercial ELISA tests conducted in a laboratory in Chad and results obtained from ELISA and indirect IFA tests for livestock and human samples respectively, performed in laboratories in Germany and Switzerland. Additionally, we evaluated the influence of demographic factors on the agreement between the two test results. The study enhances our understanding of the inter-laboratory agreement of diagnostic test results across laboratory conditions and, for humans, test types, which is crucial for accurately interpreting results from epidemiological seroprevalence studies.
2. Material and Methods

2.1 Ethics statement
The study has been submitted to and approved by the Ethics Committee of Northwest and Central Switzerland (EKNZ) (project id 2017-00884) and by the Comité National de Bioéthique du Tchad (CNB-Tchad) (project id 134/PR/MESRS/CNBT/2018). Formal written consent was obtained from study participants and animal owners after we presented our study to the community and before data collection occurred.

2.2 Sample collection and laboratory analysis in Chad
The samples analysed in this study were collected between January and February 2018 [4]. In brief, a cross-sectional study in livestock (cattle, sheep, goats, horses, and donkeys) and human populations was conducted in the two rural health districts, Yao and Danamadji, in Chad. Multistage cluster sampling was used, with villages and nomadic camps serving as cluster units. In Danamadji and Yao, respectively, blood samples were collected from apparently healthy 571 and 389 humans and 560 and 483 livestock. The samples were subsequently analysed at the Institut de Recherche en Élevage pour le Développement (IRED) in N’Djamena, Chad. Livestock and human samples were analysed using different indirect ELISAs: ID Screen® Q Fever Indirect Multi-species ELISA from IDvet for livestock and the Panbio® Coxiella burnetii IgG ELISA from Abbott for humans. For RVF, a competitive ELISA (ID Screen® Rift Valley Fever Competition Multi-species from IDvet) was used for human and livestock samples. The diagnostic test procedure and thresholds were applied according to the manufacturer’s protocols without modification (Table S1 – Table S3). Equivocal samples were retested once.

2.3 Diagnostic testing in Switzerland and Germany
Following the initial diagnostic analysis at IRED, 10% of the human and livestock samples from each region were randomly selected and sent to laboratories in Switzerland and Germany in 2021 for repeated diagnostic analysis for QF and RVF, respectively. In Switzerland, two indirect ELISAs (IDEXX Q Fever IgG Antibody and ID Screen® Q Fever Indirect Multispecies from IDvet) were used at the Center for
Zoonoses, Animal Bacterial Diseases, and Antimicrobial Resistance (ZOBA) for QF diagnostics in livestock samples (ruminants and equids, respectively). At the Institute for Infectious Diseases (IFIK) of the University of Bern, an indirect IFA (Q Fever IFA IgG assay from Focus Diagnostics, US) was used for QF diagnostics in human samples. For RVF, livestock samples were analysed using a competitive ELISA (ID Screen® Rift Valley Fever Competition Multi-species ELISA from IDvet) at the Federal Research Institute for Animal Health (FLI), and human samples were analysed using an indirect IFA (Anti-Rift-Tal-Fieber-Viren-IIFT [IgG] from EUROIMMUN) at the Robert Koch Institute (RKI). The diagnostic test procedure and thresholds were applied according to the manufacturer’s protocols without modification (Table S1, Table S3, Table S4 – Table S6). Equivocal samples were not retested.

Fig 1: Sample collection and laboratory analysis at IRED in Chad and diagnostic testing in Switzerland and Germany. The numbers of samples at each stage are indicated.

2.4 Statistical analysis
The inter-laboratory agreement of the test results from Chad and Switzerland or Germany was evaluated using concordance, Cohen’s Kappa, and PABAK for each of the four datasets: QF results in human samples, QF results in livestock samples, RVF results in human samples, and RVF results in livestock samples. Cohen's kappa and PABAK values were interpreted according to the standard scale: 'fair' agreement (kappa = 0.21–0.40), 'moderate' agreement (kappa = 0.41–0.60), 'substantial' agreement (kappa = 0.61–0.80), and 'almost perfect' agreement (kappa > 0.80) [25].

In addition, we investigated factors associated with test agreement by assigning to each sample a value of 0 if there was disagreement between the two test results (i.e., negative in Chad and positive in Switzerland/Germany, or the opposite) and 1 if the test results were consistent (i.e., both positive or both negative). This binary outcome was used as the dependent variable in logistic regression models to identify the statistical association between test agreement and demographic factors, including the district (Yao versus Danamadji) and setting (village versus camp) where the sample was collected, and sex, age, and livestock species (cattle, small ruminants, and equids) of the sampled individual. The variable age was analysed in two ways, as a continuous and as a categorical variable. For the continuous variable, a unit of 1 year was used for livestock and of 10 years for humans. For the categorical variable, samples were stratified as < 2 years (age group 1), 2–3 years (age group 2), 4 years and older (age group 3) for livestock, and < 30 years (age group 1), 30–39 years (age group 2), 40–60 years (age group 3), 61 years and older (age group 4) for humans.

Univariable logistic regressions were initially performed to assess individual predictors. In cases where the univariable model was infeasible due to perfect agreement in one group, Chi-square tests were applied. Odds ratios (OR) and their corresponding 95% confidence intervals were calculated for these analyses. To consider potential interdependencies between the variables, we applied multivariable logistic regressions to estimate adjusted coefficients and OR. We included all variables and selected age as categorical variable.

Statistical calculations, modeling, and data visualization were conducted in R (version 4.2.2). The package “irr” was used to calculate the concordance and Cohen’s kappa. The package “vcd” was used to obtain confidence intervals of Cohen’s kappa, computed using the standard method based on normal approximation [33]. The epi.kappa() function from the “epiR” package was used to calculate PABAK.
and corresponding confidence intervals. The confidence intervals for the OR were calculated using the output values of the associated regression model and the functions \(\exp(\text{coefficients(model)})\) for the upper CI and \(\exp(\text{confint.default(model)})\) for the lower CI was used.
3. Results

3.1 Study population and samples

From 103 livestock and 96 human samples sent to Switzerland and Germany, not all of them could be used for statistical analysis. Reasons for excluding them include unsuccessful matching of test identity due to labelling errors, missing serum upon arrival at the laboratory, and equivocal results (Table 1). Finally, 91 human and 102 livestock samples were tested for either one of both tests, depending on the availability of serum (Table S7 and Table 1, Fig. 1).

Table 1: Overview of diagnostic tests from Chad and Switzerland (CH) or Germany (GE).

<table>
<thead>
<tr>
<th>Disease</th>
<th>QF¹</th>
<th>RVF²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Samples</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laboratory, Chad</td>
<td>IRED³</td>
<td>IRED³</td>
</tr>
<tr>
<td>Test 1, Chad</td>
<td>ID Screen® Q Fever</td>
<td>Panbio® Coxiella burnetii IgG ELISA</td>
</tr>
<tr>
<td>Laboratory, CH/GE</td>
<td>ZOBA (CH)⁴</td>
<td>IFIK (CH)⁵</td>
</tr>
<tr>
<td>Test 2, CH/GE</td>
<td>IDEXX Q Fever Ab Test and the ID Screen® Q Fever Indirect Multispecies</td>
<td>Focus Diagnostics Q Fever IFA IgG assay</td>
</tr>
</tbody>
</table>

| Initial number of samples shipped to CH/GE | 103 | 96 | 103 | 96 |
| Excluded samples from test 2 due to missing serum | 2 | 8 | 0 | 4 |
| Equivocal results during test 2 | 8 | 14 | 2 | 0 |
| Non-matchable | 2 | 2 | 1 | 2 |
| Number of samples used for agreement testing | 91 | 72 | 100 | 90 |

¹Q Fever
²Rift Valley Fever
³Institut de Recherche en Elevage pour le Développement, Chad
⁴Center for Zoonoses, Animal Bacterial Diseases and Antimicrobial, Switzerland (CH)
Of the 91 human and 102 livestock samples that were used to perform the intra-laboratory test agreement analysis, most samples were collected in Danamadji, with 62% of human samples and 57% of livestock samples originating from this region (Table S7). Fifty-six percent of human samples and 58% of livestock samples were collected from camps. The sex distribution was uneven, with 70% of human samples being men and 70% of livestock samples being females. Among humans, age groups 1-3 were evenly represented (30%, 28%, 33%), while only 9% belonged to age group 4. In livestock, 50% of the samples were from age group 2, with 17% and 33% from age groups 1 and 3, respectively. Most livestock samples were from cattle (46%), followed by small ruminants (41%) and equids (13%).

3.2 Diagnostic tests agreement

3.2.1 Level of inter-laboratory test agreement

Concordance values ranged from 62.5% to 94% (Table 2). Cohen's kappa values, which ranged from 0.31 to 0.59, indicated that livestock QF and RVF, and human RVF tests had 'moderate' agreement, while human QF tests had 'fair' agreement (Table 2). PABAK values showed that the livestock QF and RVF tests had 'almost perfect' agreement, the human RVF test had 'substantial' agreement, and the human QF tests had 'fair' agreement (Table 2).

Table 2: Analyses of test agreement (concordance and Cohen's kappa) of samples tested in an inter-laboratory test agreement study.

<table>
<thead>
<tr>
<th>Switzerland</th>
<th>Germany</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chad</td>
<td></td>
</tr>
<tr>
<td>QF livestock</td>
<td>QF humans</td>
</tr>
<tr>
<td>Pos</td>
<td>Neg</td>
</tr>
<tr>
<td>Pos</td>
<td>6</td>
</tr>
<tr>
<td>Neg</td>
<td>4</td>
</tr>
<tr>
<td>Concordance (%)</td>
<td>92.3</td>
</tr>
<tr>
<td>Cohen's kappa value (95% CI)</td>
<td>0.59 (0.31, 0.86)</td>
</tr>
</tbody>
</table>
### 3.2.1 Influence of factors on inter-laboratory test agreement

For QF in livestock, none of the investigated demographic factors significantly impacted the agreement between the two test results in both univariable and multivariable analyses (Table S8 and Table 3). However, some notable trends ($p < 0.15$) emerged: small ruminants tended to show better agreement than cattle, and samples from Yao showed lower agreement compared to those from Danamadji (Table 3). For QF in humans, samples from villages had significantly higher agreement compared to those from camps, with odds of agreement being 13.4 times higher (Table 3). Additionally, older age groups had significantly lower agreement compared to the youngest age group (Table 3).

For RVF in humans and livestock, none of the demographic factors significantly influenced test agreement. However, some trends were observed for RVF test agreement in humans, with older age groups showing lower agreement compared to the youngest age group (Table 3).

Across diseases and populations, there was a consistent trend of lower agreement with increasing age, which was significant for human QF and almost reached significance for human RVF (Table 3, Fig. 2).

For livestock RVF tests, the odds ratio for agreement was also lower in older age groups, although the $p$-value was 0.26 (Table 3).

### Table 3: Results of multivariable logistic regression models investigating the effect of demographic factors as independent variables on the inter-laboratory test agreement of Q fever (QF) and Rift Valley Fever (RVF) in humans and livestock. Goodness of fit of the models are presented as pseudo-$R^2$-squared ($R^2$).

<table>
<thead>
<tr>
<th>Independent variable (OR and 95% CI)</th>
<th>Switzerland (QF livestock)</th>
<th>Switzerland (QF humans)</th>
<th>Germany (RVF livestock)</th>
<th>Germany (RVF humans)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>P-value (P)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>District</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Reference: Danamadji)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yao:</td>
<td>$0.14 (0.01, 1.74)$</td>
<td>$1.38 (0.38, 4.99)$</td>
<td>$4.99 (0.41, 60.49)$</td>
<td>$1.37 (0.37, 5.04)$</td>
</tr>
<tr>
<td><strong>P</strong></td>
<td>$P = 0.13$</td>
<td>$P = 0.63$</td>
<td>$P = 0.21$</td>
<td>$P = 0.64$</td>
</tr>
<tr>
<td>Setting</td>
<td>Village:</td>
<td>Village:</td>
<td>Village:</td>
<td>Village:</td>
</tr>
<tr>
<td>---------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>(Reference: camp)</td>
<td>0.34 (0.05, 2.06)</td>
<td>13.35 (3.20, 55.86)</td>
<td>0.32 (0.04, 2.43)</td>
<td>1.85 (0.58, 5.93)</td>
</tr>
<tr>
<td></td>
<td>P = 0.24</td>
<td><strong>P = 0.0004</strong></td>
<td>P = 0.27</td>
<td>P = 0.30</td>
</tr>
<tr>
<td>Species</td>
<td>Equids:</td>
<td>Equids:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Reference: cattle)</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.51 (0.29, 103.21)</td>
<td>Inf (perfect agreement)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P = 0.25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Small ruminants:</td>
<td>Small ruminants:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.26 (0.59, 65.76)</td>
<td>0.65 (0.09, 4.78)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P = 0.13</td>
<td>P = 0.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age group</td>
<td>Group 2:</td>
<td>Group 2:</td>
<td>Group 1*:</td>
<td>Group 2:</td>
</tr>
<tr>
<td>(Reference: age group 1/group 2*)</td>
<td>1.30 (0.10, 17.39)</td>
<td>0.67 (0.15, 3.07)</td>
<td>Inf (perfect agreement)</td>
<td>0.62 (0.12, 3.16)</td>
</tr>
<tr>
<td></td>
<td>P = 0.84</td>
<td>P = 0.61</td>
<td>P = 0.57</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Group 3:</td>
<td>Group 3:</td>
<td>Group 3*:</td>
<td>Group 3:</td>
</tr>
<tr>
<td></td>
<td>0.93 (0.06, 14.38)</td>
<td>0.14 (0.03, 0.82)</td>
<td>0.30 (0.04, 2.41)</td>
<td>0.32 (0.07, 1.47)</td>
</tr>
<tr>
<td></td>
<td>P = 0.96</td>
<td><strong>P = 0.03</strong></td>
<td>P = 0.26</td>
<td>P = 0.14</td>
</tr>
<tr>
<td></td>
<td>Group 4:</td>
<td>Group 4:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.09 (0.01, 0.83)</td>
<td>0.16 (0.02, 1.14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>P = 0.03</strong></td>
<td>P = 0.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>Female:</td>
<td>Female:</td>
<td>Female:</td>
<td>Female:</td>
</tr>
<tr>
<td>(Reference: male)</td>
<td>0.38 (0.03, 4.61)</td>
<td>0.82 (0.20, 3.38)</td>
<td>0 (perfect agreement in males)</td>
<td>1.01 (0.27, 3.86)</td>
</tr>
<tr>
<td></td>
<td>P = 0.45</td>
<td>P = 0.78</td>
<td>P = 0.99</td>
<td></td>
</tr>
<tr>
<td><strong>R²</strong></td>
<td>0.23</td>
<td>0.35</td>
<td>0.28</td>
<td>0.09</td>
</tr>
</tbody>
</table>
Fig. 2: In the mosaic plot illustrating the association between age groups and inter-laboratory test agreement, the widths of the adjacent rectangles represent the proportion of individuals within each age group. The heights of the rectangles represent the proportion of individuals within each age group who exhibit a specific test agreement outcome (concordant or discrepant).
QF and RVF are important zoonotic diseases in sub-Saharan Africa, for which several epidemiological questions remain open. Reliable diagnostics are highly relevant for generating prevalence data, and it has been emphasized that there is a need for cost-effective surveillance tools for low and middle-income countries [5,34].

Our study revealed varying levels of test agreement, ranging from fair to moderate (Cohen’s kappa) or almost perfect when considering PABAK. The good inter-laboratory agreement of livestock test results for RVF was in line with other studies on RVF test agreement, although it is important to note that literature is scarce [19,20]. The inter-laboratory agreement for livestock QF was slightly better than expected, based on previous studies that assessed ELISA’s agreement with different test types [15,17,18]. We observed a notably lower agreement for human tests for both diseases that can be attributed to using two different tests for human samples, ELISA in Chad and indirect IFA in Switzerland or Germany. Previous studies have shown varying sensitivities and specificities of the commercial diagnostic test for QF used in this study (Panbio®), ranging from 71% to 100% [10,35–37]. In these studies, indirect IFA was used as a reference method to evaluate the ELISA, revealing varying agreement between the results of the two tests. The variability in sensitivity for the same test raises the question of whether it is due to the scope for interpretation in indirect IFA, which is considered a challenge, even though indirect IFA is regarded as the gold standard for human QF diagnostics [5,8]. A study that compared QF indirect IFA results from different reference centres in three countries (United Kingdom, France, and Australia) found a concordance between the indirect IFA results of only 35% [38]. Our results presented here reflect this uncertainty and underline the complexity of QF diagnostics [5,8,39]. In addition, commercial QF ELISA kits, unlike indirect IFA, cannot distinguish between acute or chronic infection and vaccination, which can sometimes lead to misinterpretation and discordant test results, even between different ELISA kits [21,40]. For RVF, lower inter-laboratory agreement for human tests compared to livestock tests can also be attributed to using different tests in Chad (ELISA) and Germany (indirect IFA). The literature on RVF test agreement, particularly for human diagnostics, is notably scarce.
A possible reason for disagreement between test results could be the hemolytic quality of some of our samples. Although we lack quantitative information on the quality of the individual samples in our study, laboratory staff in Switzerland and Germany indicated that haemolysis was a concern. In recent literature, haemolysis in specimens was reported to be the most common cause of test result discrepancies in clinical laboratories [41]. There are many in vitro causes of haemolysis, mostly pre-analytical problems such as incorrect procedures and/or materials used in blood collection, while transport, processing, and storage account for only a minority of cases [42]. This limitation shows the importance of careful planning and execution of the pre-analytical phase, especially in prevalence studies where the outcome can be influenced by the quality of the sample material. Nevertheless, it is crucial to recognize the challenges associated with collecting samples under difficult field conditions, where access to centrifuges may be limited until several days after blood sampling. In addition, the samples in our study were stored for 2.5 years with repeated freeze-thaw cycles between the performance of the two tests for some of the samples, which probably affected sample quality. Therefore, extended transport to secondary laboratories for subsequent analysis can additionally adversely affect sample quality. We emphasize the importance of considering these challenges when discussing the outcomes of epidemiological prevalence studies or diagnostic test evaluation studies.

Our results demonstrate a statistical relationship between test agreement and age of the sampled individuals, with higher agreement observed in younger individuals for both diseases and in humans and livestock. This finding aligns with a study by de Bronsvoort et al. (2019), which suggested that lower agreement in older individuals may be due to their higher likelihood of previous exposure to other pathogens over their lifetime [30]. This may result in cross-reactivity in serological tests, making serological differentiation between diseases more difficult [39]. Cross-reactions caused by antibodies provoked by other pathogens, such as *C. burnetii* antigens, the agent causing QF, with antibodies produced against *Bartonella spp.*, *Legionella spp.*, and *Chlamydiae spp.* has been reported [43–45], or RVF virus antigens with antibodies produced against Rio Grande virus [46–48]. Furthermore, with increasing age, the likelihood of exposure to the causative agent of QF and RVF increases [49,50], as does the chance of having residual antibodies titres against these diseases in the blood [6,51–53]. In addition in QF, antigen-lipopolysaccharide complexes remaining in the host after infection with *C.
*burnetii* can trigger humoral and cell-mediated immune responses, producing interfering antibodies [54,55]. Such antibody titres can potentially produce ambiguous results that are neither clearly positive nor negative, leading to misinterpretation and discordant test results. These findings underscore the need for further research to develop age-specific diagnostic protocols and improve test accuracy across diverse populations.

The district-level analysis showed a trend toward higher agreement for livestock QF tests in samples from Danamadj compared to Yao. This regional variation may be influenced by varying local environmental conditions, disease prevalence, or livestock management practices. However, we did not capture such information, so we were not able to identify a potential latent factor that can explain the difference in test agreement detected for the two regions. Continuing with the focus on sampling location, the sample collection setting (village versus camp) was identified as a significant factor in the multivariable model for human QF test agreement. One hypothesis for this finding is that villages generally have shorter distances from the sampling location to the laboratory, allowing for more appropriate storage conditions compared to remote camps. This shorter travel time likely preserves sample integrity, resulting in higher test agreement. The significant role of the setting variable suggests that environmental and logistical factors may be crucial for diagnostic test result interpretation, although we did not observe the influence of the setting for RVF in humans nor for livestock species. Future studies should investigate the impact of environmental and logistical factors on test results and gather detailed data on local conditions and practices that affect sample quality and test agreement.

Small ruminants exhibited a trend towards higher agreement in QF test results compared to cattle. In a different type of QF ELISA, Stellfeld et al. (2020) observed significant differences in the ranges of OD<sub>450</sub> values obtained from sera of sheep, goats, and cattle [56]. Sheep exhibited a large range of OD<sub>450</sub> values, whereas cattle showed a smaller range [56]. Greater distribution between OD<sub>450</sub> values allows for more precise grading of ELISA results, which may explain the better inter-test agreement for small ruminants compared to cattle. These differences in OD<sub>450</sub> ranges are likely due to species-specific immune responses to *C. burnetii*, suggesting varying immune reactions among ruminant species [56]. Although females showed lower inter-laboratory test agreement compared to males for both diseases and populations, the high p-values and wide confidence intervals of OR indicate that these results are
not statistically robust and should be interpreted with caution. Epidemiological studies suggest similar QF exposure rates between genders, but symptoms are 2.5 times more common in men [5]. This variability in inflammatory immune response between genders [57–60] could lead to less clear antibody titers in samples from females, a hypothesis that should be tested by further studies investigating the influence of demographic factors on diagnostic test agreement.
Conclusion

Our study highlights the variability in inter-laboratory diagnostic test agreement for QF and RVF serology in humans and livestock based on samples collected in Chad. Despite differences in laboratories, personnel, and test types, test agreements ranged from fair to moderate (Cohen’s kappa) or almost perfect considering PABAK. Given the reliance on serological profiles for QF and RVF epidemiological studies, it is crucial to consider factors that may complicate accurate diagnosis. We identified that human QF test agreement was significantly higher in individuals living in villages and younger individuals, with the latter trend also observed in human RVF tests. Our findings emphasize the need to recognize that diagnostic tests may yield varying results, impacting the outcome and interpretation of disease prevalence studies.
Declaration of generative AI and AI-assisted technologies in the writing process

While preparing this work, the author used ChatGPT to correct the English in the revised sentences (using the command: 'correct the English'). After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

References


