SEMINA
Ciências Agrárias

DOI: 10.5433/1679-0359.2024v45n1p177

Interobserver and inter-cycle variability in the electrocardiographic assessment of dogs

Variabilidade interobservador e entre ciclos cardíacos na avaliação eletrocardiográfica de cães

Fernando Luis Cemenci Gnoatto[1]*; Estela Dall'Agnol Gianezini[2]; Tainá Minuzzo[2]; Tatiana Champion[3]

Highlights

Observers with little experience overestimate P wave duration values.
The degree of agreement does not proportionally reflect the experience of observers.
The assessment of a cardiac cycle can represent the average of cardiac cycles.

Abstract

The electrocardiogram, which graphically records the phases of the cardiac cycle, plays a crucial role in diagnosing arrhythmias and conduction abnormalities. However, the reliability of results can be significantly affected by the level of expertise of the evaluator. Acknowledging potential discrepancies in interpretations of the same electrocardiographic readings across different cardiac cycles and between analysts of varying experience, this study aimed to assess the interobserver and inter-cycle variability in electrocardiographic exams of dogs. This retrospective research was carried out at the Veterinary Hospital Superintendency Unit of the Federal University of Fronteira Sul, Realeza Campus, PR, Brazil. The study involved a random selection of electrocardiographic records from 50 dogs, collected between September 2018 and December 2019, without discrimination based on breed, sex, age, or weight. Various parameters were analyzed, including heart rate and rhythm; durations of P waves; QRS complexes; PR and QT intervals; amplitudes of P, R, and T waves; and both atrial and ventricular electrical axes. The data underwent Shapiro-Wilk normality testing, with parametric data analyzed via analysis of variance and Tukey's test, and non-parametric data assessed using the Kruskal-Wallis followed by Dunn's tests. The final step involved applying the Kappa Agreement Coefficient Test to the data. The findings revealed variability in heart rhythm interpretation, with disagreement between evaluators of low to intermediate experience levels. Specifically, low-level evaluators tended to underestimate maximum heart rates,

[1] Student of the Undergraduate Course in Veterinary Medicine, Universidade Federal da Fronteira Sul, UFFS, Realeza, PR, Brazil. E-mail: gnoatto.f.l.c.27@gmail.com

[2] Students of the Master's Course of the Postgraduate Program in Health, Welfare and Sustainable Animal Production in Fronteira Sul, UFFS, Realeza, PR, Brazil. E-mail: egianezini@gmail.com; minuzzotaina@gmail.com

[3] Prof[a] Dr[a], Undergraduate Course in Veterinary Medicine and Postgraduate Program in Health, Welfare and Sustainable Animal Production on the Southern Border, UFFS, Realeza, PR, Brazil. E-mail: tatiana.champion@uffs.edu.br

* Author for correspondence

Received: Mar. 06, 2023 Approved: Feb. 13, 2024

recording a median rate of 105, compared to 122 by high-level evaluators (p=0.0100). Additionally, P wave durations were consistently overestimated by low- and intermediate-level evaluators, who reported median durations of 52 and 51, respectively, against 48 by their highly experienced counterparts (p=0.0064). No significant differences emerged in analyses comparing a single cardiac cycle to either three cycles or the average of three cycles. The study highlighted the impact of evaluator expertise on the variability of electrocardiographic interpretations, particularly concerning P wave duration, for which the evaluators tended to overestimate measurements, and the degrees of agreement, which did not represent their experience level. The findings stress the importance of training and routine practice in electrocardiographic examinations to mitigate the risk of mistaken interpretations and analyses.
Key words: Heart rate. Electrocardiogram. Level of experience. Observation errors.

Resumo

O eletrocardiograma representa o registro gráfico das fases do ciclo cardíaco e é utilizado para o diagnóstico de arritmias ou distúrbios de condução. No entanto, a confiabilidade dos resultados pode ser influenciada pelo grau de experiência dos avaliadores. Devido à possibilidade de variação de resultados de um mesmo exame entre os ciclos cardíacos e entre analisadores de diferentes níveis de experiência, o estudo teve como objetivo avaliar a variabilidade interobservador e entre ciclos cardíacos durante a avaliação de exames eletrocardiográficos de cães. A pesquisa foi conduzida de forma retrospectiva na Superintendência Unidade Hospitalar Veterinária Universitária da Universidade Federal da Fronteira Sul, Campus Realeza – PR. Para a análise eletrocardiográfica foram selecionados aleatoriamente exames de 50 cães, gravados em arquivos computadorizados durante o período de setembro de 2018 a dezembro de 2019, independente da raça, sexo, idade e peso. Foram avaliados frequência e ritmo cardíaco, durações de onda P, complexo QRS, intervalo PR e intervalo QT, amplitudes de onda P, R e T, e eixo elétrico atrial e ventricular. Os dados foram analisados por teste de normalidade de Shapiro-Wilk, sendo os dados paramétricos analisados por análise de variância seguido por Teste de Tukey e os dados não paramétricos pelo Teste de Kruskal-Wallis, seguido pelo Teste de Dunn. Como última etapa, os dados foram submetidos ao Teste de Coeficiente de Concordância de Kappa. Evidenciou-se variabilidade na interpretação de ritmos cardíacos com discordância de avaliadores de nível baixo e intermediário, diferença da frequência cardíaca máxima do avaliador de baixo nível de experiência que subestimou os valores, obtendo mediana de 105, enquanto o de alto nível obteve mediana de 122 (p=0,0100). Ainda, evidenciou-se superestimação da duração da onda P pelos avaliadores de nível baixo e intermediário, que encontraram medianas de 52 e 51 respectivamente, enquanto o de alto nível obteve mediana de 48 (p=0,0064). Em relação a diferentes ciclos cardíacos, não houve diferenças entre a análise de apenas um ciclo em relação a três ciclos ou a média de três ciclos. Por meio do estudo foi possível evidenciar a variabilidade na interpretação entre avaliadores com diferentes níveis de experiência, em relação a duração da onda P, em que os avaliadores superestimaram os valores e em relação aos graus de concordância que não representaram os níveis de experiência dos avaliadores na análise do todas as variáveis, ressaltando a importância do treinamento e rotina de exames eletrocardiográficos para evitar interpretações e análises equivocadas.
Palavras-chave: Ritmo cardíaco. Eletrocardiograma. Nível de experiência. Erros de observação.

Introduction

Electrocardiography serves as a crucial diagnostic method that charts the electrical potentials produced by the heart, showcasing myocardial depolarization and repolarization phases (Porsani et al., 2020). By virtue of its properties, this technique is capable of identifying alterations in electrical impulse flow, impulse origins, and their frequencies (Silveira et al., 2018).

Owing to its diagnostic capabilities, the electrocardiogram has become a staple in veterinary practice, effectively identifying arrhythmias, cardiac conduction blocks, myocardial oxygenation issues, electrolyte imbalances, pleural effusions, and cardiac chamber enlargements (Figueiredo et al., 2016; Porsani et al., 2020). It is routinely employed in clinical assessments, particularly for geriatric patients and during pre-surgical evaluations (Figueiredo et al., 2016; Ferreira et al., 2017; Silveira et al., 2018; Porsani et al., 2020).

Electrocardiographic readings are characterized by specific wave durations, amplitudes, and directions, with potential variances attributable to the size of the animal and, most importantly, its species (Macêdo et al., 2019). Deviations from established norms in the electrocardiogram can indicate or hint at conduction blocks, arrhythmias, and atrioventricular overload (Bombardelli et al., 2021).

However, the reliability of electrocardiographic findings can be compromised by interobserver variability, particularly due to subjective interpretations when the analysis is strongly linked to qualitative aspects (Porsani et al., 2020). The assessment of such qualitative variables can be challenging (Leroy et al., 2013).

The expertise of the observer significantly influences the interpretation of various aspects of electrocardiographic examination, with diagnostic proficiency closely tied to the level of training of the evaluator (Leroy et al., 2013). Consequently, inexperienced observers often fail to diagnose conditions that are readily identified by their trained counterparts (Bagardi et al., 2021).

Porsani et al. (2020) highlight the potential for interobserver discrepancies, especially concerning the identification, measurement, and interpretation of electrocardiographic data. Moreover, variations in measurements can arise independently of an observer's experience level, but rather due to the evaluator themselves, and may also occur among evaluators with comparable expertise. Previous studies have shown a 4% variability in electrocardiogram interpretations among cardiologists and a 9% variability among sports medicine specialists (Berte et al., 2015; Bagardi et al., 2021).

Given the limited research on the optimal number of cardiac cycles to analyze in a tracing for accurate detection of electrical conduction abnormalities, this study explores the hypothesis that differences may exist between analyzing a single cycle, three cycles, or the average of three cycles. The objective is to ascertain whether evaluating multiple cycles is necessary for routine assessments.

Additionally, we considered the potential impact of the experience level of the evaluator on the analysis, examining variability across three distinct levels of expertise in electrocardiographic interpretation. Therefore, the study aims to assess both interobserver variability and inter-cycle variability during the electrocardiographic evaluation of dogs.

Materials and Methods

The study was carried out as a retrospective analysis at the Veterinary Hospital Superintendency Unit (SUHVU) of the Federal University of Fronteira Sul (UFFS), located at the Realeza Campus, PR, Brazil. G*Power software (version 3.1.9.7) was utilized to determine the necessary sample number, employing the configuration for F tests – ANOVA with repeated measurements, across different factors. This analysis identified a cohort of 45 animals, achieving a statistical power of 0.95 (with an effect size of 0.5, α = 0.05, and a power of 95%). The electrocardiographic assessments were based on the computerized records of canine patients seen by the Veterinary Cardiology Service at SUHVU – UFFS. These records, collected from September 2018 to December 2019, were chosen randomly and included dogs of various breeds, sexes, ages, and weights. The only requirement for inclusion was the complete registration of each animal's data, while any records with missing information were excluded.

The initial data collection focused on fundamental information such as breed, sex, age, and weight from the cardiology service records at SUHVU. Electrocardiographic recordings were conducted using bipolar leads (I, II, and III) and augmented unipolar leads (aVR, aVL, and aVF), employing a multichannel, computerized system (ECGPC TEB®). The calibration standard was set at 1 cm/mV, with a recording speed of 50 mm/s, and the recordings lasted around five minutes each.

The electrocardiograms were assessed by three evaluators with different levels of expertise in electrocardiographic interpretation. These included a low-level (an undergraduate in their final year of Veterinary Medicine), an intermediate-level (a Veterinary professional pursuing graduate studies with 1.6 years of experience), and a high-level evaluator (a Veterinary professional and professor with 20 years of experience in the field). Each evaluator analyzed the electrocardiographic tracings independently and at separate times.

Analyses focused on the bipolar II lead, examining aspects such as heart rate, rhythm, and the durations of the P wave, QRS complex, PR interval, and QT interval in milliseconds, as well as the amplitudes of the P, R, and T waves in millivolts. Observations also included the polarity of the T wave (positive, negative, or biphasic) and the ST segment amplitude across three consecutive cardiac cycles. To calculate the mean atrial and middle ventricular electrical axes in the frontal plane, the amplitude of the P wave and QRS complex values in leads I and III were measured. This was followed by calculation using a graphical method, which consists of measuring the deflection of the waves in leads I and III, followed by comparison with the hexaxial system and determining the direction of the axes, as presented by Santilli et al. (2018).

For the interpretation of the electrocardiographic tracings, the findings were compiled in spreadsheets and compared with established normal values from the literature. The parameters for evaluating the P wave and QRS complex durations were based on those proposed by Wolf et al. (2000), while the criteria for other durations, wave amplitudes, and electrical axis degrees were adopted from Santilli et al. (2018).

Qualitative variables (sex, breed, T wave polarity, and electrocardiographic interpretations) were presented as percentages reflecting their total occurrences. The Shapiro-Wilk test was applied to assess the normality of quantitative variables. Parametric data were expressed as mean ± standard deviation, while non-parametric data were presented as medians accompanied by interquartile ranges (25%-75%). For parametric variables, an analysis of variance (ANOVA) followed by Tukey's post hoc test was employed, whereas non-parametric variables were analyzed using the Kruskal-Wallis test with subsequent Dunn's test, adopting a significance level of $p < 0.05$.

As a last step, Cohen's Kappa Agreement Coefficient Test was performed to assess the concordance between evaluations made by the low- and intermediate-level evaluators compared to those made by the high-level evaluator. The Kappa agreement coefficient (κ) values were interpreted as follows: <0 indicated no agreement, 0-0.20 slight agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, 0.61-0.80 substantial agreement, and 0.81-1.0 almost perfect agreement (Landis & Koch, 1977). All statistical analyses were conducted using Graphpad Prism 5.0 software.

Results and Discussion

In the analysis encompassing three cardiac cycles evaluated by each of the three observers, a total of 450 interpretations were made. These interpretations covered various parameters: durations of the P wave, QRS complex, PR interval, QT interval, amplitudes of the P, R, and T waves, as well as atrial and ventricular electrical axes. On the other hand, assessments of heart rhythm, heart rate, and T wave polarity were conducted only once for each examination. Electrocardiography, a non-invasive diagnostic tool, is extensively utilized in veterinary practice for its utility in diagnosing cardiac conditions. Nevertheless, factors such as the use of different devices can introduce variability in electrocardiographic tracings, potentially affecting interpretation accuracy (Figueiredo et al., 2016; Ferreira et al., 2017; Silveira et al., 2018; Porsani et al., 2020).

To mitigate the impact of equipment variability on the electrocardiographic tracings, this study ensured that all examinations were performed using the same device. It is crucial to approach the evaluation of electrocardiographic tracings with diligence, particularly for computerized electrocardiography, where the durations of the P wave and QRS complex may deviate from the standards established for traditional methods. Reference values specific to computerized electrocardiography should be consulted due to their differences from those of the conventional approach (Arias et al., 2021).

From the database of the SUHVU-UFFS cardiology service, 50 electrocardiographic exams were reviewed. The breed distribution included 12 Poodles, 11 mixed breeds, eight Pinschers, four Shih Tzus, three Yorkshire Terriers, two Border Collies, three Dachshunds, two Boxers, one Lhasa Apso, one Pug, one Cocker Spaniel, one French Bulldog, and one Golden Retriever.

Regarding the sex distribution of the subjects, 62% (31/50) of the examinations were of female animals, while 38% (19/50) were of male animals. The age of the subjects was 8 ± 4.3 years, and their average weight was 6 (4.5-9.6) kg.

Table 1 presents the heart rhythms identified by the highly experienced evaluator in the electrocardiographic examinations, accompanied by their absolute and relative frequencies.

Table 1
Absolute and relative frequencies of heart rhythms found by the evaluator with a high level of experience in the electrocardiographic analysis of 50 patients

Heart rhythm	Absolute frequency (50)	Relative frequency (%)
Sinus arrhythmia with wandering pacemaker	12	24
Sinus arrhythmia	10	20
Sinus arrhythmia with wandering pacemaker and sinus arrest	8	16
Sinus rhythm	7	14
Sinus tachycardia	3	6
Sinus arrhythmia with first-degree atrioventricular block	2	4
Sinus arrhythmia with ventricular premature complexes	2	4
Sinus arrhythmia with atrial premature complexes	1	2
Sinus rhythm with paroxysmal supraventricular tachycardia	1	2
Sinus arrhythmia with sinus arrest and ventricular premature complexes	1	2
Sinus arrhythmia with wandering pacemaker, isolated premature atrial complexes, and bigeminy	1	2
Sinus arrhythmia with second-degree Mobitz type II atrioventricular block	1	2
Sinus arrhythmia with ventricular extrasystole and accelerated idioventricular rhythm	1	2

Out of 150 analyses conducted by three observers, agreement on the basic rhythm of the electrocardiographic recordings was reached in 68% (101) of cases. This left 32% (49) of the heart rhythms with discrepancies among the evaluators. Differences were noted particularly from the least experienced observer and the one with intermediate experience. Disagreements were more frequent with the least experienced observer, accounting for 51% (25) of cases, and 49% (24) by the intermediate-level observer, when compared to the findings of the most experienced observer. Many of these disagreements were attributed to a lack of experience in identifying specific electrocardiographic features throughout the recordings, such as atrioventricular blocks, wandering pacemaker, ventricular premature complexes, and sinus arrest.

Table 2 presents the electrocardiographic variables measured by observers with varying levels of experience in dogs, showing mean values with standard deviations or median values with interquartile ranges.

Table 2

Mean (standard deviations) or median (interquartile ranges) values of electrocardiographic variables measured by observers with different levels of experience in dogs

	Observer 1	Observer 2	Observer 3	p
Min. HR	97 (83-107)[a]	86 (74-94)[a]	87 (78-110)[a]	0.1329
Max. HR	105 (91-132)[a]	111 (99-144)[ab]	122 (108-152)[b]	0.0100
Avg. HR	151 (122-160[a]	146 (128-169)[a]	150 (124-178)[a]	0.8476
P (ms)	**52 (44-57)[a]**	**51 (46-57)[a]**	**48 (43-50)[b]**	**0.0064**
PR (ms)	94 (77-105)[a]	86 (76-103)[a]	87 (76-100)[a]	0.7102
QRS (ms)	55 (48-60)[a]	57 (51-63)[a]	54 (48-60)[a]	0.3617
QT (ms)	198 ± 24[a]	206 ± 22[a]	203 ± 21[a]	0.2676
P (mV)	0.28 ± 0.10[a]	0.26 ± 0.11[a]	0.29 ± 0.11[a]	0.2599
Q (mV)	0.07 (0.00-0.14)[a]	0.09 (0.00-0.16)[a]	0.08 (0.00-0.14)[a]	0.7843
R (mV)	1.13 (0.82-1.41)[a]	1.15 (0.82-1.45)[a]	1.13 (0.87-1.47)[a]	0.9465
T (mV)	0.30 (0.20-0.47)[a]	0.35 (0.18-0.53)[a]	0.23 (0.16-0.37)[a]	0.3157
ST (mV)	0.04 (0.02-0.08)[a]	0.03 (0.00-0.08)[a]	0.05 (0.00-0.08)[a]	0.3766
º P	60 (49-72)[a]	62 (44-75)[a]	61 (49-72)[a]	0.9616
º QRS	70 (54-79)[a]	69 (53-78)[a]	70 (55-77)[a]	0.9667

HR: Heart rate. Min.: Minimum. Max.: Maximum. Avg.: Average. Observer 1: Low-experience-level evaluator. Observer 2: Intermediate-experience-level evaluator. Observer 3: High-experience-level evaluator. Medians in the row followed by different lowercase letters are statistically different by the Test.

In dogs exhibiting sinus arrhythmia, both minimum and maximum heart rates were calculated, followed by an average rate, whereas for dogs with regular rhythms, only the average rate was determined. There was no significant difference in the minimum (p=0.1329) and average (p=0.8476) heart rates among the observers. However, the maximum heart rate identified by the least experienced observer was significantly different from that identified by the most experienced observer (p=0.0100), likely due to the inexperience of the former in selecting the electrocardiographic segment with the highest heart rate during sinus arrhythmia.

Furthermore, Table 2 indicates a significant discrepancy in the measurement of P wave duration (p=0.0064) between the less experienced and intermediate-level observers compared to the most experienced observer. This discrepancy arose because the less experienced and intermediate-level observers tended to overestimate the duration of the P wave, reporting median values of 52 and 51 ms, respectively, whereas the most experienced observer reported a median value of 48 ms. Such overestimation can lead to incorrect diagnoses of left atrial overload by less experienced observers, as prolonged P wave duration may indicate this condition. However, the sensitivity of electrocardiography in detecting atrial or ventricular overload is limited (Santilli et al., 2018; Arias et al., 2021; Bombardelli et al., 2021).

This large difference is primarily due to the differing levels of experience among the evaluators, as experience can impact measurements in other diagnostic tests, such as echocardiography (Leroy et al., 2013). In human medicine, acceptable intra- and interobserver variability in P wave measurement in electrocardiographic tracings has been reported, with high-resolution monitors enhancing the analysis quality and helping to reduce observer variability (Dilaveris et al., 1999). For the authors, the tendency of less experienced evaluators to overestimate P wave measurements stems from their uncertainty in accurately determining the start and end points of the wave.

Variability in electrocardiographic interpretation can lead to unnecessary additional tests, which may cause stress for the animal and incur significant costs (Dores et al., 2017; Porsani et al., 2020). Other electrocardiographic parameters such as the QRS complex, PR and QT intervals, and the amplitudes of the P, R, Q, and T waves did not show significant differences among the evaluators.

Table 3 reveals that 64% (288/450) of the measured P wave durations exceeded the established normal ranges for dogs (Santilli et al., 2018; Arias et al., 2021). Of these, 34.72% (100/288) were identified by the least experienced evaluator, 36.8% (106/288) by the intermediate-level evaluator, and 28.47% (82/288) by the most experienced evaluator. The P wave on an electrocardiogram signifies depolarization of the left and right atria. The duration of a normal P wave should not exceed 44.54 ± 1.04 milliseconds for dogs under 9.9 kg, 46.42 ± 0.86 milliseconds for dogs between 10 and 20 kg, and 48.50 ± 1.70 milliseconds for dogs over 20 kg. The amplitude should not surpass 0.4 mV, as higher values may indicate right atrial overload (Santilli et al., 2018; Arias et al., 2021).

Table 3
Absolute and relative frequencies of electrocardiographic alterations found in the analysis of 450 records from 50 patients, frequency of alterations found by observers with different levels of experience (low, intermediate, and high), and description of electrocardiogram reference values for dogs

Electrocardiographic variable	Reference values†	Absolute frequency (n=450)	Relative frequency (%)	Observer 1	Observer 2	Observer 3
P (ms)	< 44.54 ± 1.04* < 46.42 ± 0.86** < 48.50 ± 1.70***	288	64	34.7% (100)	36.8% (106)	28.4% (82)
PR (ms)	60-130	24	5.3	33.3% (8)	41.6% (10)	25% (6)
QRS (ms)	< 54.78 ± 1.61* < 59.70 ± 1.18** < 63.30 ± 1.54***	192	42.6	32.8% (63)	35.4% (68)	31.7% (61)
QT (ms)	150-240	18	4	33.3% (6)	50% (9)	16.6% (3)
P (mV)	< 0.4	60	13.3	31.6% (19)	25% (15)	43.3% (26)
R (mV)	0.5 - 3	48	10.6	31.2% (15)	37.5% (18)	31.2% (15)
P axis (º)	-18º to +90º	11	2.4	36.3% (4)	18.1% (2)	45.4% (5)
QRS axis (º)	+40º to +100º	66	14.6	37.8% (25)	34.8% (23)	27.2% (18)

Observer 1: Low-experience-level evaluator. Observer 2: Intermediate-experience-level evaluator. Observer 3: High-experience-level evaluator. * Up to 9.9 kg; ** From 10 to 19.9 kg; *** Over 19.9 kg. † (Wolf et al., 2000; Santilli et al., 2018).

For the QRS complex duration, 42.6% (192/450) of the readings were above the normal range. Of these, 32.81% (63/192) were reported by the least experienced evaluator, 35.41% (68/192) by the intermediate-level evaluator, and 31.77% (61/192) by the most experienced evaluator. The QRS complex represents the depolarization of the left and right ventricles. The acceptable durations for the QRS complex vary with dog size (Arias et al., 2021), with durations beyond 54.78 ± 1.61 milliseconds for dogs under 9.9 kg, 59.70 ± 1.18 milliseconds for dogs between 10 and 20 kg, and 63.30 ± 1.54 milliseconds for dogs over 20 kg suggesting left ventricular overload. R wave amplitudes should range from 0.5 mV to 2.5 mV in small- to medium-sized dogs, and up to 3.0 mV in larger breeds (Santilli et al., 2018; Arias et al., 2021).

It is important to note that computerized methods may yield higher P wave and QRS complex duration values compared to traditional methods. Therefore, the type of recording equipment should be taken into account when interpreting these durations to avoid interpretation errors. Additionally, human observers tend to be more accurate in detecting anomalies on computer screens than with traditional methods, which may explain the increased duration values for the P wave and QRS complex (Oliveira et al., 2013; Porsani et al., 2020; Arias et al., 2021). Nevertheless, such measurements should be approached cautiously, with the expertise of the observer playing a crucial role in accurately determining the start and end points of the electrocardiographic waveform.

As shown in Table 4, for P wave duration, there was fair agreement (κ=0.278) between the low- and high-level evaluators and moderate agreement (κ=0.488) between the intermediate- and high-level evaluators. In the case of QRS complex duration, there was moderate agreement (κ=0.431) between the low- and high-level evaluators and fair agreement (κ=0.382) between the intermediate- and high-level evaluators. Unlike what was observed for the duration of the P wave, the agreement between evaluators was greater in relation to the low-level evaluator.

Table 4
Kappa values and degrees of agreement of electrocardiographic results obtained among evaluators with low and intermediate levels of experience relative to the results obtained by the evaluator with a high level of experience

Electrocardiographic variable	Kappa	Agreement	Kappa	Agreement
P (ms)	0.278	Fair	0.488	Moderate
PR (ms)	0.431	Moderate	0.382	Fair
QRS (ms)	0.790	Substantial	0.790	Substantial
QT (ms)	0.658	Substantial	0.370	Fair
P (mV)	0.852	Almost perfect	0.766	Substantial
R (mV)	0.779	Substantial	0.898	Almost perfect
QRS axis (º)	0.623	Substantial	0.811	Almost perfect

Observer 1: Low-experience-level evaluator. Observer 2: Intermediate-experience-level evaluator. Kappa agreement coefficient (κ), κ<0 indicates no agreement, 0-0.20 slight agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, 0.61-0, 80 substantial agreement, and 0.81-1.0 almost perfect agreement (Landis & Koch, 1977).

It is important to interpret with caution the findings where 64% of P wave duration measurements indicated potential atrial overload and 42.6% of QRS complex duration measurements suggested possible left ventricular enlargement (Bombardelli et al., 2021). These findings in smaller animals, particularly changes related to the P wave, may be evidenced with increasing age, such as the deposition of fibrous tissue in the heart due to apoptosis and an increase in connective tissue within the cardiac interstitium, which are typical of aging. These changes may also coincide with prevalent cardiac conditions like degenerative valve disease, which can cause atrial and ventricular remodeling (Oliveira et al., 2013; Mendes et al., 2019; Arias et al., 2021; Bombardelli et al., 2021).

Regarding PR interval duration, 5.3% (24/450) of the readings fell outside the normal range for dogs. Analysis by the evaluators showed that 87.5% (21/24) of these were prolonged and 12.5% (3/24) were shortened, relative to the normal range of 60 to 130 milliseconds. The distribution of these measurements among the evaluators was 33.3% (8) by the low-level, 41.6% (10) by the intermediate-level, and 25% (6) by the high-level evaluator. Both the low- and intermediate-level evaluators achieved substantial agreement (κ=0.790) regarding PR interval duration. Accurate measurement of the PR interval is critical, as a shortened PR interval may indicate conditions such as ventricular pre-excitation, while a prolonged PR interval can be indicative of atrioventricular conduction delays or blocks (Santilli et al., 2018). It is important to recognize that an observer's limited experience can result in the selection of PR intervals that are inaccurately short or long, failing to reflect the true duration on the electrocardiographic tracing. Changes in the PR interval may be evident paroxysmally, and could be overlooked by less experienced observers due to their lack of expertise (Santilli et al., 2018).

In the evaluation of the QT interval duration, 4% (18/450) of the measurements deviated from the normal range. The distribution of these measurements was as follows: 33.3% (6/18) were assessed by the least experienced evaluator, achieving substantial agreement (κ=0.658); 50% (9/18) by the intermediate-level evaluator, with fair agreement (κ=0.370); and 16.6% (3/18) by the most experienced evaluator. Thus, it is crucial for evaluators, particularly those with less experience, to carefully identify the start of repolarization and the end of myocardial depolarization. QT interval variations may indicate electrolyte imbalances, and according to Martin (2007), prolongations potentially signal conditions such as hypocalcemia, hypokalemia, and hypothermia, whereas shortenings can be indicative of hypercalcemia and hyperkalemia.

For the amplitudes of the P and R waves, 13.3% (60 out of 450) and 10.6% (48/450) of the readings, respectively, were found to be outside the accepted norms for the canine species. The distribution of these deviations among the observers is detailed in Table 1. The low-level evaluator achieved an almost perfect agreement (κ=0.852) for P wave amplitude, while the intermediate-level evaluator reached substantial agreement (κ=0.766). In contrast, for the R wave amplitude, the low-level evaluator had substantial agreement (κ=0.779), and the intermediate-level evaluator achieved an almost perfect agreement (κ=0.898). The variations in P wave amplitude are justified, particularly in cases with a wandering pacemaker, in which the greater amplitude is more likely to be recognized by the more experienced observer.

In terms of T wave polarity, there was a 12% discrepancy (19/150) with the assessments of the highly experienced evaluator. Among these divergent findings, 47.3% (9/19) were made by the least experienced evaluator and 52.7% (10/19) by the intermediate-level evaluator. The least experienced evaluator reached an almost perfect agreement (κ=0.825), while the intermediate-level evaluator had substantial agreement (κ=0.615). T wave changes are often non-specific and challenging to interpret. Selecting a section of the electrocardiogram with minimal artifacts and uniformity is essential for accurately determining T wave polarity (Santilli et al., 2018). It is important to note that the presence of artifacts and a non-linear baseline can lead to misinterpretation of biphasic waves, especially by observers with a lower level of experience.

While myocardial depolarization occurs omnidirectionally, depolarization vectors, measured in degrees, delineate the direction of depolarization at various stages of the cardiac cycle (Santilli et al., 2018). Concerning the average atrial and ventricular electrical axes, disagreement with established norms for canines was observed in 2.4% (11/450) and 14.6% (66/450) of results, respectively. For the middle ventricular electrical axis, the least experienced evaluator achieved a substantial agreement level (κ=0.623), whereas the moderately experienced evaluator reached an almost perfect agreement (κ=0.811). This indicates that the electrocardiographic tracings, as assessed by novices and semi-experienced evaluators, exhibited minimal alterations in myocardial electrical transmission (Santilli et al., 2018). Generally, wave and interval durations displayed lower agreement coefficients, in contrast to wave amplitudes and the ventricular electrical axis, which showed higher scores, indicating greater consensus among evaluators.

Table 5 presents the mean values and standard deviations or median (interquartile ranges) of electrocardiographic variables from three cardiac cycles and their average in dogs. The study revealed a lack of significant differences in electrocardiographic assessments across three cardiac cycles, suggesting that the measurement of a single cycle by a well-trained or minimally trained evaluator could represent the findings of additional cycles within the electrocardiogram.

Table 5
Mean values and standard deviations or median (interquartile ranges) of electrocardiographic variables measured in three cardiac cycles and the average of three cardiac cycles in dogs

	Cycle 1	Cycle 2	Cycle 3	Average of three cycles	p
P (ms)	48 (43-53)[a]	50 (43-53)[a]	50 (43-54)[a]	49 (44-54)[a]	0.8092
PR (ms)	90 (77-103)[a]	89 (75-103)[a]	88 (76-103)[a]	88 (77-103)[a]	0.9770
QRS (ms)	53 (50-60)[a]	55 (50-62)[a]	53 (50-63)[a]	55 (49-60)[a]	0.9690
QT (ms)	202 ± 23[a]	202 ± 23[a]	202 ± 22[a]	202 ± 22[a]	0.9981
P (mV)	0.27 (0.20-0.34)[a]	0.27 (0.20-0.34)[a]	0.26 (0.20-0.34)[a]	0.27 (0.20-0.34)[a]	0.9999
Q (mV)	0.08 (0.00-0.14)[a]	0.07 (0.00-0.14)[a]	0.08 (0.00-0.15)[a]	0.08 (0.00-0.14)[a]	0.9988
R (mV)	1.2 (0.8-1.5)[a]	1.1 (0.8-1.4)[a]	1.1 (0.8-1.4)[a]	1.1 (0.8-1.4)[a]	0.9967
T (mV)	0.29 (0.19-0.46)[a]	0.27 (0.18-0.45)[a]	0.28 (0.19-0.43)[a]	0.28 (0.18-0.46)[a]	0.9711
ST (mV)	0.03 (0.00-0.09)[a]	0.04 (0.00-0.09)[a]	0.04 (0.00-0.08)[a]	0.04 (0.00-0.08)[a]	0.9577
P axis (º)	60 (49-74)[a]	60 (46-74)[a]	60 (49-71)[a]	61 (48-72)[a]	0.9937
QRS axis (º)	70 (53-79)[a]	69 (53-78)[a]	70 (53-78)[a]	70 (54-78)[a]	0.9990

Medians in the row followed by different lowercase letters are statistically different by the Kruskal-Wallis Test.

The data imply that assessing more than one cardiac cycle during electrocardiographic exams might be unnecessary, as one cycle could typify the average of those encountered in the evaluation. Nevertheless, vigilance is advised to detect any alterations that might influence the accurate measurement of the waves. Despite no variance observed across cycle analyses, measuring multiple cycles is recommended to verify any deviations from the norm, thus solidifying the interpretation within the electrocardiographic report and ensuring the validation of the results.

Electrocardiographic recordings were analyzed by various evaluators at different instances to preclude bias and skewed interpretations. Yet, the choice of different cardiac cycles partially reflects agreement or erroneous measurements, given the potential for variations within the same electrocardiogram.

Conclusion

The study revealed variability in electrocardiographic interpretation among observers of differing expertise levels. Observers with a lower level of experience showed variable agreement with the most seasoned professional. Interpretative errors, such as overestimation of heart rhythms and P wave duration by low- and intermediate-level evaluators, could lead to inaccuracies in electrocardiographic assessments and reports. Although no difference was noted in the measurement of other waves, the degrees of agreement varied. Despite the lack of significant differences in interpretations across various cardiac cycles, where one cycle could represent the average of the recordings, electrocardiographic tracings must be approached with caution, particularly by those with little experience.

References

Arias, G. B., Oliveira, M. K. G., Schnitzer, J. F., Di Santis, G. W., Martins, M. I. M., & Gava, F. N. (2021). Eletrocardiografia computadorizada em cadelas com tumores mamários. *Acta Scientiae Veterinariae, 49*(1821), 1-6. doi: 10.22 456/1679-9216.114172

Bagardi, M., Manfredi, M., Zani, D. D., Brambilla, P. G., & Locatelli, C. (2021). Interobserver variability of radiographicmethods for the evaluation of left atrial size in dogs. *Veterinary Radiology & Ultrasound, 62*(2), 161-174. doi: 10.1111/vru.12930

Berte, B., Duytschaever, M., Elices, J., Kataria, V., Timmers, L., Heuverswyn, F. V., Stroobandt, R., Neve, J. D, Watteyne, K., Vandensteen, E., Vandekerckhove, Y., & Tavernier, R. (2015). Variability in interpretation of the electrocardiogram in young athletes: an unrecognized obstacle for electrocardiogram based screening protocols. *European Society of Cardiology, 17*(9), 1435-1440. doi: 10.1093/europace/euu385

Bombardelli, M. M. L., Champion, T., Fischborn, J. C. J., & Gusso, A. B. F. (2021). Diagnostic accuracy of the electrocardiogram for detection of atrial and ventricular overloads in dogs. *Acta Scientiae Veterinariae, 49*(1803), 1-8. doi: 10.22456/1679-9216.105274

Dilaveris, P., Batchvarov, V., Gialafos, J. & Malik, M. (1999), Comparison of Different Methods for Manual P Wave Duration Measurement in 12-Lead Electrocardiograms. *Pacing and Clinical Electrophysiology, 22*, 1532-1538. doi 10.1111/j.1540-8159.1999.tb00358.x

Dores, H., Santos, J. F., Dinis, P., Costa, F. M., Mendes, L., Monge, J., Freitas, A., Gonçalves, P. A., Cardim, N., & Mendes, M. (2017). Variabilidade na interpretação do eletrocardiograma do atleta: mais uma limitação na avaliação pré-competitiva. *Revista Portuguesa de Cardiologia, 36*(6), 443-449. doi: 10.1016 /j.repc.2016.07.013

Ferreira, G. B., Filippi, M. G., Paes, A. C., & Lourenço, M. L. G. (2017). Avaliação eletrocardiográfica em cães com erliquiose monocítica. *Revista de Educação Continuada em Medicina Veterinária e Zootecnia do CRMV-SP, 15*(2), 38-44. doi: 10.36440/recmvz.v15 i2.37335

Figueiredo, V. C., Pereira, C. S., Muzzi, R. A. L., Borges, J. C., Muzzi, L. A. L., Oberlender, G., Oliveira, M. M., & Abreu, C. B. (2016). Importância da eletrocardiografia como um exame pré-cirúrgico em cães. *Pesquisa Veterinária Brasileira, 36*(11), 1091-1094. doi: 10.1590/S0100-736X2016001100006

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *International Biometric Society, 33*(2), 363-374. doi: 10.2307/2529786

Leroy, C., Conchou, F., Layssol-lamour, C., Deviers, A., Sautet, J., Concordet, D., & Mogicato, G. (2013). Normal canine prostate gland: repeatability, reproducibility, observer-dependent variability of ultrasonographic measurements of the prostate in healthy intact beagles. *Anatomia Histologia Embryologia, 42*(5), 355-361. doi: 10.1111/ahe.12022

Macêdo, H. J. R., Silva, J. M. C., Mendes, I. L., Lopes, R. V., Vasconcelos, A. L. C. F., & Almeida, A. P. (2019). Principais alterações no eletrocardiograma em cães. *Ciência Animal, 29*(3), 38-49. https://pesquisa.bvsalud.org/portal/resource/pt/vti-25877

Martin, D. (2007) *The Complete Study Guide to Learning the Electrocardiogram.* Mind-Forge Education Services.

Mendes, T. G., Sousa, P. R., Moura, R. S., & Carvalho, R. O. A. (2019). Biomarcadores cardíacos e sua aplicabilidade clínica em cães e gatos. *Enciclopédia Biosfera, 16*(29), 412-436. doi: 10.18677/EnciBio_2019A31

Oliveira, L. S., Santos, R. R. B., Melo, M. B., Laranjeira, D. F., & Barrouin-Melo, S. M. (2013). Eletrocardiografia computadorizada em cães: estudo comparativo. *Pesquisa Veterinária Brasileira, 33*(7), 949-953. doi: 10.1590/S0100-736X2013000700018

Porsani, M. Y. H., Oliveira, L. E. D., Muzzi, R. A. L., & Abreu, C. B. (2020). Variabilidade interobservador entre três métodos de eletrocardiografia (convencional, digital e digital impresso) em gatos saudáveis. *Acta Scientiae Veterinariae, 48*(1761), 1-6. doi: 10.22456/1679-9216.97775

Santilli, R., Perego, M., Sydney, M., & Pariaut, R. (2018). *Electrocardiography of the dog and cat: diagnosis of arrhythmias* (2nd ed.). Edra.

Silveira, S. D., Gheller, B. G., & Meirelles, A. C. F. (2018). Preoperative electrocardiographic study of dogs at the veterinary hospital of pontifícia universidade católica do paraná. *Ciência Animal Brasileira, 19*(1), 1-12. doi: 10.1590/1809-6891v19e-49000

Wolf, R., Camacho, A. A., & Souza, R. C. A. (2000). Eletrocardiografia computadorizada em cães. *Arquivo Brasileiro de Medicina Veterinária e Zootecnia, 52*(6), 1-6. doi: 10.1590/S0102-09352000000600010