BSAVA BRITISH SMALL ANIMAL VETERINARY ASSOCIATION
JSAP JOURNAL OF SMALL ANIMAL PRACTICE

PAPER

Diagnostic accuracy of computer aided electrocardiogram analysis in dogs

A. H. Estrada*[,1], A. Spake[†], M. E. Kleman[†], D. Leeder[†], D. Blischok-Lapekas[†], M. Margiocco[†], J. Gentile-Solomon[†], N. Piscitelli[†] and D. Szlosek[†]

*Department of Small Animal Clinical Sciences, University of Florida, College of Veterinary Medicine, Gainesville, FL 32610, USA
[†]IDEXX Laboratories, Inc, One IDEXX Drive, Westbrook, ME 04092, USA

[1]Corresponding author email: estradaa@ufl.edu

Objectives: **Evaluation of a computerised electrocardiogram algorithm compared to the interpretation of a team of board-certified veterinary cardiologists.**

Materials and Methods: **This was a cross-sectional retrospective cohort study. A total of 399 electronic canine electrocardiogram recordings screened from 1391 electrocardiograms were enrolled in the study. A panel of seven cardiologists, masked to patient information, evaluated electrocardiograms for the following: P-wave amplitude and duration; PR-interval; R-wave amplitude; QRS duration; heart rate; mean electrical axis; and final overall diagnosis for the detection of arrhythmia and any abnormal electrocardiogram anomaly.**

Results: **The sensitivity of the electrocardiogram algorithm for detecting arrhythmias was 99.7% (95% confidence intervals, CI: 98.5 to 99.9) and the specificity was 99.5% (95% CI: 98.0 to 99.9) compared to the consensus result created by panel of cardiologists. The sensitivity of the algorithm for the detection of any electrocardiogram anomaly, including abnormal measurements, was 71.3% (95% CI: 65.5 to 76.7) and the specificity was 35.1% (95% CI: 27.0 to 43.8) compared to the panel of cardiologists.**

Clinical Significance: **The electrocardiogram algorithm was shown to have high sensitivity for the detection of arrhythmias, but not all electrocardiogram anomalies. The results support the use of this algorithm as a tool to aid in the triage of the electrocardiogram workflow.**

Journal of Small Animal Practice (2020) **62,** 145–149
DOI: 10.1111/jsap.13267

Accepted: 26 October 2020; Published online: 01 December 2020

INTRODUCTION

Electrocardiograms (ECGs) in veterinary medicine are performed for a variety of indications: as a pre-operative screening tool, to diagnose arrhythmias noted on auscultation, as part of a full cardiac workup for patients with known or suspected heart disease, as part of screening for arrhythmias during chemotherapy or as a baseline for breeds at risk of developing clinically relevant arrhythmias (Boxers, Doberman pinchers).

Over the two last decades, there has been an increase in the volume of visits for dogs presented to a veterinarian (AVMA 2018). A 2007 American Veterinary Medical Association (AVMA) survey of USA pet owning households estimated 119.4 million veterinary visits for dogs in 2006 (Shepherd 2008). If these patients require consultation with a cardiologist, this can prove difficult as there are only 361 board certified cardiologists (ACVIM and ECVIM), making direct access to specialists somewhat limited. Teleconsulting offers an opportunity to increase access to board certified cardiologists and in turn improve the quality of care delivered to patients by the family veterinarian.

There have been many challenges in the implementation of teleconsulting, including cost, quality of equipment and reliable connectivity (Kahn 2015). Advances in technology have eliminated many of those hurdles. The most important challenge for teleconsulting in veterinary medicine is professional acceptance and understanding of the role this tool may play in day-to-day practice. In 2016, AVMA created a telemedicine advisory panel, highlighting the emergence of using teleconsulting in the veterinary-client-patient relationship ("Telemedicine: Report of the AVMA Practice Advisory Panel (2016)"). The use of computer aided ECG analysis can be applied to teleconsulting. In addition to increased access to care, the use of automated computer-aided analysis of ECGs offers advantages in exact repeatability of measurement, lack of observer error (*e.g.* observer fatigue, lapse of attention), and an increase in efficiency and productivity compared to solely manual methods (Moody *et al.* 2006).

For telemedicine devices like automated computer aided ECG analysis to gain acceptance, continued validation of computer-aided ECG algorithms is needed. The CardioPet ECG Algorithm is a decision support technology that aides in the interpretation of ECGs by cardiologists and is designed to accurately measure the waveforms of an ECG complex. The aim of this study was to validate the performance of the ECG Algorithm against a panel of board-certified veterinary cardiologists on the measurement of seven ECG parameters: heart rate, P-wave amplitude, R-wave amplitude, P-wave duration, QRS-interval duration, PR-interval duration, mean electrical axis (MEA) axis; and to assess whether this ECG Algorithm could correctly classify ECGs as normal or abnormal as defined by a panel of board-certified cardiologists.

MATERIALS AND METHODS

Study design

A cross sectional, retrospective, cohort design was used in the study. ECGs were obtained from the IDEXX Vetmedstat System during the month of March 2017. Only canine ECGs that passed a signal to noise ratio filter were considered qualified and eligible for review. A total of 1391 ECGs were screened, of which 15.9% (221/1391) were excluded because they were feline. Out of the remaining 1170 canine ECGs, 12.5% (146/1170) were not qualified, leaving 1024 ECGs eligible for review (Fig 1). Of the 1024 ECGs that met our criteria, a total of 399 ECGs were chosen for this study based upon pairing to pair up each of our cardiologists three times and wanting each ECG file to be read in triplicate. Each ECG was evaluated separately by three cardiologists, who were all masked to the results of the ECG algorithm. In total, each of the seven cardiologists evaluated 171 of the 399 ECGs.

The 399 cases were randomly assigned to three cardiologists out of a panel of seven board-certified cardiologists, such that each of the cardiologists were paired with each other for three different cases (Fig 2). Both the ECG algorithm and the panel of cardiologists evaluated the following characteristics of each case: heart rate, P-wave amplitude, R-wave amplitude, P-wave duration, QRS-interval duration, PR-interval duration, MEA axis. For the waveform measurements, the reference method was defined as the mean of the cardiologist's measurements. The ECG algorithm and the cardiologists also rendered a normal or abnormal result for each ECG file. The ECG result could be deemed abnormal for an abnormal waveform parameter or for any arrhythmias. The final decision by the cardiologists for the detection of an arrhythmia or other abnormal ECG result were defined as at least two of the three cardiologists reviewing the ECG file agreeing on the ECG result. Throughout the rest of this manuscript, the grouped agreement shall be called the "majority result."

Statistical analysis

Bland–Altman plots, with mean difference, 95% confidence intervals for the mean difference and limits of agreement (1.96*sd) were reported. The mean difference was calculated using the mean of the three cardiologist's ECG measurements (as the reference) for each waveform subtracted from the algorithm's measurement. Measurement of MEA was conducted with several different techniques at the preference of the cardiologist. The techniques utilised for evaluation of the MEA by the cardiologists include the lead graphing method, the isoelectric lead method and the tallest R wave method. Sensitivity and specificity analyses were calculated using the majority result as the reference method to compare against the ECG algorithm's classification of arrhythmias and all ECG anomalies (including arrhythmias).

RESULTS

The top 5 breeds represented in the study population were Yorkshire terrier (n=21), Chihuahua (n=19), dachshund (n=19), Maltese (n=18), and shih-tzu (n=16). Mixed breed dogs made up 21.1% (n=84) of the study population. The median age for dogs was 8 years (range: 3 months to 10 years). Weight was available for 383 dogs (96.0%) with a median weight of 11.1 kg (range: 1.5 to 61.3 kg).

Bland–Altman plots for all seven of the ECG parameters, along with the mean difference, 95% confidence interval around mean difference, and limits of agreement can be found in Fig 2. The mean difference for the seven ECG parameters can also be found in Table 1. For the Bland–Altman plot of MEA axis, a spaced interval for the measurements among the readers was observed due to rounding to the nearest tenth place during manual measurement (Fig 3A).

Sensitivity of the algorithm for the detection of arrhythmias and all ECG anomalies

The sensitivity of the ECG algorithm for detecting arrhythmias was 99.7% (95% confidence intervals, CI: 98.5 to 99.9) and the

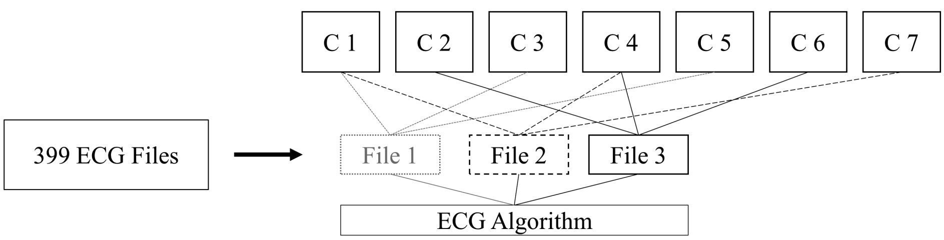

FIG 1. Consort Diagram for ECG cases eligible to be enrolled in the study

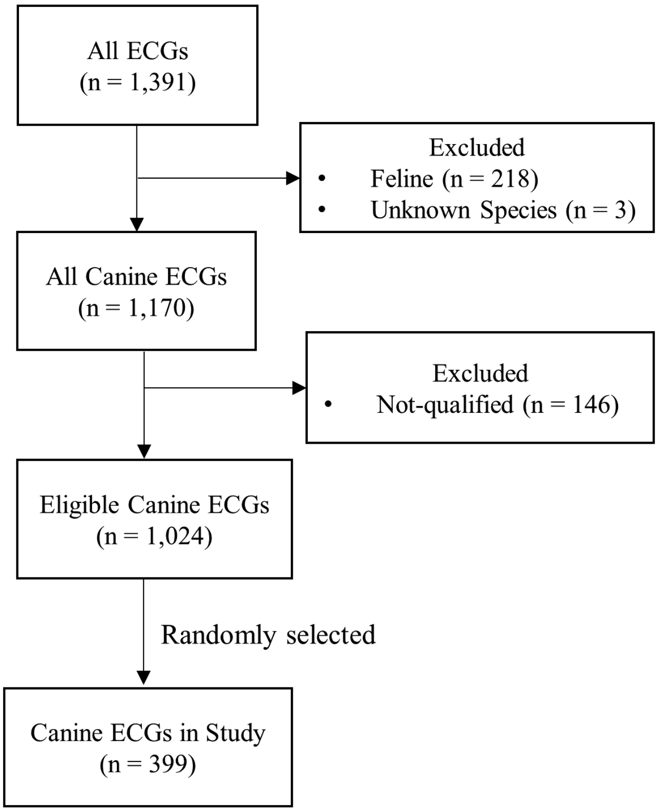

FIG 2. Study Design for ECG Evaluation. Each of the 399 ECG files was evaluated by 3 randomly selected cardiologists out of a panel of seven cardiologists (C1-7). For example, File 1 (in dotted grey) was randomly assigned to cardiologist reviewers 1, 3 and 5 for evaluation and was concurrently evaluated by the ECG algorithm. File 2 (in dotted black) was randomly assigned to cardiologist reviewers 1, 4 and 7 for evaluation and was concurrently evaluated by the ECG algorithm. Every ECG file was read by the ECG Algorithm

Table 1. Interclass correlation of ECG parameters measured by seven cardiologist and ECG algorithm

ECG parameter	Between cardiologist	Algorithm *versus* cardiologist
	ICC (95% CI)	ICC (95% CI)
MEA	0.96 (0.87 to 0.98)	0.96 (0.85 to 0.98)
QRS	0.97 (0.93 to 0.98)	0.98 (0.95 to 0.99)
PR-Interval	0.99 (0.95 to 1.00)	0.98 (0.98 to 0.99)
P-Width	0.96 (0.95 to 0.96)	0.82 (0.79 to 0.84)
Heart Rate	0.99 (0.99 to 1.00)	0.97 (0.93 to 0.99)
P Height	0.98 (0.96 to 1.00)	0.97 (0.79 to 0.98)
R Height	0.99 (0.97 to 1.00)	0.95 (0.92 to 0.97)

All P-values found to be >0.995, for Yuen-TOST test for statistical equivalence. Equivalency bounds can be found in Table S1. CI: confidence intervals.

specificity was 99.5% (95% CI: 98.0 to 99.9) compared to the majority result (Table 2). The ECG algorithm identified two false positives and one false negative detection of arrhythmias. Upon further investigation, the false negative finding was a sinus pause at the very end of an ECG tracing, which was outside the range of timing in which the algorithm measures. In addition, two sinus arrhythmias were incorrectly identified as pathologic arrhythmias. A full list of the types of arrhythmias in this study can be found in Table S1. The sensitivity of the algorithm for the detection of any ECG anomaly, including abnormal measurements, was 71.3% (95% CI: 65.5 to 76.7) and the specificity was 35.1% (95% CI: 27.0 to 43.8) compared to the majority result (Table 3).

Exploratory discordant ECG analysis

Due to the high level of discordance for the detection of all ECG abnormalities between the majority result and ECG algorithm, an exploratory secondary analysis was conducted to adjudicate the false positives and false negatives.

Out of the 87 false positives, the secondary analysis found that 53 were truly positive and 34 were true false positives. The reasons for the false positive results were as follows: false positive due to the algorithm seeing artefact in the tracing and incorrectly labelling this as an arrhythmia or abnormal beat (24 cases); the reference beat generated to determine measurements had an incorrect assessment of a notched P-wave (three); the reference beat generated to determine measurements had an averaged wider P-wave than any one ECG complex (two); the algorithm incorrectly assigned a small dog or toy configuration parameters to a larger dog (two); the MEA was assessed incorrectly (one); a sinus arrhythmia was incorrectly flagged for a true arrhythmia (two).

Of the 76 false negatives, the secondary analysis found that 62 were truly positives and 14 were true false negatives. The reasons for the false negatives were as follows: the algorithm incorrectly assigned large breed dog parameters to small or toy breeds (eight); cases had notched P-waves that the algorithm did not assess as notched (two); the tolerance assigned for an abnormal MEA is 4° and therefore it did not assign an abnormal result to a case which the three cardiologist deemed was an abnormal MEA (one); did not measure the QRS complex correctly (one); did not measure the P-wave correctly (one); and missed one sinus pause at the very end of an ECG tracing and was deemed a true missed assessment of an arrhythmia (one).

DISCUSSION

The objective of this study was to compare the effectiveness of an ECG algorithm in evaluating an ECG to the performance of a team of board-certified cardiologists. One of the most important functions of the ECG algorithm is the detection of arrhythmias. When comparing the ability to identify arrhythmias, the ECG algorithm had high agreement with the team of cardiologist manually reviewing the ECGs, showing both a high sensitivity and high specificity. The ECG algorithm missed one arrhythmia. Upon further investigation, this arrhythmia was a sinus pause at the very end of an ECG tracing, which was outside the range of timing in which the algorithm measures. In addition, two sinus arrhythmias were incorrectly identified as pathologic arrhythmias. Even with high sensitivity and specificity, further improvements of the ECG algorithm are needed to ensure that there are fewer misclassifications of arrhythmias.

The ECG algorithm the algorithm showed poor sensitivity and specificity for labelling an ECG as "abnormal." With this specific algorithm, an ECG can be graded "abnormal" due to the presence

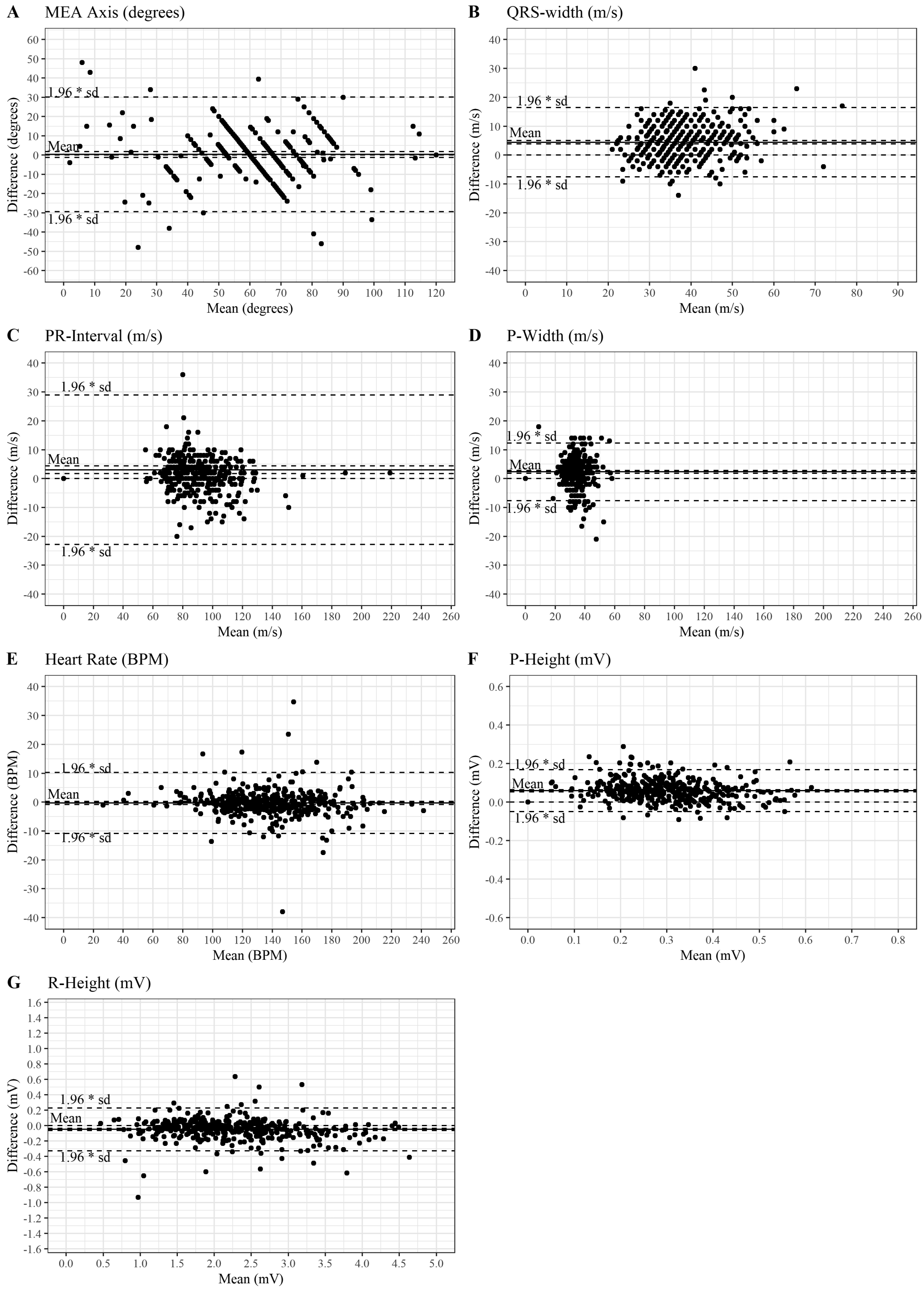

FIG 3. Bland–Altman plots for the seven ECG waveforms. The mean difference, 95% confidence interval around mean difference, and limits of agreement (defined as 1.96 * sd of the difference) are shown. The mean difference was calculated using the mean of the three cardiologist's ECG measurements (as the reference) for each waveform subtracted from the algorithm's measurement

of an arrhythmia or based upon wave measurements detected outside of established intervals. A possible reason for this discrepancy could be that the unmasked cardiologist had available the waveform measurements from the ECG algorithm, which takes the average of multiple complexes, thus obtaining a better representation of the overall ECG than a cardiologist who measures a single complex. For example, the algorithm may identify an ECG with small or isoelectric P-waves to be abnormal because of the parameters initially set to determine normal *versus* abnormal P-wave height identification and classification. Additionally, the

Table 2. Agreement between consensus cardiologist result and ECG algorithm for the detection of an arrhythmia

		Consensus cardiologist results (for the detection of an arrhythmia)		
		Abnormal	Normal	Total
ECG algorithm	Abnormal	31 (7.8%)	2 (0.5%)	33 (8.3%)
	Normal	1 (0.3%)	365 (91.5%)	366 (91.7%)
	Total	32 (8.0%)	367 (92.0%)	399

Table 3. Agreement between consensus cardiologist result and ECG algorithm for the detection of any ECG anomaly

		Consensus cardiologist results (for the detection of any ECG anomaly)		
		Abnormal	Normal	Total
ECG algorithm	Abnormal	189 (47.4%)	87 (21.8%)	276 (69.2%)
	Normal	76 (19.0%)	47 (11.8%)	123 (30.8%)
	Total	265 (66.4%)	134 (33.6%)	399

algorithm follows the textbook definition of what an abnormal waveform to the hundredth of a millisecond, while a cardiologist manually screens the complex (Burke *et al.* 2014). Thus, a result could technically meet the textbook definition of an abnormal results to the hundredth of a millisecond that could have been overlooked by the cardiologist. The authors' opinion is that over-interpretation of abnormal ECGs which are subsequently flagged and sent to a cardiologist for manual review is preferred to under-interpretation and possibly missing an important arrhythmia for a patient. However, these abnormalities do not carry equivalent medical relevance. Arrhythmias identify functional abnormalities and can represent potentially serious underlying cardiac disease; whereas, ECG measurements alone are not considered reliable to detect underlying cardiac disease. A study evaluating the P-wave width to detect underlying left atrial enlargement found this measurement to be unreliable in the detection of underlying left atrial enlargement (Savarino *et al.* 2012). Therefore, the sensitivity for detection of abnormal measurements correctly identifying functional abnormalities is not considered as medically relevant when evaluating the performance of the ECG algorithm.

There are a number of limitations with this study which should be considered. The population used in this study only includes dogs that received an ECG from a single telemedicine service which could lead to sampling bias. Repeatability and reliability of both the cardiologists and the algorithm were not measured. With the team of cardiologist being allowed to use their preferred method for measuring MEA axis, this could have contributed to some of the discrepancy noted. The adjudication panel only reviewed the false positives and false negatives, thus there in an imbalance in the secondary review of the ECG results. Due to this, minimal inferences can be made from the adjudicated results. In addition, there were a total of 31 arrhythmias identified in a total of 399 ECGs of which only a single supraventricular tachycardia was found and no ventricular ectopy, or ventricular tachycardia were present. With a small number of events, it was difficult to draw conclusions on the accuracy of the ECG algorithm for identifying arrhythmias.

The authors are optimistic that further studies will continue to address challenges related to computer aided ECG analysis. The next series of studies should seek to provide more insight into the utilisation of computer-aided algorithms in ECG diagnosis. The variety of breeds, dog size and body composition are all factors that could influence the ECG and its interpretation. Additional studies that aid in the identification of normal parameters for these different characteristics will help increase the accuracy and benefit of computer aided ECG analysis. Further studies should also attempt to quantify the benefit of the utilisation of this tool. The quantification could be in tangible and intangible benefits.

The ECG algorithm can be used as a clinical decision support system to aid in ECG analysis for a patient. We believe the use of automated systems for the measurement of ECG parameters can, over time, improve the accuracy and reliability of ECG interpretation. While a cardiologist may only measure one complex and determine the ECG measurements, the algorithm can average the beats of multiple complexes to determine the average measurement for each ECG parameter. The averaging the each of the complexes allows for a more representative evaluation of the ECG waveforms.

Computer-aided ECG analysis is a valuable tool that employed in a manner that is consistent with guidelines established by the AVMA advisory panel on telemedicine, can improve the practice of veterinary medicine ("Telemedicine: Report of the AVMA Practice Advisory Panel (2016)"). The authors of this study do not support the use of an ECG algorithm as a definitive diagnosis tool, but rather as a tool to aid in the triage of the ECG workflow. The computer-aided ECG algorithm is used as a tool to aid the practitioner in this manner and facilitate appropriate cardiology consultation. As teleconsulting continues to grow and advance in its application to human medicine, it is not a question of whether teleconsulting will be utilised in veterinary medicine, but rather when it will be fully incorporated in standardised practice.

Conflict of interest

All the authors except Dr Estrada are employed by IDEXX.

References

AVMA. (2018) 2017-2018 Pet Ownership and Demographics Sourcebook. *American Veterinary Medical Association*. https://www.avma.org/resources-tools/reports-statistics. Accessed April 28, 2020

Burke, G. M., Wang, N., Blease, S., *et al*. (2014) Assessment of reproducibility – automated and digital caliper ECG measurement in the Framingham heart study. *Journal of Electrocardiology* **47**, 288-293

Kahn, J. M. (2015) Virtual visits – confronting the challenges of telemedicine. *New England Journal of Medicine* **372**, 1684-1685

Moody, G., Koch, H. & Steinhoff, U. (2006) The PhysioNet/Computers in Cardiology Challenge 2006: QT Interval Measurement. 4

Savarino, P., Borgarelli, M., Tarducci, A., *et al*. (2012) Diagnostic performance of P wave duration in the identification of left atrial enlargement in dogs. *Journal of Small Animal Practice* **53**, 267-272

Shepherd, A. J. (2008) Results of the 2007 AVMA survey of US pet-owning households regarding use of veterinary services and expenditures. *Journal of the American Veterinary Medical Association* **233**, 727-728

Telemedicine: Report of the AVMA Practice Advisory Panel. (2016) https://www.avma.org/KB/Resources/Reports/Pages/Telemedicine-Report-AVMA-Practice-Advisory-Panel.aspx. Accessed 6 August 6, 2018

Supporting Information

The following supporting information is available for this article:

Table S1. Types of Arrhythmias found (an ECG file can have more than one type).