ORIGINAL ARTICLE

Reliability and Validity of Clinician ECG Interpretation for Athletes

Charles Magee, M.D., M.P.H.,* Joshua Kazman, M.S.,† Mark Haigney, M.D.,* Ralph Oriscello, M.D.,‡ Kent J. DeZee, M.D., M.P.H.,* Patricia Deuster, Ph.D., M.P.H.,† Patrick Depenbrock, M.D.,§ and Francis G. O'Connor, M.D., M.P.H.†

From the *Department of Medicine, Uniformed Services University of the Health Sciences, Bethesda, MD; †Department of Military and Emergency Medicine, Uniformed Services University of the Health Sciences, Bethesda, MD; ‡Cardiology Division, East Orange Veterans Affairs Hospital, East Orange, NJ; and §Family Medicine Clinic, Fort Belvoir Community Hospital, Fort Belvoir, VA

Background: Electrocardiogram (ECG) with preparticipation evaluation (PPE) for athletes remains controversial in the United States and diagnostic accuracy of clinician ECG interpretation is unclear. This study aimed to assess reliability and validity of clinician ECG interpretation using expert-validated ECGs according to the 2010 European Society of Cardiology (ESC) interpretation criteria.

Methods: This is a blinded, prospective study of diagnostic accuracy of clinician ECG interpretation. Anonymized ECGs were validated for normal and abnormal patterns by blinded expert interpreters according to the ESC interpretation criteria from October 2011 through March 2012. Six pairs of clinician interpreters were recruited from relevant clinical specialties in an academic medical center in March 2012. Each clinician interpreted 85 ECGs according to the ESC interpretation guidelines. Cohen and Fleiss' kappa, sensitivity, and specificity were calculated within specialties and across primary care and cardiology specialty groups.

Results: Experts interpreted 189 ECGs yielding a kappa of 0.63, demonstrating "substantial" inter-rater agreement. A total of 85 validated ECGs, including 26 abnormalities, were selected for clinician interpretation. The kappa across cardiology specialists was "substantial" and "moderate" across primary care (0.69 vs 0.52, respectively, $P < 0.001$). Sensitivity and specificity to detect abnormal patterns were similar between cardiology and primary care groups (sensitivity 93.3% vs 81.3%, respectively, $P = 0.31$; specificity 88.8% vs 89.8%, respectively, $P = 0.91$).

Conclusions: Clinician ECG interpretation according to the ESC interpretation criteria appears to demonstrate limited reliability and validity. Before widespread adoption of ECG for PPE of U.S. athletes, further research of training focused on improved reliability and validity of clinician ECG interpretation is warranted.

Ann Noninvasive Electrocardiol 2014;19(4):319–329

ECG screening; preparticipation screening; athlete ECG

Sudden cardiac death (SCD) is a tragic occurrence in athletes. Although rare, SCD accounts for more deaths in athletes than asthma, trauma, or any nonaccidental cause.^{1–4} The reported overall incidence of SCD in athletes within the United

States has remained stable at nearly 1 per 100,000 athlete-years. However, it varies considerably by intensity of athletic activity, and is reported as high as 32 per 100,000 athlete-years in division 1 National Collegiate Athletic Association (NCAA)

Address for correspondence: Charles Magee, M.D., M.P.H., Department of Medicine, Uniformed Services University of the Health Sciences, 4301 Jones Bridge Road, Bethesda, MD 20814. Fax: (301) 295-5792; E-mail: charles.magee@usuhs.edu

Financial Disclosure: There are no financial disclosures or conflicts of interest to report.

Disclaimer: The opinions expressed herein are those of the author(s), and are not necessarily representative of those of the Uniformed Services University of the Health Sciences (USUHS), the Department of Defense (DOD); or, the United States Army, Navy, or Air Force.

basketball players.^{3,5-8} The current standard of care for screening U.S. athletes for SCD includes a history and physical examination, although many have argued that electrocardiogram (ECG) screening may have merit in mitigating risk of SCD.⁹⁻¹⁶ Presently, the American Heart Association does not recommend ECG in preparticipation evaluations (PPE) of U.S. athletes to enhance screening for SCD, citing low SCD prevalence, false-positives, significant cost, and potential risks generated through false positive tests and risks associated with additional diagnostic testing and treatments.¹⁷

In 2010, the European Society of Cardiology (ESC) published a consensus statement for cardiologists and sports medicine physicians recommending compulsory ECG interpretation with PPE of athletes.⁹ The statement provides a framework to dichotomize ECG patterns related to benign physiologic adaptations, often described as "*The Athlete's Heart*,"¹⁸⁻²⁰ and those suggestive of underlying cardiac disease requiring further clinical investigation. The guideline stems from over 25 years of observational data trending a reduction in the incidence of SCD from 3.6/100,000 down to 0.4/100,000 athlete-years within Italian athlete populations screened with PPE that included an ECG.¹⁰ Recently, the impact of adding ECG to PPE of athletes in Israel failed to demonstrate a reduction in SCD incidence.²¹ A cross-sectional study of screening with and without ECG in NCAA athletes reported an increase in PPE sensitivity from 45% to over 90% (confidence interval 58.7% to 99.8%), at the expense of generating nearly 17% false positives.²²

If the United States were to consider implementing ECG as part of the PPE as recently recommended in the "Seattle Criteria,"²³ understanding the reliability and validity of clinician ECG interpretations would be necessary before widespread adoption. A recent cohort study by Hill et al. found the sensitivity and specificity of pediatric cardiologists in 2010 to be 68% and 70%, respectively, when asked to detect abnormal patterns in ECGs for PPE of athletes.²⁴ Drezner et al. demonstrated improved accuracy of ECG interpretation following an educational intervention to train U.S. clinicians to use the 2010 ESC interpretation guidelines: this resulted in an overall sensitivity of 91% and specificity of 94% for identifying abnormal patterns.²⁵ However, the study design is limited by retest bias and

it did not assess inter-rater reliability. To date, no study has addressed reliability, a necessary component of diagnostic accuracy according to the *Standards for the Reporting of Diagnostic Accuracy Studies* (STARD),²⁶ of clinician ECG interpretation. Furthermore, the validity of reference-standard expert ECG interpretation has not been appropriately characterized according to the 2010 ESC interpretation guidelines for PPE of athletes. Therefore, the primary objective of this study was to assess reliability and validity of clinician ECG interpretation across relevant clinical specialties. For the reference standard, ECG patterns were validated using expert interpretations according to the 2010 ESC interpretation criteria.

METHODS

This was a prospective, blinded study of clinician ECG interpretation compared to expert interpretation according to the 2010 ESC interpretation guidelines for PPE of athletes. The 2010 ESC interpretation guidelines and criteria have been reported previously⁹ and summarized concisely by Drezner et al.²⁵ (Appendix). Following validation by expert interpreters, clinicians across six specialties performed ECG interpretations. This study was designed and reported in accordance with the STARD statement.

ECG Set Derivation

To obtain validated ECGs, normal ECGs were selected retrospectively from a patient database, and abnormal ECGs were obtained from an academic teaching file. Two expert raters interpreted the ECGs and those with perfect pattern agreement were selected for the ECG set, as detailed below.

ECG Selection

Standard 12-lead ECGs were obtained from the East Orange VA hospital in East Orange, NJ, Department of Cardiology; originating from patients presenting for routine and referral indications from 2009–2010. ECGs included patients ranging between 20 and 35 years of age, with 20 years representing the youngest age available and 35 years as a cutoff to minimize risk of atherosclerotic disease manifestations on ECGs. Preliminary normal ECGs were identified from

an electronic database by the following terms: "Within normal limits"; "Normal tracing"; "Sinus tachycardia, otherwise normal"; "sinus bradycardia, otherwise normal." Preliminary abnormal ECGs were derived from an academic teaching file of clinically confirmed cardiac disease, identified by the presence of the following patterns: hypertrophic cardiomyopathy (HCM), Brugada, complete right and left bundle branch block, Wolff-Parkinson-White (WPW) syndrome, long QT (LQT) and short QT (SQT), Arrhythmogenic Right Ventricular Cardiomyopathy (ARVC), and others. Each ECG was anonymized prior to study use. Information relating to age, gender, height, weight, test indication, machine-generated intervals, axis, and interpretations was also removed from the ECGs to reduce interpretation bias.

Expert Interpretation

Experts consisted of two independent, blinded cardiologists, each with active board certification in cardiology and either subspecialty fellowship training in cardiac electrophysiology or certification of special qualifications in cardiac electrocardiography. Each expert independently interpreted the same ECGs according to the ESC 2010 interpretation guidelines until perfect pattern agreement was achieved on a sufficient number of ECGs with respect to specific abnormal pattern(s) or absence of abnormal patterns to create a validated set for clinician interpretations. ECGs demonstrating perfect expert agreement for abnormal patterns of cardiac diseases most commonly associated with SCD in U.S. athletes were preferentially selected for use in the validated set of ECGs. ECGs demonstrating expert agreement for absence of abnormal patterns were randomly selected for use in the validated set of ECGs.

Clinician Interpretation

Clinicians were recruited from clinical practice within a single academic medical center, different than the source of the ECGs, from the following specialties: Family Medicine (FM), Internal Medicine (IM), Pediatrics (Ped), Sports Medicine (SM), Cardiology (Card), and Pediatric Cardiology (PCard). Eligibility requirements included active board certification and clinical practice (>20 hours/week for the preceding 6 months). Participants received no compensation

and remained blind to expert interpretation. All participants provided written informed consent at study enrollment. Each participant received educational materials on the 2010 ESC guidelines in the form of a two-page reference guide (Appendix). Participants were instructed to interpret ECGs according to the 2010 ESC guidelines within a two-week period and offered ECG calipers to assist in interpretation.

Statistical Analysis

ECG Derivation

Inter-rater agreement of expert ECG interpretation is reported using Cohen's kappa according to the detection of any abnormal pattern(s) for all interpreted ECGs.²⁷

Clinician Interpretation

Reliability and validity of clinician interpretation were assessed using expert ECG interpretations as reference standard. Clinician data were analyzed according to individual specialty and specialty groups (primary care: FM, IM, Ped, SM; Cardiology: Card, PCard). Inter-rater reliability was computed using Cohen's and Fleiss' kappa,^{27,28} including 95% confidence intervals. Chi-square tests were calculated to identify statistically significant differences from a null kappa of 0.4, selected as a minimal level of clinically relevant inter-rater agreement. Validity was assessed by calculating sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for each specialty and specialty group by using expert interpretation as the reference standard for identifying both the presence of abnormal pattern(s) as well as correctly identifying specific abnormal pattern(s). 95% confidence intervals were determined and Wilcoxon rank-sums tests performed to identify statistically significant differences between specialty groups for sensitivity and specificity. For all tests, statistical significance was set at $P = 0.05$.

The study was powered for analysis of inter-rater agreement in the *clinician-interpretation step*. The number of ECGs for interpretation was determined using an ECG abnormality prevalence of 30% based on previous reports^{11,29} from well-trained clinician interpreters to mitigate prevalence bias in the kappa statistic. A set of at least 85 ECGs with 59 normal ECGs and 26 abnormal ECGs

Figure 1. Flow diagram of expert ECG interpretation. Blinded experts independently interpreted 189 anonymized ECGs. Each expert interpreted all ECGs as either normal or abnormal. ECGs were classified according to expert detection of abnormal patterns according to the 2010 ESC interpretation criteria.

was necessary to identify a clinically significant difference between a kappa of 0.70 (representing “substantial” inter-rater agreement) and a null kappa of 0.40 (representing “fair” inter-rater agreement) with alpha set at 0.05 and 80% power.³⁰ All analyses were performed using SAS v9.3 (SAS Institute, 2011, Cary, NC, USA). This study was approved by the Institutional Review Boards of the Walter Reed National Military Medical Center and Uniformed Services University of the Health Sciences in Bethesda, MD. No external funding was provided. The principal investigator had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

RESULTS
ECG Set Derivation

Each expert independently interpreted 189 ECGs from October 2011 through March 2012 (Fig. 1).

Table 1. Expert Interpretation for Classification of ECGs by the Presence of Abnormal Patterns

Expert #2	Expert #1		
	Abnormal	Normal	
Abnormal	40	15	55
Normal	14	120	134
	54	135	189
Observed agreement	84.66%		
Kappa	0.63		“substantial” ^a

^a“Substantial” agreement based on Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159–174. All interpretations according to 2010 ESC interpretation guidelines.

Experts agreed 40 ECGs contained abnormal and 120 normal pattern(s) according to ESC 2010 criteria (Table 1). Experts disagreed regarding the presence of abnormal patterns in the remaining 29 ECGs. Observed agreement between experts was 84.7% with a kappa of 0.63 (SE 0.064, 95%

Table 2. Expert Agreement on Specific Abnormal Patterns When Each Classified the ECG as Abnormal

Abnormal Pattern ^a	Expert 1 (agreement/ total interpreted)	Expert 2 (agreement/ total interpreted)
T-wave inversion	7/10	7/12
ST depression	1/5	1/2
Pathologic Q waves	4/4	4/9
LBBB	1/1	1/1
RBBB	4/4	4/4
LAE	6/13	6/11
LAD	7/8	7/8
RVH	5/5	5/6
Mobitz 2 2AVB	1/1	1/1
3AVB	1/1	1/1
Preexcitation	4/4	4/5
LQT interval	2/3	2/3
SQT interval	0/2	0/1
Brugada pattern	2/3	2/5
Epsilon wave	0/3	0/0
Atrial tach	1/8	1/2
PVCs	0/1	0/0
HCM-related (LAE, LAD, ST depression, T-wave inversion, path Q waves)	25/40	25/42
Overall	46/76	46/71

LBBB = complete left bundle branch block; RBBB = complete right bundle branch block; LAE = left atrial enlargement; LAD = left axis deviation; RVH = right ventricular hypertrophy; Mobitz 2 2AVB = Mobitz Type II 2° atrioventricular block; 3AVB = 3° atrioventricular block; LQT interval = long-QT interval; SQT interval = short-QT interval; PVCs = 2 or more premature ventricular contractions; HCM = hypertrophic cardiomyopathy.

^aAbnormal pattern indicates that the expert identified the pattern was present in the ECG.
All interpretations according to 2010 ESC interpretation guidelines.

CI 0.50 to 0.75; $K_{\max} = 0.987$) for detecting the presence and absence of abnormal pattern(s); this demonstrated "substantial" agreement.²⁷ In 37 of the 40 ECGs that experts agreed contained abnormal patterns, there was agreement on 46 specific patterns (Table 2). In the remaining 3 ECGs, experts disagreed on the specific abnormal patterns. Of the 29 ECGs where experts disagreed on the presence of abnormal patterns, 16 included abnormal patterns associated with HCM: left atrial enlargement (N = 7), T wave inversion (N = 5), and pathologic Q-waves (N = 4) (Table 3).

ECGs selected for *clinician interpretation* included 59 normal ECGs selected at random from among the 120 expert-validated normal ECGs and 26 abnormal ECGs selected from the 40 expert-validated abnormal ECGs for a total of 85 ECGs. The selected abnormal ECGs demonstrated the following patterns (3 ECGs contained more than 1 abnormal pattern): T wave inversions (5), left atrial enlargement (3), left axis deviation (3), preexcitation pattern (e.g. WPW) (3), LQT

Table 3. Expert Disagreement of Abnormal Pattern Interpretations

14 ECGs with 14 Abnormal Patterns Identified Only by Expert #1:	15 ECGs with 16 Abnormal Patterns Identified Only by Expert #2:
TWI (2)	TWI (3)
LAE (3)	LAE (4)
LQT interval (1)	LQT interval (1)
SQT interval (2)	SQT interval (1)
Brugada pattern (1)	Brugada pattern (2)
Atrial tachyarrhythmia (2)	Atrial tachyarrhythmia (1)
Epsilon wave (3)	Pathologic Q waves (4)

TWI = T-wave inversion; LAE = left atrial enlargement; LQT interval = long-QT interval; SQT interval = short-QT interval.
All interpretations according to 2010 ESC interpretation guidelines.

intervals (2), Brugada patterns (2), pathologic Q waves (2), complete right bundle branch block (2), right ventricular hypertrophy (2), ST-segment depressions (1), complete left bundle branch block (1), Mobitz type-2 atrioventricular block (1),

third-degree heart block or complete atrioventricular block (1), atrial tachyarrhythmia (1).

Clinician Characteristics

Clinician recruitment was begun and completed in March, 2012. The first two board-certified clinicians from each specialty agreed to participate following informed consent and all provided an interpretation for every study ECG. The mean age for clinicians was 39.5 years in both primary care and cardiology specialty groups. All cardiology specialists were male as well as 62.5% of primary care physicians. All physicians were graduates of U.S. medical schools and credentialed to interpret ECGs. Specialists reported a mean of 5.5 years in their current specialty and primary care physicians reported a mean of 9.4 years which was not a statistically significant difference (Independent T-test, two-sided, $P = 0.20$). The mean number of ECGs interpreted per year was 195 (range = 5–1200; SD = 407.7) for primary care physicians and 1050 (range = 1000–1200; SD = 100) for cardiology specialty physicians (Independent T-test, two-sided, $P = 0.0004$). All primary care physicians reported no formal certification in ECG interpretation and all cardiology specialists provided board certification as formal certification.

Clinician Reliability

Kappa was calculated between each individual clinician and the expert interpretation for the detection of any abnormal pattern(s) (Table 4, top). The mean kappa across all study participants was 0.72, ranging from 0.45 for one pediatrician to 0.89 for one pediatric cardiologist (95% CI 0.65, 0.80; SD = 0.11). Eight individual clinicians achieved kappa of ≥ 0.70 with expert interpretations and P-values were < 0.05 for all comparisons against a null kappa of 0.40, except for a single pediatrician with a kappa of 0.45, which was not significantly greater than the null.

Within specialty kappa was calculated to assess inter-rater agreement for the detection of abnormal ECG pattern(s) (Table 4, bottom). The mean kappa across all specialties was 0.55, ranging from 0.37 within pediatrics to 0.75 within cardiology (95% CI 0.40, 0.70; SD = 0.14). Within-specialty kappas were significantly greater than the null within all disciplines except Pediatrics and Sports Medicine.

Fleiss' kappas for inter-rater agreement on the presence of abnormal ECG patterns within primary care and cardiology specialty groups were 0.52 and 0.69, respectively ($P < 0.001$, chi-square; Table 5).

Clinician Validity

Sensitivity and specificity to detect an abnormal pattern(s) were calculated for each specialty and specialty group (Table 6, top). Sensitivities ranged from 65.4% in pediatrics to 96.2% in cardiology. Specificity ranged from 86.4% in sports medicine to 93.2% in pediatric cardiology. Sensitivity and specificity for primary care were 81.3% and 88.8%, respectively. Sensitivity and specificity for cardiology specialties were 93.3% and 89.8% respectively. No statistically significant difference in sensitivity or specificity was noted with regard to identifying the presence of abnormal patterns between primary care and cardiology specialties. The PPV and NPV were 76.1% and 91.5% for primary care and 80.2% and 96.8% for cardiology specialties, respectively.

The sensitivity to *correctly identify* at least one abnormal pattern according to expert interpretation was calculated by specialty and specialty groups (Table 6, bottom). Sensitivities ranged from 42.3% in pediatrics to 84.6% in pediatric cardiology, with 60.1% and 83.7% for primary care and cardiology specialty groups, respectively. A statistically significant difference (Wilcoxon rank-sum test, two-sided exact, $P = 0.0485$) was noted by specialty. The positive and negative predictive value of primary care specialties were 70.2% and 83.5% whereas they were 78.4% and 92.6%, respectively, for cardiology specialties. The positive and negative likelihood ratios for ECG interpretation, that is, measures employing sensitivity and specificity to determine whether the test result effectually alters the pretest likelihood based on the calculated probability that an underlying pattern truly exists, for cardiology specialties were 8.23 and 0.18, respectively.

DISCUSSION

Increasingly, U.S. physicians—from sports medicine to cardiology specialists—are being asked to interpret ECGs on athletes of all ages to enhance screening for SCD, in particular for PPE yet no formal certification process exists in the United States for interpreting athlete ECGs. Before

Table 4. Inter-Rater Agreement (Cohen's Kappa) for Detection of Abnormal ECG Patterns

Agreement of Interpretation between Individual Clinicians and Expert ECG Interpretation				
Individual	Kappa	Level of Agreement^a	95% CI	P Value^b
FM-1	0.80	Substantial	(0.66, 0.93)	<0.0001
FM-2	0.69	Substantial	(0.51, 0.86)	0.0006
IM-1	0.84	Almost perfect	(0.72, 0.96)	<0.0001
IM-2	0.59	Moderate	(0.40, 0.77)	0.0233
Ped-1	0.45	Moderate	(0.25, 0.66)	0.3053
Ped-2	0.70	Substantial	(0.53, 0.86)	0.0002
SM-1	0.65	Substantial	(0.49, 0.81)	0.0013
SM-2	0.75	Substantial	(0.59, 0.90)	<0.0001
Card-1	0.79	Substantial	(0.66, 0.93)	<0.0001
Card-2	0.74	Substantial	(0.60, 0.89)	<0.0001
Pcard-1	0.89	Almost perfect	(0.78, 1.00)	<0.0001
Pcard-2	0.76	Substantial	(0.62, 0.91)	<0.0001

Agreement of Interpretation within Clinical Specialties				
Specialty	Kappa	Level of Agreement^a	95% CI	P Value^b
FM	0.56	Moderate	(0.38, 0.74)	0.0396
IM	0.56	Moderate	(0.38, 0.74)	0.0433
Ped	0.37	Fair	(0.16, 0.59)	0.5935
SM	0.42	Moderate	(0.23, 0.61)	0.4091
Card	0.75	Substantial	(0.61, 0.90)	<0.0001
Pcard	0.65	Substantial	(0.48, 0.82)	0.0017

FM = Family Medicine; IM = Internal Medicine; Ped = Pediatrics; SM = Sports Medicine; Card = Cardiology; Pcard = Pediatric Cardiology.

^aLevel of agreement based on Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33(1):159-174.

^bNull hypothesis kappa = 0.4; one-sided z-test, statistical significance set at P = 0.05.

All interpretations according to 2010 ESC interpretation guidelines.

Table 5. Clinician Agreement (Fleiss' Kappa) for the Detection of abnormal ECG Patterns across Multiple Raters

Multi-Rater Agreement (Fleiss' Kappa) for the Detection of Abnormal ECG Patterns			
Specialty Group	Kappa	Level of Agreement	P Value^a
Primary care	0.52	Moderate	<0.0001
Specialty care	0.69	Substantial	<0.0001
All	0.57	Moderate	<0.0001

Specialty groups: "Primary Care" = Family Medicine, Internal Medicine, Pediatrics, and Sports Medicine; "Specialty Care" = Cardiology and Pediatric Cardiology; "All" = Primary Care and Specialty Care.

^aNull hypothesis kappa = 0; 1-sided z-test, statistical significance set at P = 0.05.

Chi-square test Primary & Specialty (Ho: no difference), P < 0.001.

All interpretations according to 2010 ESC interpretation guidelines.

considering widespread adoption of ECG screening as part of the PPE, understanding the diagnostic accuracy of clinician ECG interpretations is critical. In the present study, independent, blinded cardiology experts performed ECG interpretation to validate a set of ECGs; to assess the reliability and validity of ECG interpretation skills by primary care and cardiology specialists. The results of this study indicate inter-rater agreement between

experts was substantial/good (kappa = 0.63). Our study participants were experienced, clinically active board certified physicians with credentials to interpret ECGs at a single academic medical center. Clinicians were similar across specialties except in number of ECGs interpreted in the past year, which demonstrated important variation in primary care physicians. Measures of reliability were significantly higher in cardiology specialties

Table 6. Validity Measures for Clinician ECG Interpretation by Specialty

Validity Measures for the Detection of Abnormal Patterns by Specialty								
Specialty	Sensitivity (%)	95% CI		Specificity (%)	95% CI		PPV (%)	NPV (%)
FM	86.5	77.3	95.8	89.8	84.4	95.3	79.0	93.8
IM	86.5	77.3	95.8	88.1	82.3	94.0	76.3	93.7
Ped	65.4	52.5	78.3	90.7	85.4	95.9	75.6	85.6
SM	86.5	77.3	95.8	86.4	80.3	92.6	73.8	93.6
Card	96.2	90.9	100.0	86.4	80.3	92.6	75.8	98.1
Pcard	90.4	82.4	98.4	93.2	88.7	97.8	85.5	95.7
Primary	81.3	76.0	86.6	88.8	85.9	91.6	76.1	91.5
Specialty	93.3	88.5	98.1	89.8	86.0	93.7	80.2	96.8
All	85.3	81.3	89.2	89.1	86.8	91.4	77.6	93.2

Validity Measures for the Correct Identification of Specific Abnormal Pattern(s) in Abnormal ECGs by Specialty

Specialty	Sensitivity (%)	95% CI		Specificity (%)	95% CI		PPV (%)	NPV (%)
FM	67.3	54.6	80.1	89.8	84.4	95.3	66.0	86.2
IM	71.2	58.8	83.5	88.1	82.3	94.0	72.6	87.4
Ped	42.3	28.9	55.7	90.7	85.4	95.9	66.7	78.1
SM	59.6	46.3	73.0	86.4	80.3	92.6	66.0	82.9
Card	82.7	72.4	93.0	86.4	80.3	92.6	72.9	91.9
Pcard	84.6	74.8	94.4	93.2	88.7	97.8	84.6	93.2
Primary	60.1	53.4	66.8	88.8	85.9	91.6	70.2	83.5
Specialty	83.7	76.6	90.8	89.8	86.0	93.7	78.4	92.6
Overall	68.0	62.8	73.1	89.1	86.8	91.4	73.4	86.3

FM = Family Medicine; IM = Internal Medicine; Ped = Pediatrics; SM = Sports Medicine; Card = Cardiology; Pcard = Pediatric Cardiology; Primary = FM, IM, Ped, SM; Specialty = Card, Pcard; Overall = all clinicians.

Hypothesis testing for detecting the presence of an abnormal pattern: Wilcoxon rank-sum, two-sided exact test: Primary & Specialty sensitivity: (Ho: difference = 0), $Pr > |S - \text{Mean}| = 0.3111$.

Primary & Specialty specificity: (Ho: difference = 0), $Pr > |S - \text{Mean}| = 0.9111$.

Hypothesis testing for correct identification of specific abnormal patterns: Wilcoxon rank-sum, two-sided exact test: Primary & Specialty sensitivity: (Ho: difference = 0), $Pr > |S - \text{Mean}| = \mathbf{0.0485}$.

Primary & Specialty specificity: (Ho: difference = 0), $Pr > |S - \text{Mean}| = 0.9111$.

All interpretations according to 2010 ESC interpretation guidelines.

than primary care specialties; yet reliability across both groups fell below the experts. This is a concerning finding given the clinicians were interpreting a validated ECG set. The sensitivity for cardiology specialists to detect any abnormal patterns was 93.3% and higher than primary care specialties (81.3%); this difference was statistically significant when specific abnormal patterns were evaluated (83.7% vs 60.1%, $P = 0.0485$). Overall, false positive rates were above 10% for clinicians across all specialties in this study.

This study included important biases. The validated ECG set, containing 26 abnormal ECGs with 14 specific abnormal patterns, was selected to represent ECG patterns associated with common causes of SCD in U.S. athletes in a justified frequency. Validation of abnormal patterns by

independent, blinded experts biases the ECG set to include more "clear-cut" patterns than may be present in a general athlete population. Combined with an increased prevalence of abnormal ECGs, these biases systematically optimize clinician reliability and validity measures beyond what might be expected in screening of a general athlete population. Our findings therefore raise legitimate concerns for the use of ECG in PPE of U.S. athletes by physicians similar to those in our study population.

Compared to the reference standard, sensitivity and specificity to detect abnormal patterns were not significantly different between primary care and cardiology specialties. The 95% confidence intervals suggest cardiology specialties may have a higher sensitivity to detect abnormal patterns (88.5

to 98.1 vs 76.0 to 86.6), though with limited power this trend is not significant. Cardiology specialties demonstrate significantly higher sensitivity to *correctly* identify the specific abnormal ECG patterns when present (83.7% vs 60.1%, $P = 0.0485$). Thus, it appears primary care and cardiology specialties may be similarly able to identify when an abnormal pattern is present, yet cardiologists demonstrate significantly improved ability to correctly identify the specific abnormal patterns according to the 2010 ESC criteria. This may represent an important difference when performing confirmatory testing, but somewhat less important in screening tests.

The validity measures reported in this study are similar to data reported by Drezner et al.,²⁵ particularly for specificity measures, falling close to those obtained for respective specialties after receiving the two-page ECG criteria tool. This is reassuring as the quick reference guide in this study was adapted with permission from the ECG criteria tool. Important differences exist between these studies. The study by Drezner et al. did not assess for reliability measures and may have encountered test-retest bias. Taken together, these studies highlight a need for adequate training to optimize validity measures, and perhaps, to improve measures of reliability as well.

Limitations

The ESC interpretation criteria are not widely utilized in the United States for athlete screening and not all primary care specialties currently interpret high-volume ECGs in routine practice. To familiarize study experts and clinicians with the criteria, each was provided a straightforward, easy to use reference guide. Study participants received no formal training. Each participant was provided the ESC criteria in a quick reference guide, similar to the ECG criteria tool used by Drezner et al.²⁵ This study was powered for inter-rater agreement as an assessment of reliability between pairs of clinicians within each specialty. In contrast, the study was not adequately powered for validity measures and the observed sensitivity and specificity may not accurately reflect the abilities within each specialty. Future study of larger

clinician populations would optimize power for these validity measures and should be considered. Although the variation in the numbers of ECGs interpreted in the past 12 months across primary care physicians may have impacted both reliability and validity in this group, it is not known if this variation is commonplace across these specialties in the United States and therefore remains an important finding of this study. The ECGs used in this study are not from athletes, yet clearly demonstrate the patterns of diagnostic concern according to the ESC guidelines. Lastly, expert ECG interpretation demonstrated limitations in agreement, but remains the accepted reference standard.

CONCLUSIONS

Screening athletes for SCD with ECG remains controversial in the United States. To date, insufficient research has addressed the diagnostic accuracy of clinician ECG interpretation, particularly with regard to the reliability of clinicians responsible for PPE in athletes, including cardiology specialties. Previous reports have raised awareness of the limited utility of ECG interpretation for PPE of athletes when performed by inadequately trained physicians.^{12,16} The current study supports this caution, though cardiology specialists appear to demonstrate improved accuracy measures. Interpreting a validated set of ECGs with a reference guide, clinicians were less reliable than experts, with primary care and cardiology specialties each generating concerning false positive and false negatives. Placed back in the context of athlete PPE, such shortcomings could result in failure to prevent SCD in athletes and restriction from sport while further cardiac evaluation is pursued. Improved understanding of the factors influencing accuracy of ECG interpretation in athletes is clearly warranted. Likewise, efforts to improve reliability and validity, similar to those piloted by Drezner et al.,²⁵ should continue to be developed and tested prior to implementation of ECG interpretation by primary care specialties for PPE in U.S. athletes.

Appendix A: 2010 ESC guidelines Reference Guide

Abnormality Code	Abnormal ECG Finding	Definition
A	T-wave inversion	1 mm in depth from baseline in two or more adjacent leads not including aVR or V ₁ (¹ note exception below – Figure 1)
B	ST segment depression	≥ 1 mm in depth in two or more adjacent leads
C	Pathologic Q waves	≥ 3 mm in depth or > 0.04 sec in duration in two or more leads
D	Complete left bundle branch block	QRS > 0.12 sec, predominantly negative QRS complex in lead V ₁ (QS or rS), and upright monophasic R wave in leads I and V ₆ (Figure 2)
E	Complete right bundle branch block	QRS > 0.12 sec, terminal R wave in lead V ₁ (rsR'), and wide terminal S wave in leads I and V ₆ (Figure 3)
F	Intraventricular conduction delay	Nonspecific, QRS > 0.12 sec
G	Left atrial enlargement	Prolonged P wave duration of > 0.12 sec in leads I or II with negative portion of the P wave ≥ 1 mm in depth and ≥ 0.04 sec in duration in lead V ₁
H	Left axis deviation	–30° to –90°
I	Right atrial enlargement	High/pointed P wave ≥ 2.5 mm in leads II and or V ₁
J	Right ventricular hypertrophy	Right axis deviation ≥ 120°, tall R wave in V ₁ + persistent precordial S waves (R-V ₁ + S-V ₅ > 10.5 mm)
K	Mobitz type II 2° AV block	Intermittently nonconducted P waves not preceded by PR prolongation and not followed by PR shortening
L	3° AV block	Complete heart block
M	Ventricular preexcitation	PR interval < 0.12 sec with a delta wave (slurred upstroke in the QRS complex [Figure 4])
N	Long QT interval	QTc ≥ 0.47 sec (99% males) ; QTc ≥ 0.48 sec (99% females); [QTc ≥ 0.50 sec (unequivocal LQTS) [Figure 5]
O	Short QT interval	QTc ≤ 0.34 sec
P	Brugada-like ECG pattern	High take-off and downsloping ST segment elevation in V ₁ -V ₃
Q	Epsilon wave	Small negative deflection just beyond the QRS in V ₁ or V ₂ (Figure 7)
R	Profound sinus bradycardia	< 30 BPM or sinus pauses ≥ 3 sec
S	Atrial tachyarrhythmias	Supraventricular tachycardia, atrioventricular nodal reentrant tachycardia, atrial fibrillation, atrial flutter
T	Premature ventricular contractions	≥ 2 per tracing
U	Ventricular arrhythmias	Couplets, triplets, nonsustained ventricular tachycardia

REFERENCES

1. Maron BJ. Sudden death in young athletes. *N Engl J Med* 2003;349:1064–1075.
2. Maron BJ, Shirani J, Poliac LC, Mathenge R, Roberts WC, Mueller FO. Sudden death in young competitive athletes. Clinical, demographic, and pathological profiles. *JAMA* 1996;276:199–204.
3. Van Camp SP, Bloor CM, Mueller FO, Cantu RC, Olson HG. Nontraumatic sports death in high school and college athletes. *Med Sci Sports Exerc* 1995;27:641–647.
4. Burke AP, Farb A, Virmani R, Goodin J, Smialek JE. Sports-related and non-sports-related sudden cardiac death in young adults. *Am Heart J* 1991;121:568–575.

5. Driscoll DJ, Edwards WD. Sudden unexpected death in children and adolescents. *J Am Coll Cardiol* 1985;5:118B-121B.
6. Maron BJ, Gohman TE, Aeppli D. Prevalence of sudden cardiac death during competitive sports activities in Minnesota high school athletes. *J Am Coll Cardiol* 1998;32:1881-1884.
7. Maron BJ, Doerer JJ, Haas TS, et al. Sudden deaths in young competitive athletes: Analysis of 1866 deaths in the United States, 1980-2006. *Circulation* 2009;119:1085-1092.
8. Harmon KG, Asif IM, Klossner D, et al. Incidence of sudden cardiac death in national collegiate athletic association athletes. *Circulation* 2011;123:1594-1600.
9. Corrado D, Pelliccia A, Heidbuchel H, et al. Recommendations for interpretation of 12-lead electrocardiogram in the athlete. *Eur Heart J* 2010;31:243-259.
10. Corrado D, Basso C, Pavei A, et al. Trends in sudden cardiovascular death in young competitive athletes after implementation of a preparticipation screening program. *JAMA* 2006;296:1593-1601.
11. Pelliccia A, Culasso F, Di Paolo FM, et al. Prevalence of abnormal electrocardiograms in a large, unselected population undergoing pre-participation cardiovascular screening. *Eur Heart J* 2007;28:2006-2010.
12. Corrado D, McKenna WJ. Appropriate interpretation of the athlete's electrocardiogram saves lives as well as money. *Eur Heart J* 2007;28:1920-1922.
13. Corrado D, Biffi A, Basso C, et al. 12-lead ECG in the athlete: Physiological versus pathological abnormalities. *Br J Sports Med* 2009;43:669-676.
14. Corrado D, Pelliccia A, Bjørnstad HH, et al. Cardiovascular pre-participation screening of young competitive athletes for prevention of sudden death: Proposal for a common European protocol. Consensus Statement of the Study Group of Sport Cardiology of the Working Group of Cardiac Rehabilitation and Exercise Physiology and the Working Group of Myocardial and Pericardial Diseases of the European Society of Cardiology. *Eur Heart J* 2005;26:516-524.
15. Bessem B, Groot FP, Nieuwland W. The Lausanne recommendations: A Dutch experience. *Br J Sports Med* 2009;43:708-715.
16. Corrado D, Basso C, Schiavon M, et al. Pre-participation screening of young competitive athletes for prevention of sudden cardiac death. *J Am Coll Cardiol* 2008;52:1981-1989.
17. Maron BJ, Thompson PD, Ackerman MJ, et al. Recommendations and considerations related to preparticipation screening for cardiovascular abnormalities in competitive athletes: 2007 update: A scientific statement from the American Heart Association Council on Nutrition, Physical Activity, and Metabolism: Endorsed by the American College of Cardiology Foundation. *Circulation* 2007;115:1643-1655.
18. Oakley DG, Oakley CM. Significance of abnormal electrocardiograms in highly trained athletes. *Am J Cardiol* 1982;50:985-989.
19. Huston TP, Puffer JC, Rodney WM. The athletic heart syndrome. *N Engl J Med* 1985;313:24-32.
20. Fagard R. Athlete's heart. *Heart* 2003;89:1455-1461.
21. Steinvil A, Chundadze T, Zeltser D, et al. Mandatory electrocardiographic screening of athletes to reduce their risk for sudden death proven fact or wishful thinking? *J Am Coll Cardiol* 2011;57:1291-1296.
22. Baggish AL, Hutter AM, Wang F, et al. Cardiovascular screening in college athletes with and without electrocardiography: A cross-sectional study. *Ann Intern Med* 2010;152:269-275.
23. Drezner JA, Ackerman MJ, Anderson J, et al. Electrocardiographic interpretation in athletes: The 'Seattle Criteria'. *Br J Sports Med* 2013;47:122-124.
24. Hill AC, Miyake CY, Grady S, et al. Accuracy of interpretation of preparticipation screening electrocardiograms. *J Pediatr* 2011;159:783-788.
25. Drezner JA, Asif IM, Owens DS, et al. Accuracy of ECG interpretation in competitive athletes: The impact of using standised ECG criteria. *Br J Sports Med* 2012;46:335-340.
26. Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: Explanation and elaboration. *Clin Chem* 2003;49:7-18.
27. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-174.
28. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull* 1971;76:378-382.
29. Pelliccia A, Maron BJ, Culasso F, et al. Clinical significance of abnormal electrocardiographic patterns in trained athletes. *Circulation* 2000;102:278-284.
30. Sim J, Wright CC. The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Phys Ther* 2005;85:257-268.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web site:

Figure 1. Normal variant of T wave inversion in athletes of African-Caribbean descent.

Figure 2. Left Bundle Branch Block: QRS > 0.12 sec, predominantly negative QRS complex in lead V₁ (QS or rS), and upright monophasic R wave in leads I and V₆.

Figure 3. Right Bundle Branch Block: QRS > 0.12 sec, terminal R wave in lead V₁ (rsR'), and wide terminal S wave in leads I and V₆.

Figure 4. Delta Wave: Suggestive of ventricular preexcitation; PR interval < 0.12 sec with or without a delta wave (slurred upstroke in the QRS complex).

Figure 5. QTc Interval: LONG QT : QTc ≥ 0.47 sec (99% males) or QTc ≥ 0.48 sec (99% females), or QTc ≥ 0.50 sec (unequivocal LQTS); SHORT QT: QTc ≤ 0.34 sec.

Figure 6. Brugada ECG: High take-off and downsloping ST segment elevation in V₁-V₃.

Figure 7. Epsilon Wave: Small negative deflection just beyond the QRS in V₁ or V₂.