Original Article

Inter-Rater Reliability and Downstream Financial Implications of Electrocardiography Screening in Young Athletes

Harshil Dhutia, BSc, MRCP; Aneil Malhotra, BSc, MRCP, PhD; Tee Joo Yeo, MRCP; Irina Chis Ster, MSc, PhD; Vincent Gabus, MD; Alexandros Steriotis, MD; Helder Dores, MD; Greg Mellor, MB BChir, MRCP; Carmen García-Corrales, MD; Bode Ensam, MRCP; Viknesh Jayalapan, MRCP; Vivienne Anne Ezzat, MBChB, PhD; Gherardo Finocchiaro, MD; Sabiha Gati, BSc, MRCP, PhD; Michael Papadakis, MD, MRCP; Maria Tome-Esteban, PhD; Sanjay Sharma, BSc, MD

Background**—**Preparticipation screening for cardiovascular disease in young athletes with electrocardiography is endorsed by the European Society of Cardiology and several major sporting organizations. One of the concerns of the ECG as a screening test in young athletes relates to the potential for variation in interpretation. We investigated the degree of variation in ECG interpretation in athletes and its financial impact among cardiologists of differing experience.

Methods and Results**—**Eight cardiologists (4 with experience in screening athletes) each reported 400 ECGs of consecutively screened young athletes according to the 2010 European Society of Cardiology recommendations, Seattle criteria, and refined criteria. Cohen κ coefficient was used to calculate interobserver reliability. Cardiologists proposed secondary investigations after ECG interpretation, the costs of which were based on the UK National Health Service tariffs. Inexperienced cardiologists were more likely to classify an ECG as abnormal compared with experienced cardiologists (odds ratio, 1.44; 95% confidence interval, 1.03–2.02). Modification of ECG interpretation criteria improved interobserver reliability for categorizing an ECG as abnormal from poor (2010 European Society of Cardiology recommendations; κ=0.15) to moderate (refined criteria; κ=0.41) among inexperienced cardiologists; however, interobserver reliability was moderate for all 3 criteria among experienced cardiologists (κ=0.40–0.53). Inexperienced cardiologists were more likely to refer athletes for further evaluation compared with experienced cardiologists (odds ratio, 4.74; 95% confidence interval, 3.50–6.43) with poorer interobserver reliability (κ=0.22 versus κ=0.47). Interobserver reliability for secondary investigations after ECG interpretation ranged from poor to fair among inexperienced cardiologists (κ=0.15–0.30) and fair to moderate among experienced cardiologists (κ=0.21–0.46). The cost of cardiovascular evaluation per athlete was $175 (95% confidence interval, $142–$228) and $101 (95% confidence interval, $83–$131) for inexperienced and experienced cardiologists, respectively.

Conclusions**—**Interpretation of the ECG in athletes and the resultant cascade of investigations are highly physician dependent even in experienced hands with important downstream financial implications, emphasizing the need for formal training and standardized diagnostic pathways. **(*Circ Cardiovasc Qual Outcomes*. 2017;10:e003306. DOI: 10.1161/CIRCOUTCOMES.116.003306.)**

Key Words: athletes ■ cardiologists ■ death, sudden, cardiac ■ heart disease ■ sports

Preparticipation cardiovascular screening of young athletes with electrocardiography (ECG) is effective for detecting potentially serious cardiac disease and is endorsed by the European Society of Cardiology (ESC) and several international sporting bodies.[1–3] However, the accuracy of the ECG is dependent on the individual interpretation of the test, which may vary considerably among cardiologists of differing experience.[4,5]

See Editorial by Prutkin and Drezner

Received November 13, 2016; accepted June 27, 2017.

From the Cardiology Clinical and Academic Group (H.D., A.M., G.F., S.G., M.P., M.T., S.S.) and Institute of Infection and Immunity (I.C.S.), St. George's University of London, United Kingdom; Department of Cardiology, National University Heart Centre Singapore, Singapore, China (T.J.Y.); Service of Cardiology, Department of Medicine, Lausanne University Hospital, Lausanne, Switzerland (V.G.); Department of Cardiology, St Bartholomew's Hospital, London, United Kingdom (A.S., V.A.E.); Armed Forces Hospital, Luz Hospital and CEDOC-NOVA Medical School, Lisbon, Portugal (H.D.); Heart Rhythm Services, Division of Cardiology, Department of Medicine, University of British Columbia, Vancouver, Canada (G.M.); Infanta Cristina Hospital, Badajoz, Spain (C.G.-C.); University Hospital Birmingham, Birmingham, United Kingdom (B.E.); and William Harvey Hospital, Ashford, United Kingdom (V.J.).

The Data Supplement is available at http://circoutcomes.ahajournals.org/lookup/suppl/doi:10.1161/CIRCOUTCOMES.116.003306/-/DC1.

Correspondence to Sanjay Sharma, BSc, MD, Division of Cardiovascular Sciences, St. George's University of London, Cranmer Terrace, London, SW17 0RE, United Kingdom. E-mail sasharma@sgul.ac.uk

WHAT IS KNOWN

- The inclusion of the ECG to a health questionnaire and physical examination screening protocol in young athletes improves sensitivity to detect serious cardiac disease; however, a concern of the ECG as a screening tool relates to the potential for variation in interpretation especially in inexperienced hands.

WHAT THE STUDY ADDS

- There is only moderate interobserver reliability for ECG interpretation even among cardiologists with experience in the cardiovascular evaluation of young athletes.
- Modification of ECG interpretation criteria improves reliability among inexperienced cardiologists.
- The decision to propose secondary investigations after ECG interpretation varies among inexperienced and experienced cardiologists, respectively, with significant downstream financial implications.
- The findings of this study highlight that formal training and development of standardized diagnostic pathways are essential to support cardiologists involved in cardiovascular screening of young athletes.

Recent modification of ECG interpretation recommendations has improved the efficacy of the ECG as a screening tool by reducing the false-positive rate and cost of screening.[3,6–8] Whether such modification has an impact on the variation of ECG interpretation in young athletes is unknown. Furthermore, whether experience of reporting ECGs in athletes affects variability of interpretation and recommended secondary testing among cardiologists is also unknown. This study evaluated the variation of interpretation of the athlete's ECG and its financial impact between experienced and inexperienced cardiologists using 3 internationally recognized ECG interpretation recommendations.[3,6,7]

Methods

Study Population

The charitable organization Cardiac Risk in the Young has an established cardiac screening program for young individuals aged 14 to 35 years, which serves many professional sporting organizations in the United Kingdom (www.c-r-y.org.uk). The Cardiac Risk in the Young screening protocol consists of a health questionnaire pertaining to symptoms suggestive of cardiac disease or a family history of cardiac disease, a physical examination, and a 12-lead ECG. The first 400 consecutively assessed athlete's ECGs from the program in 2014 were used for the primary analysis of interobserver agreement. These athletes have been presented previously as part of a nationwide ECG screening cost analysis.[8] None of the athletes were considered to have symptoms suggestive of cardiovascular disease, and none had a significant family history of cardiovascular disease. All had a normal physical examination. These athletes were evaluated by different experienced sports cardiologists. Purely as a reference, 24 (6.0%) of the athletes were referred for further evaluation, 23 (5.8%) underwent echocardiography, 6 (1.5%) exercise stress testing, 8 (2.0%) Holter, and 5 (1.3%) cardiac magnetic resonance imaging.

Participants

Eight cardiologists independently participated in the interpretation of the ECGs, of whom 4 were experienced in evaluating the ECG in athletes. For the purposes of the study, we defined cardiologists with experience as those who were working in a specialist sports cardiology unit for >2 years and had independently conducted preparticipation ECG screening with ECG in ≥1000 athletes. Conversely, inexperienced cardiologists were defined as those who did not routinely report on athlete's ECG. Both groups consisted of 3 general cardiologists and 1 electrophysiologist.

ECG Interpretation

All cardiologists were provided with the 400 anonymized ECGs in random order and in a digital printable format, which included the age, sex, and ethnicity of the athlete. The cardiologists were informed that all athletes had normal history and physical examination findings. Digital measurements of heart rate, QRS duration, PR interval, and QT interval were omitted.

The cardiologists were provided with a copy of published documents detailing the 2010 ESC recommendations, the Seattle criteria, and the refined criteria 1 month before commencement of ECG interpretation (Table 1).[3,6,7] Each cardiologist was instructed to assign the ECGs as normal or abnormal per criterion and specify the abnormities.

All cardiologists calculated the QT interval manually. Instructions were provided on measuring the QT interval using the tangent method.[9] Cardiologists were advised to report the longest QT interval value as the absolute QT and to correct the QT interval for heart rate using the Bazett formula, where the corrected QT interval (QTc)=QT/$\sqrt{\text{RR}}$ interval.[10]

Secondary Investigations and Financial Analysis

The initial preparticipation screening tests (history, physical examination, and ECG) were performed at a subsidized cost of $53 per athlete screened.

In the event of an abnormal ECG, the cardiologists were instructed to propose specific secondary investigations based on their usual clinical practice. The cost of secondary investigations was calculated based on the 2014/2015 UK National Health Service tariff payment system (Table I in the Data Supplement). Genetic testing was not included in the cost analysis because it is usually reserved for individuals with disease phenotype for the purposes of cascade screening.

Statistics

The data were graphically explored and summarized accordingly, that is, means, SDs, median interquartile range, and range for continuous data and proportions for categorical or binary independent data.

Raw indices of interobserver agreement are presented as the overall and specific proportions of agreement among the groups of cardiologists. Cohen κ coefficient was used to calculate the overall interobserver reliability in ECG interpretation between groups of cardiologists (experienced and inexperienced) with κ<0.20 representing poor interobserver reliability, 0.20 to 0.40 representing fair reliability, 0.40 to 0.60 representing moderate reliability, 0.60 to 0.80 representing good agreement, and 0.80 to 1.00 representing very good reliability. To disentangle the potential heterogeneities in the κ values across age, sex, and ethnicity, 2 novel binary variables were constructed for each group of clinical experts in a similar fashion: 1 if all 4 clinicians perfectly agreed, 0 otherwise. Then, a bivariate logistic regression was applied to the joined binary outcome for a simultaneous flavor of the odds of perfect agreement within the 2 clinical groups. As this is not an indication of agreement because some information is lost, the Cohen agreement coefficient was recalculated across heterogeneous groups in the population indicated by this analysis and subsequently presented. Further referrals comprised 5 destinations (echocardiography, exercise stress test, Holter, cardiac magnetic resonance imaging, and family screening) with the possibility that a patient required >1. Given the binary nature of this

Table 1. Summary of Definition of ECG Abnormalities in Athletes According to the 2010 ESC Recommendations, Seattle Criteria, and Refined Criteria[3,6,7]

All 3 criteria	ST-segment depression
	Pathological Q waves
	Complete left bundle branch block
	Ventricular pre-excitation
	Brugada-like early repolarization pattern
	Premature ventricular contractions
	Atrial or ventricular arrhythmia
2010 ESC recommendations	T-wave inversion
	Long-QT interval >440 ms (male) or >460 ms (female)
	Short-QT interval <380 ms
	Right ventricular hypertrophy
	Right- or left-axis deviation
	Right or left atrial enlargement
	Complete right bundle branch block
	Nonspecific intraventricular delay (QRS >120 ms)
Seattle criteria	T-wave inversion beyond V2 in white athletes
	T-wave inversion beyond V4 in black athletes
	Long-QT interval ≥470 ms (male) or ≥480 ms (female)
	Short-QT interval ≤320 ms
	Right ventricular hypertrophy (in presence of right-axis deviation)
	Left-axis deviation
	Right or left atrial enlargement
	Nonspecific intraventricular delay (QRS ≥140 ms)
Refined criteria	T-wave inversion beyond V1 in white athletes
	T-wave inversion beyond V4 in black athletes
	Long-QT interval ≥470 ms (male) or ≥480 ms (female)
	Short-QT interval ≤320 ms
	Complete right bundle branch block
	Borderline variants (requiring investigation if >1 present)
	T-wave inversion up to V4 in black athletes
	Right ventricular hypertrophy
	Right- or left-axis deviation
	Right or left atrial enlargement

ESC indicates European Society of Cardiology.

multivariate response, we adopted a simpler yet easily interpretable approach. Each destination was considered a separate binary outcome, and a 2-level logistic regression applied to account for the inherent dependencies in the data arising from multiple measurements for the same athlete.

P values <0.05 were considered statistically significant, and the uncertainty of the estimates is expressed as their 95% confidence intervals (CI). Marginal predictions, that is, predicted proportions summarized according to the clinical relevance, are also presented. The analyses were performed in Stata (StataCorp 2015, Stata Statistical Software).

Ethics

Ethical approval was granted by the Essex 2 Research Ethics Committee. Written consent was obtained from individuals ≥16 years of age and from a parent/guardian for those <16 years of age.

Results

Demographics

Athletes were aged 20.5±4.8 years. Two hundred and eighty-five (71%) were male. Three hundred and eighteen (79%) athletes were white and 43 (11%) were of African/Afro-Caribbean origin (black). Thirty-nine (10%) athletes consisted of other ethnicities including mixed race, Asian or Polynesian. Athletes competed in a total of 18 different sporting disciplines—predominantly soccer (29%), rugby (16%), and cycling (15%)—and exercised for 16.6±6.0 hours per week.

Identification of ECGs Suggestive of Cardiac Disease

One (0.3%) athlete was diagnosed with potentially serious cardiac disease, notably long-QT syndrome (QTc 520 ms). The ECG of this athlete was classified as requiring further evaluation by all 8 cardiologists.

Categorization of ECG Abnormalities in Accordance to ECG Interpretation Criteria

Frequency of ECG Abnormalities

Inexperienced cardiologists more frequently categorized an ECG as abnormal compared with experienced cardiologists for all 3 criteria (Figure 1). Compared with both the 2010 ESC recommendations and the Seattle criteria, the refined criteria reduced the proportion of ECGs categorized as abnormal among all cardiologists.

Interobserver Reliability

Interobserver reliability for categorizing an ECG as abnormal among inexperienced cardiologists was poor for the 2010 ESC recommendations (κ=0.15; 95% CI, 0.12–0.20), fair for the Seattle criteria (κ=0.25; 95% CI, 0.16–0.32), and moderate for the refined criteria (κ=0.41; 95% CI, 0.24–0.50; Figure 2). Among experienced cardiologists, there was moderate reliability for categorizing an ECG as abnormal for all 3 criteria (2010 ESC recommendations: κ=0.40; 95% CI, 0.37–0.45; Seattle criteria: κ=0.53; 95% CI, 0.39–0.64; and refined criteria: κ=0.43; 95% CI, 0.21–0.51).

Interobserver reliability for the presence of a long-QT interval was only fair to moderate (κ=0.21–0.44) among inexperienced cardiologists (Table II in the Data Supplement). Conversely, interobserver reliability for the presence of a long-QT interval among experienced cardiologists improved from fair (κ=0.31) with the 2010 ESC recommendations to good (κ=0.60) with the Seattle and refined criteria. There was

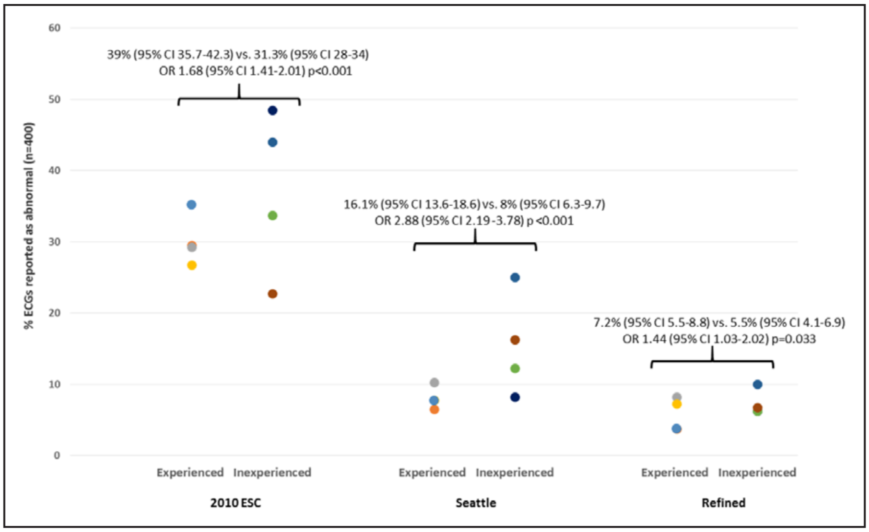

Figure 1. ECGs categorized as abnormal by each cardiologist. Inexperienced cardiologists are more likely to categorize an ECG as abnormal compared with experienced cardiologists irrespective of criteria used. CI indicates confidence interval; and OR, odds ratio.

moderate and good reliability for the presence of abnormal T-wave inversion among inexperienced (κ=0.43–0.54) and experienced cardiologists (κ=0.54–0.64), respectively. The degree of reliability for the presence for other ECG abnormalities ranged from poor to moderate among the cardiologists (Table II in the Data Supplement).

Proportion of Overall and Specific Interobserver Agreement

Proportion of overall interobserver agreement was higher among experienced cardiologists compared with inexperienced cardiologists for all 3 criteria (Table 2). Among both groups, overall agreement was driven specifically by agreement in ECGs categorized as normal. Modification of ECG interpretation criteria improved the overall proportion of interobserver agreement for both inexperienced and experienced cardiologists.

Bivariate Binary Outcome Model

On the basis of the 2010 ESC recommendations, sex (male athletes) and ethnicity (black athletes) proved to be the strongest predictors of disagreement among inexperienced cardiologists as suggested by the bivariate binary mode, whereas sex (male athletes) was the main source of difference in agreement among experienced cardiologists (Table III in the Data Supplement).

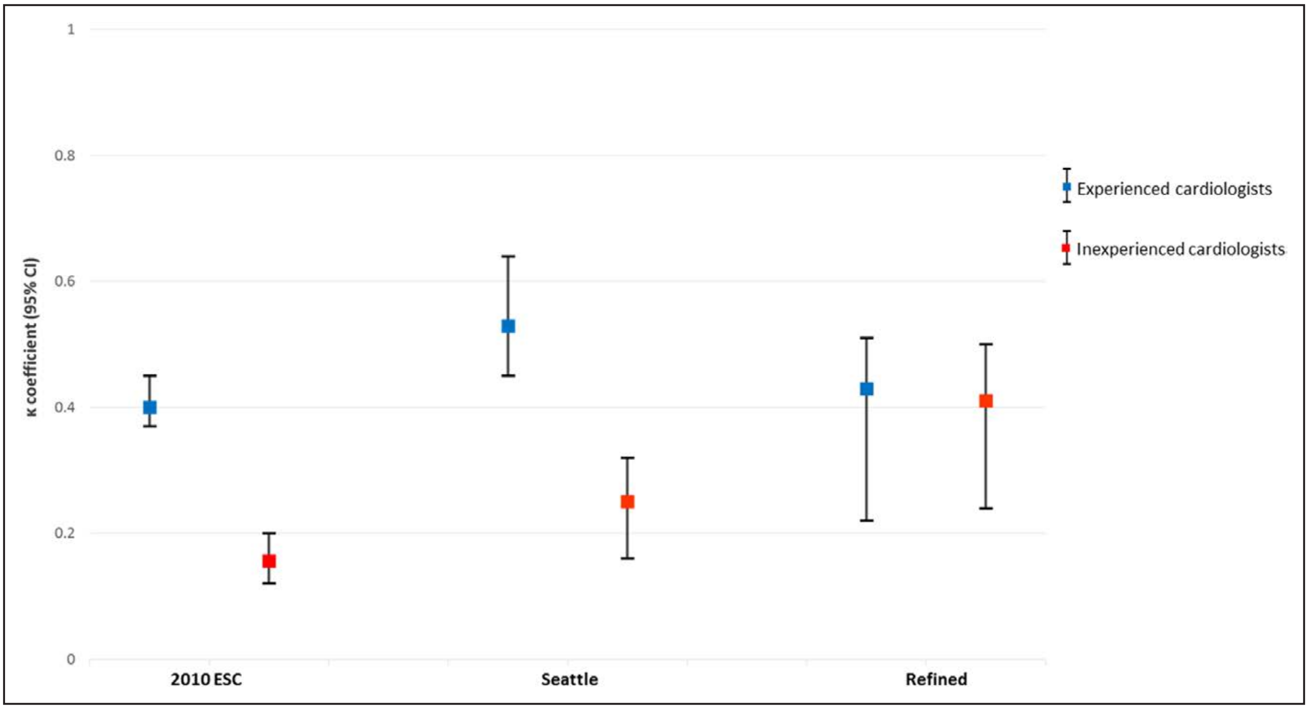

Figure 2. The estimated interobserver reliability among cardiologists for categorizing an ECG as abnormal. Interobserver reliability for an abnormal ECG among experienced cardiologists was moderate for all 3 criteria. Among inexperienced cardiologists, modification of ECG criteria improved interobserver reliability from poor to moderate. CI indicates confidence interval; and ESC, European Society of Cardiology.

Table 2. Proportion of Overall and Specific Agreement for ECG Interpretation Among Cardiologists

	Cardiologists	Overall Agreement (+95% CI)	Agreement for Abnormal (+95% CI)	Agreement for Normal (+95% CI)
2010 ESC recommendations	Inexperienced	0.60 (0.58–0.63)	0.46 (0.44–0.52)	0.68 (0.65–0.71)
	Experienced	0.75 (0.72–0.77)	0.58 (0.53–0.63)	0.82 (0.79–0.82)
Seattle criteria	Inexperienced	0.80 (0.78–0.83)	0.44 (0.29–0.48)	0.88 (0.87–0.90)
	Experienced	0.93 (0.92–0.95)	0.57 (0.45–0.67)	0.96 (0.95–0.97)
Refined criteria	Inexperienced	0.92 (0.90–0.95)	0.46 (0.35–0.56)	0.96 (0.95–0.97)
	Experienced	0.94 (0.92–0.95)	0.46 (0.32–0.58)	0.97 (0.96–0.98)

Further Evaluation and Secondary Investigations

Frequency of Investigations and Interobserver Reliability

Inexperienced cardiologists recommended further evaluation in 17.7% (95% CI, 15.0%–20.0%) of the 400 ECGs and showed fair interobserver reliability with respect to which ECGs required further evaluation (κ=0.23; 95% CI, 0.14–0.30). Experienced cardiologists recommended further evaluation in 7.0% (95% CI, 5.3%–8.8%) of the 400 ECGs, with moderate interobserver reliability (κ=0.40; 95% CI, 0.31–0.53).

Inexperienced cardiologists were likely to recommend a higher proportion of secondary investigations compared with experienced cardiologists, with poor to fair interobserver reliability (Table 3; Figure 3). Interobserver reliability for secondary investigations ranged from fair to moderate among experienced cardiologists.

Proportion of Overall and Specific Interobserver Agreement

Proportion of overall and specific agreement for recommending further evaluation and secondary investigations after ECG interpretation was higher among experienced cardiologists (Table 4). Agreement for familial evaluation was comparable among both groups. In both groups, overall agreement was specifically driven by agreement in not recommending investigations.

Costs Generated by Secondary Investigation

On the basis of the predicted proportions for each group, the cost of secondary investigation among inexperienced cardiologists amounted to $48 697 (95% CI, $35 583–$69 896) and equated to $122 (95% CI, $89–$175) per athlete screened. For experienced cardiologists, the total cost of secondary investigation amounted to $19 123 (95% CI, $11 878–$30 726) and equated to $48 (95% CI, $30–$78) per athlete screened.

Accounting for the initial preparticipation screening costs, the overall cost per athlete equated to $175 (95% CI, $142–$228) and $101 (95% CI, $83–$131) for inexperienced and experienced cardiologists, respectively.

Discussion

The ECG is a relatively cheap investigation that improves the sensitivity for detecting potentially serious cardiac disease in athletes compared with history and examination alone.[11] Recent modification of ECG interpretation recommendations in athletes has significantly reduced false-positive rates without compromising sensitivity.[7,12,13] However, as with any subjective investigation, the effectiveness of the ECG is dependent on the individual interpretation of the test. This study conveys important data pertaining to variation in interpretation of the ECG in a relatively large cohort of highly trained athletes and reveals that there is only moderate reliability in ECG interpretation in athletes among experienced cardiologists. Intuitively, such variation will have significant financial implications on downstream costs of systematic evaluation of athletes.

Impact of Experience in Interpretation of the ECG in Athletes

An important concern about the ECG as a screening tool is the potential for erroneous diagnosis in inexperienced hands. This study reveals that inexperienced cardiologists were at least

Table 3. Estimated Proportions of Further Evaluation and Secondary Investigations Proposed After ECG Interpretation

	Proportion of ECGs Requiring Further Evaluation and Secondary Investigations (n=400) (Marginal Proportions [95% CI])		Odds Ratio (95% CI)	*P* Value
	Inexperienced Cardiologists, %	Experienced Cardiologists, %		
Further evaluation after screening	17.7 (15.0–20.0)	7.0 (5.5–8.8)	4.7 (3.5–6.4)	$P<0.001$
Echocardiography	16.4 (13.8–19.1)	6.5 (4.9–8.1)	4.7 (3.4–6.5)	$P<0.001$
Exercise stress test	6.9 (5.3–8.6)	1.7 (0.9–2.5)	5.9 (3.6–9.6)	$P<0.001$
Holter	6.1 (4.4–7.7)	1.7 (0.9–2.6)	5.5 (3.2–9.2)	$P<0.001$
Cardiac MRI	5.5 (3.8–7.1)	0.9 (0.2–1.5)	12.2 (6.1–24.6)	$P<0.001$
Family screening	0.8 (0.4–1.4)	0.9 (0.4–1.5)	0.9 (0.4–2.0)	$P=0.84$

CI indicates confidence interval; and MRI, magnetic resonance imaging.

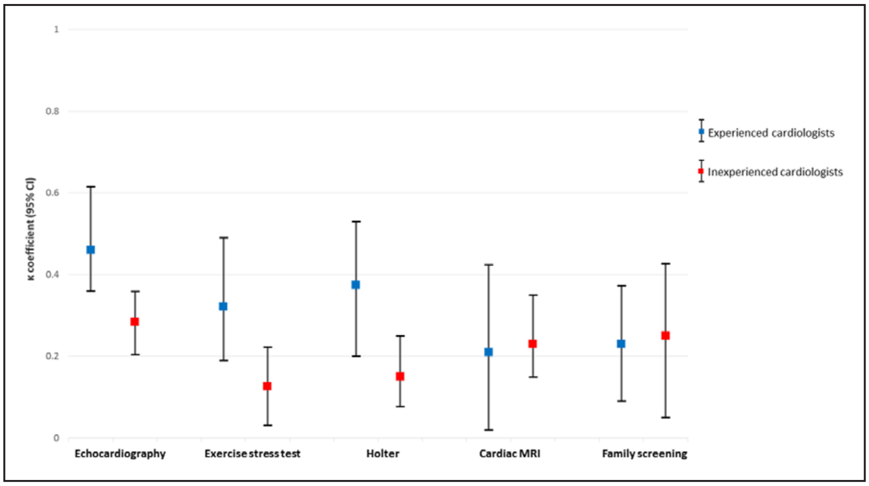

Figure 3. The estimated interobserver reliability among cardiologists for secondary investigations after ECG interpretation. Inexperienced cardiologists demonstrated poorer interobserver reliability for secondary investigations compared with experienced cardiologists. CI indicates confidence interval.

44% more likely to categorize an ECG as abnormal compared with experienced cardiologists (Figure 1). Furthermore, inexperience was associated with poorer interobserver reliability for categorizing an ECG as abnormal.

Reassuringly, the ECG of the athlete harboring potentially serious cardiac disease was identified by all 8 cardiologists. From the clinical point of view, it is arguable that this is certainly the most important aspect of the ECG as a screening tool.

Impact of Modification of Standardized ECG Interpretation Criteria

The impact of recent modification of ECG interpretation criteria on variation in interpretation in highly trained athletes has not been ascertained in a large cohort of highly trained young adult athletes although Berte et al[4] reported a higher overall agreement with the Seattle criteria compared with the 2010 ESC recommendations in a small cohort of adolescent male soccer players. We observed that contemporary ECG interpretation guidelines (Seattle and refined criteria) improve the proportion of overall interobserver agreement by ≤35% and 20% among inexperienced and experienced cardiologists, respectively (Table 2), and reduce the interobserver reliability gap between experienced and inexperienced cardiologists when categorizing an ECG as abnormal (Figure 2).

Impact of ECG Interpretation Variation on Workload and Costs

The workload and cost of secondary investigations required to confirm or refute the diagnosis of cardiac disease after an ECG abnormality are cited as important obstacles to screening young athletes with ECG.[14] We observed that inexperienced cardiologists were 5 times more likely to refer an athlete for further evaluation compared with experienced cardiologists based on ECG interpretation. Specifically, screening by inexperienced cardiologists resulted in a 5-fold increase in number of echocardiograms requested, 6-fold increase in exercise stress tests and Holter monitors, and a 12-fold increase in cardiac magnetic resonance imaging scans compared with experienced cardiologists (Table 3). In addition to the increased

Table 4. Proportions of Overall and Specific Agreement Among Cardiologists for Recommending Further Evaluation and Secondary Investigations After ECG Interpretation

	Cardiologists	Overall Agreement (+95% CI)	Agreement to Recommend (+95% CI)	Agreement Not to Recommend (+95% CI)
Further evaluation after screening	Inexperienced	0.77 (0.74–0.79)	0.38 (0.31–0.44)	0.86 (0.83–0.87)
	Experienced	0.92 (0.89–0.93)	0.45 (0.35–0.55)	0.95 (0.94–0.96)
Echocardiography	Inexperienced	0.81 (0.78–0.84)	0.40 (0.32–0.47)	0.89 (0.87–0.90)
	Experienced	0.94 (0.92–0.95)	0.50 (0.37–0.61)	0.97 (0.96–0.98)
Exercise stress test	Inexperienced	0.89 (0.87–0.91)	0.19 (0.10–0.29)	0.94 (0.93–0.95)
	Experienced	0.98 (0.96–0.99)	0.33 (0.11–0.53)	0.99 (0.98–0.99)
Holter	Inexperienced	0.91 (0.89–0.92)	0.20 (0.12–0.29)	0.95 (0.94–0.96)
	Experienced	0.98 (0.97–0.99)	0.38 (0.07–0.63)	0.99 (0.98–0.99)
Cardiac MRI	Inexperienced	0.92 (0.90–0.94)	0.27 (0.17–0.37)	0.96 (0.95–0.97)
	Experienced	0.99 (0.98–0.99)	0.24 (0.11–0.41)	0.99 (0.98–0.99)
Family screening	Inexperienced	0.99 (0.98–0.99)	0.31 (0.00–0.67)	0.99 (0.99–0.99)
	Experienced	0.98 (0.97–0.99)	0.12 (0.00–0.35)	0.99 (0.99–0.99)

CI indicates confidence interval; and MRI, magnetic resonance imaging.

number of investigations requested, interobserver reliability among inexperienced cardiologists for these secondary investigations ranged from poor to fair (Figure 3). Consequently, ECG-based preparticipation screening conducted by inexperienced cardiologists resulted in an ≈2-fold increase in cost compared with experienced cardiologists.

Experience was associated with a lower frequency of secondary investigations and improvement in proportion of overall agreement (Tables 3 and 4); nevertheless, interobserver reliability among experienced cardiologists for these investigations was only fair to moderate (Figure 3). In real-life practice, cardiovascular screening in athletes is conducted by physicians of varying experience, ranging from general cardiologists, electrophysiologists, and sports physicians. Therefore, the variation in interpretation is likely to fall between the 2 ranges above and will have significant implications on health resources that may preclude financial planning and sustainability of nationwide ECG screening of young athletes.

Despite demonstrating a reduction in the number of positive ECGs with contemporary ECG criteria, inexperienced cardiologists proposed further investigations in 17.7% of the cohort, which is considerably higher than expected based on the authors' experience and existing publications. This finding indicate that modification of ECG interpretation criteria may be associated with a lower positive ECG rate but may still not influence the personal practice among inexperienced cardiologists without further expert guidance.

Strategies to Reduce Variation in ECG Interpretation and Clinical Practice

Despite the success of the Italian athletic screening program in reducing sudden cardiac death, preparticipation screening with ECG remains a contentious issue given the absence of randomized control study evidence demonstrating that early detection of disease translates to lives saved, and consequently, ECG screening is not universally practiced. Nevertheless, ECG screening is endorsed by several major sporting organizations including Fédération Internationale de Football Association and the International Olympic Committee.[2,15] Although our study has shown only moderate reliability for interpretation of the athlete's ECG among experienced cardiologists, we do not aim to deter sporting organizations from screening athletes with ECG. Indeed, the ECG is associated with interobserver reliability rates that are comparable to well-established and generally accepted screening tests such as Papanicolaou smear testing for cervical carcinoma and mammography for carcinoma of the breast.[16–19] Furthermore, the only serious condition conferring increased risk of sudden cardiac death was identified on the basis of an abnormal ECG. By acknowledging the degree and impact of variation in ECG interpretation in athletes, our findings should herald the development of effective educational approaches aimed at reducing this variation. Modification of guidelines for the interpretation of the ECG in athletes from the 2010 ESC recommendations to the Seattle and refined criteria seem to be useful in improving agreement and reliability in ECG interpretation especially in inexperienced cardiologists; however, a better understanding of physiological versus pathological ECG patterns requires appropriate training and education of physicians including cardiologists to potentially minimize variation regardless of whether ECG analysis is being conducted for screening purposes or for diagnostic purposes. Recent small studies have demonstrated significant improvement in ECG interpretation in athletes after online training among inexperienced physicians and hold promise for the future.[20,21]

In comparison to other established and endorsed screening programs in the UK National Health Service (www.gov.uk/topic/population-screening-programmes) with similar rates of positive screening tests, there were no standardized diagnostic pathways for asymptomatic young athletes exhibiting ECG anomalies.[7,8,13,22–28] In this regard, the recently published international recommendations for ECG screening in athletes are unique as they provide guidance to physicians on the minimal set of investigations for each electric abnormality.[29,30] Such guidance will hopefully reduce variation in clinical practice among screening cardiologists, which may improve efficiency.

Limitations

This study has several limitations warranting mention. The definition of experience in interpreting the ECG in athletes was arbitrary, but there is currently no formal accreditation available to quantify this more accurately. We only included adult cardiologists for this study, and hence, the findings may not be readily applicable to organizations whose athletes are screened by pediatric cardiologists, sports physicians, and other healthcare providers. Only 400 athlete ECGs were included to make the study feasible. All 400 athletes were not assessed with echocardiography, and consequently there is no gold-standard reference on which to calculate the interpreter's accuracy for detecting of structural heart disease; however, our study aimed to investigate the level of interobserver variability in ECG interpretation and clinical practice rather than detection of disease. Although all cardiologists were supplied with published documents for all 3 ECG interpretation criteria, we cannot exclude that some may not have strictly adhered to them and reported ECGs based on their own experience. Machine-generated intervals were removed to aptly test the knowledge of the cardiologists in this study, but the impact they present on variation is not known.

Conclusions

Interpretation of the ECG in young athletes and the resultant cascade of downstream investigations is highly physician dependent even in experienced hands, which markedly impacts on the workload and cost of ECG screening. Formal training and development of a standardized diagnostic pathway is essential to support cardiologists involved in cardiovascular screening of young athletes.

Sources of Funding

H. Dhutia, A. Malhotra, and Dr Finocchiaro were funded by research grants from the Cardiac Risk in the Young organization.

Disclosures

Dr Sharma has been an applicant on previous grants from Cardiac Risk in the Young and British Heart Foundation to study athletes. The other authors report no conflicts.

References

1. Corrado D, Pelliccia A, Bjørnstad HH, Vanhees L, Biffi A, Borjesson M, Panhuyzen-Goedkoop N, Deligiannis A, Solberg E, Dugmore D, Mellwig KP, Assanelli D, Delise P, van-Buuren F, Anastasakis A, Heidbuchel H, Hoffmann E, Fagard R, Priori SG, Basso C, Arbustini E, Blomstrom-Lundqvist C, McKenna WJ, Thiene G; 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. 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. doi: 10.1093/eurheartj/ehi108.
2. Ljungqvist A, Jenoure P, Engebretsen L, Alonso JM, Bahr R, Clough A, De Bondt G, Dvorak J, Maloley R, Matheson G, Meeuwisse W, Meijboom E, Mountjoy M, Pelliccia A, Schwellnus M, Sprumont D, Schamasch P, Gauthier JB, Dubi C, Stupp H, Thill C. The International Olympic Committee (IOC) Consensus Statement on periodic health evaluation of elite athletes March 2009. *Br J Sports Med*. 2009;43:631–643. doi: 10.1136/bjsm.2009.064394.
3. Corrado D, Pelliccia A, Heidbuchel H, Sharma S, Link M, Basso C, Biffi A, Buja G, Delise P, Gussac I, Anastasakis A, Borjesson M, Bjørnstad HH, Carrè F, Deligiannis A, Dugmore D, Fagard R, Hoogsteen J, Mellwig KP, Panhuyzen-Goedkoop N, Solberg E, Vanhees L, Drezner J, Estes NA 3rd, Iliceto S, Maron BJ, Peidro R, Schwartz PJ, Stein R, Thiene G, Zeppilli P, McKenna WJ; Section of Sports Cardiology, European Association of Cardiovascular Prevention and Rehabilitation. Recommendations for interpretation of 12-lead electrocardiogram in the athlete. *Eur Heart J*. 2010;31:243–259. doi: 10.1093/eurheartj/ehp473.
4. Berte B, Duytschaever M, Elices J, Kataria V, Timmers L, Van Heuverswyn F, Stroobandt R, De Neve J, Watteyne K, Vandensteen E, Vandekerckhove Y, Tavernier R. Variability in interpretation of the electrocardiogram in young athletes: an unrecognized obstacle for electrocardiogram-based screening protocols. *Europace*. 2015;17:1435–1440. doi: 10.1093/europace/euu385.
5. Brosnan M, La Gerche A, Kumar S, Lo W, Kalman J, Prior D. Modest agreement in ECG interpretation limits the application of ECG screening in young athletes. *Heart Rhythm*. 2015;12:130–136. doi: 10.1016/j.hrthm.2014.09.060.
6. Drezner JA, Ackerman MJ, Anderson J, Ashley E, Asplund CA, Baggish AL, Börjesson M, Cannon BC, Corrado D, DiFiori JP, Fischbach P, Froelicher V, Harmon KG, Heidbuchel H, Marek J, Owens DS, Paul S, Pelliccia A, Prutkin JM, Salerno JC, Schmied CM, Sharma S, Stein R, Vetter VL, Wilson MG. Electrocardiographic interpretation in athletes: the 'Seattle criteria'. *Br J Sports Med*. 2013;47:122–124. doi: 10.1136/bjsports-2012-092067.
7. Sheikh N, Papadakis M, Ghani S, Zaidi A, Gati S, Adami PE, Carré F, Schnell F, Wilson M, Avila P, McKenna W, Sharma S. Comparison of electrocardiographic criteria for the detection of cardiac abnormalities in elite black and white athletes. *Circulation*. 2014;129:1637–1649. doi: 10.1161/CIRCULATIONAHA.113.006179.
8. Dhutia H, Malhotra A, Gabus V, Merghani A, Finocchiaro G, Millar L, Narain R, Papadakis M, Naci H, Tome M, Sharma S. Cost implications of using different ECG criteria for screening young athletes in the United Kingdom. *J Am Coll Cardiol*. 2016;68:702–711. doi: 10.1016/j.jacc.2016.05.076.
9. Al-Khatib SM, LaPointe NM, Kramer JM, Califf RM. What clinicians should know about the QT interval. *JAMA*. 2003;289:2120–2127. doi: 10.1001/jama.289.16.2120.
10. Funck-Brentano C, Jaillon P. Rate-corrected QT interval: techniques and limitations. *Am J Cardiol*. 1993;72:17B–22B.
11. Harmon KG, Zigman M, Drezner JA. The effectiveness of screening history, physical exam, and ECG to detect potentially lethal cardiac disorders in athletes: a systematic review/meta-analysis. *J Electrocardiol*. 2015;48:329–338. doi: 10.1016/j.jelectrocard.2015.02.001.
12. Brosnan M, La Gerche A, Kalman J, Lo W, Fallon K, MacIsaac A, Prior D. The Seattle Criteria increase the specificity of preparticipation ECG screening among elite athletes. *Br J Sports Med*. 2014;48:1144–1150. doi: 10.1136/bjsports-2013-092420.
13. Riding NR, Sheikh N, Adamuz C, Watt V, Farooq A, Whyte GP, George KP, Drezner JA, Sharma S, Wilson MG. Comparison of three current sets of electrocardiographic interpretation criteria for use in screening athletes. *Heart*. 2015;101:384–390. doi: 10.1136/heartjnl-2014-306437.
14. Maron BJ, Friedman RA, Kligfield P, Levine BD, Viskin S, Chaitman BR, Okin PM, Saul JP, Salberg L, Van Hare GF, Soliman EZ, Chen J, Matherne GP, Bolling SF, Mitten MJ, Caplan A, Balady GJ, Thompson PD; American Heart Association Council on Clinical Cardiology, Advocacy Coordinating Committee, Council on Cardiovascular Disease in the Young, Council on Cardiovascular Surgery and Anesthesia, Council on Epidemiology and Prevention, Council on Functional Genomics and Translational Biology, Council on Quality of Care and Outcomes Research, and American College of Cardiology. Assessment of the 12-lead ECG as a screening test for detection of cardiovascular disease in healthy general populations of young people (12-25 Years of Age): a scientific statement from the American Heart Association and the American College of Cardiology. *Circulation*. 2014;130:1303–1334. doi: 10.1161/CIR.0000000000000025.
15. Dvorak J, Grimm K, Schmied C, Junge A. Development and implementation of a standardized precompetition medical assessment of international elite football players–2006 FIFA World Cup Germany. *Clin J Sport Med*. 2009;19:316–321. doi: 10.1097/JSM.0b013e3181b21b6e.
16. Stoler MH, Schiffman M; Atypical Squamous Cells of Undetermined Significance-Low-grade Squamous Intraepithelial Lesion Triage Study (ALTS) Group. Interobserver reproducibility of cervical cytologic and histologic interpretations: realistic estimates from the ASCUS-LSIL Triage Study. *JAMA*. 2001;285:1500–1505.
17. Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR. Variability in radiologists' interpretations of mammograms. *N Engl J Med*. 1994;331:1493–1499. doi: 10.1056/NEJM199412013312206.
18. Berg WA, Campassi C, Langenberg P, Sexton MJ. Breast imaging reporting and data system: inter- and intraobserver variability in feature analysis and final assessment. *AJR Am J Roentgenol*. 2000;174:1769–1777. doi: 10.2214/ajr.174.6.1741769.
19. Beales L, Wolstenhulme S, Evans JA, West R, Scott DJ. Reproducibility of ultrasound measurement of the abdominal aorta. *Br J Surg*. 2011;98:1517–1525. doi: 10.1002/bjs.7628.
20. Drezner JA, Asif IM, Owens DS, Prutkin JM, Salerno JC, Fean R, Rao AL, Stout K, Harmon KG. Accuracy of ECG interpretation in competitive athletes: the impact of using standised ECG criteria. *Br J Sports Med*. 2012;46:335–340. doi: 10.1136/bjsports-2012-090612.
21. Exeter DJ, Elley CR, Fulcher ML, Lee AC, Drezner JA, Asif IM. Standardised criteria improve accuracy of ECG interpretation in competitive athletes: a randomised controlled trial. *Br J Sports Med*. 2014;48:1167–1171. doi: 10.1136/bjsports-2013-093360.
22. Smith-Bindman R, Chu PW, Miglioretti DL, Sickles EA, Blanks R, Ballard-Barbash R, Bobo JK, Lee NC, Wallis MG, Patnick J, Kerlikowske K. Comparison of screening mammography in the United States and the United kingdom. *JAMA*. 2003;290:2129–2137. doi: 10.1001/jama.290.16.2129.
23. Jacomelli J, Summers L, Stevenson A, Lees T, Earnshaw JJ. Impact of the first 5 years of a national abdominal aortic aneurysm screening programme. *Br J Surg*. 2016;103:1125–1131. doi: 10.1002/bjs.10173.
24. Ashton HA, Buxton MJ, Day NE, Kim LG, Marteau TM, Scott RA, Thompson SG, Walker NM; Multicentre Aneurysm Screening Study Group. The Multicentre Aneurysm Screening Study (MASS) into the effect of abdominal aortic aneurysm screening on mortality in men: a randomised controlled trial. *Lancet*. 2002;360:1531–1539.
25. Altobelli E, Scarselli G, Lattanzi A, Fortunato C, Profeta VF. A comparison between Pap and HPV screening tests and screening methods. *Mol Clin Oncol*. 2016;5:348–354. doi: 10.3892/mco.2016.909.
26. Bassal R, Schejter E, Bachar R, Shapira H, Sandbank J, Supino Rosin L, Schvimer M, Cohen D, Keinan-Boker L. Cervical Pap screening among Israeli women, 2005-2010. *Arch Gynecol Obstet*. 2014;289:615–622. doi: 10.1007/s00404-013-3045-z.
27. Logan RF, Patnick J, Nickerson C, Coleman L, Rutter MD, von Wagner C; English Bowel Cancer Screening Evaluation Committee. Outcomes of the Bowel Cancer Screening Programme (BCSP) in England after the first 1 million tests. *Gut*. 2012;61:1439–1446. doi: 10.1136/gutjnl-2011-300843.
28. Seeff LC, Royalty J, Helsel WE, Kammerer WG, Boehm JE, Dwyer DM, Howe WR Jr, Joseph D, Lane DS, Laughlin M, Leypoldt M, Marroulis SC, Mattingly CA, Nadel MR, Phillips-Angeles E, Rockwell TJ, Ryerson AB, Tangka FK. Clinical outcomes from the CDC's Colorectal Cancer Screening Demonstration Program. *Cancer*. 2013;119(suppl 15):2820–2833. doi: 10.1002/cncr.28163.
29. Sharma S, Drezner JA, Baggish A, Papadakis M, Wilson MG, Prutkin JM, La Gerche A, Ackerman MJ, Borjesson M, Salerno JC, Asif IM, Owens DS, Chung EH, Emery MS, Froelicher VF, Heidbuchel H, Adamuz C, Asplund CA, Cohen G, Harmon KG, Marek JC, Molossi S, Niebauer J,

Pelto HF, Perez MV, Riding NR, Saarel T, Schmied CM, Shipon DM, Stein R, Vetter VL, Pelliccia A, Corrado D. International recommendations for electrocardiographic interpretation in athletes. *J Am Coll Cardiol*. 2017;69:1057–1075. doi: 10.1016/j.jacc.2017.01.015.

30. Sharma S, Drezner JA, Baggish A, Papadakis M, Wilson MG, Jordan M. Prutkin JM, La Gerche A, Ackerman MJ, Borjesson M, Salerno JC, Asif I, Owens D, Chung EH, Emery MS, Froelicher VF, Heidbuchel H, Adamuz C, Asplund CA, Cohen G, Harmon KG, Marek JC, Molossi S, Niebauer J, Pelto HF, Perez MV, Riding NR, Saarel T, Schmied CM, Shipon DM, Stein R, Vetter V, Pelliccia A, Corrado D. International recommendations for electrocardiographic interpretation in athletes. *J Am Coll Cardiol.* 2017;69:1057–1075. doi: 10.1016/j.jacc.2017.01.015.

Circulation
Cardiovascular Quality and Outcomes

Inter-Rater Reliability and Downstream Financial Implications of Electrocardiography Screening in Young Athletes

Harshil Dhutia, Aneil Malhotra, Tee Joo Yeo, Irina Chis Ster, Vincent Gabus, Alexandros Steriotis, Helder Dores, Greg Mellor, Carmen García-Corrales, Bode Ensam, Viknesh Jayalapan, Vivienne Anne Ezzat, Gherardo Finocchiaro, Sabiha Gati, Michael Papadakis, Maria Tome-Esteban and Sanjay Sharma

Circ Cardiovasc Qual Outcomes. 2017;10:e003306
doi: 10.1161/CIRCOUTCOMES.116.003306
Circulation: Cardiovascular Quality and Outcomes is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Print ISSN: 1941-7705. Online ISSN: 1941-7713

SUPPLEMENTAL MATERIAL

Supplemental Table 1: Cost of secondary investigations

	Cost * ($)
Specialist cardiology outpatient clinic attendance for further evaluation and secondary investigations†	249
Echocardiography	112
Exercise stress test	258
Holter monitor	258
Cardiac MRI	319
Family screening ‡	996

* NHS payment system tariffs. https://www.england.nhs.uk/resources/pay-syst/. Costs are incurred in Great Britain pounds (£), but are presented in US dollars ($) with a currency exchange rate of £1=$1.52 at the time of manuscript preparation.

† Includes repeat 12-lead ECG.

‡ Specialist cardiology clinic outpatient attendance and 12-lead ECG for 4 family members.

Supplemental Table 2: Inter-observer reliability among cardiologists for the presence of specific ECG abnormalities

		Long QT interval	Short QT interval	T-wave inversion	ST depression *	Q waves	Axis deviation/ atrial enlargement/ RVH †
Inexperienced cardiologists	2010 ESC	0.21 (0.08-0.29)	0.15 (0.01-0.31)	0.43 (0.21-0.54)	0.12 (0.01-0.32)	0.10 (0.01-0.34)	0.15 (0.02-0.38)
	Seattle	0.44 (0.21-0.57)	n/a	0.51 (0.28-0.71)		0.17 (0.01-0.39)	
	Refined	0.44 (0.21-0.57)	n/a	0.54 (0.30-0.67)		0.16 (0.02-0.41)	
Experienced cardiologists	2010 ESC	0.31 (0.15-0.52)	0.43 (0.23-0.67)	0.54 (0.41-0.74)	0.13 (0.01-0.30)	0.24 (0.05-0.41)	0.43 (0.21-0.58)
	Seattle	0.60 (0.43-0.78)	n/a	0.61 (0.45-0.84)		0.18 (0.01-0.37)	
	Refined	0.60 (0.43-0.78)	n/a	0.56 (0.43-0.81)		0.24 (0.03-0.45)	

Inter-observer agreement presented as Cohen kappa [κ (95%CI)]. * ST depression applies to all criteria; † Axis deviation/atrial enlargement/RVH pooled together; RVH: Right ventricular hypertrophy

Supplemental Table 3: Kappa agreement for the 2010 ESC recommendations by population groups as suggested by the heterogeneities identified by the bivariate logistic model fitted for perfect agreement as the positive variables.

	Inexperienced cardiologists in agreement				Experienced cardiologists in agreement
	White	Other	Black	All ethnicities	All ethnicities
Female	κ = 0.21 (0.14-0.37)	κ = 0.00 (-0.26-0.38)	κ = 0.05 (-0.10-0.28)	κ = 0.20 (0.10-0.30)	κ = 0.57 (0.51-0.69)
Male	κ = 0.14 (0.06-0.20)	κ = 0.34 (0.25-0.40)	κ = 0.07 (0.06-0.13)	κ = 0.13 (0.09-0.19)	κ = 0.34 (0.27-0.40)