









Should lethal arrhythmias in hypertrophic cardiomyopathy be predicted using non-electrophysiological methods?

Richard Saumarez ^{1*}, John Silberbauer ², Jack Scannell ³,
Mariusz Pytkowski ⁴, Elijah R. Behr ⁵, Timothy Betts ⁶, Paulo Della Bella ⁷,
and Nicholas S. Peters ⁸

¹Retired, Waterbeach, CB25 9HW, UK; ²Department Cardiology, Royal Sussex Hospital, Eastern Road, Brighton BN2 5BE, UK; ³The Bayes Centre, University of Edinburgh, Edinburgh EH8 9BT, UK; ⁴Department of Cardiology, Narodowy Instytut Kardiologii, ul Alpejska 42, 04-628 Warsaw, Poland; ⁵St George's, Blackshaw Road, London SW17 0RE, UK; ⁶Department of Cardiology, John Radcliffe Hospital, Oxford OX3 9DU, UK; ⁷Department of Cardiology, San Raffaele Hospital, IT 20133, Milan, Italy; and ⁸Department of Cardiology, Hammersmith Hospital, Imperial College, London W12 0HS, UK

Received 24 November 2022; accepted after revision 20 January 2023

Abstract

While sudden cardiac death (SCD) in hypertrophic cardiomyopathy (HCM) is due to arrhythmias, the guidelines for prediction of SCD are based solely on non-electrophysiological methods. This study aims to stimulate thinking about whether the interests of patients with HCM are better served by using current, 'risk factor', methods of prediction or by further development of electrophysiological methods to determine arrhythmic risk. Five published predictive studies of SCD in HCM, which contain sufficient data to permit analysis, were analysed to compute receiver operating characteristics together with their confidence bounds to compare their formal prediction either by bootstrapping or Monte Carlo analysis. Four are based on clinical risk factors, one with additional MRI analysis, and were regarded as exemplars of the risk factor approach. The other used an electrophysiological method and directly compared this method to risk factors in the same patients. Prediction methods that use conventional clinical risk factors and MRI have low predictive capacities that will only detect 50–60% of patients at risk with a 15–30% false positive rate [area under the curve (AUC) = ~0.7], while the electrophysiological method detects 90% of events with a 20% false positive rate (AUC = ~0.89). Given improved understanding of complex arrhythmogenesis, arrhythmic SCD is likely to be more accurately predictable using electrophysiologically based approaches as opposed to current guidelines and should drive further development of electrophysiologically based methods.

Keywords

Hypertrophic cardiomyopathy • Sudden death • Prediction • Receiver operating characteristic

Introduction

The majority of sudden cardiac deaths (SCD) in hypertrophic cardiomyopathy (HCM) are due to malignant arrhythmias¹ and so are an electrophysiological problem, but the European Society of Cardiology guidelines recommends prediction of an arrhythmia, and implantable cardioverter defibrillator (ICD) implantation, through non-electrophysiological means, a combination of clinical 'risk factors'² (RFs) based on a regression model.³

If RFs are used to predict SCD in HCM, there presumably must be a close relationship between the mechanism of arrhythmogenesis and the RFs in question. The components of the electrophysiological substrate for ventricular arrhythmias, including ventricular fibrillation

(VF), have been known for many years. They are temporary conduction block with slowed conduction that gives tissue ahead of the re-entrant wavefront sufficient time to repolarize. It is also known that the substrate arises dynamically.⁴ In contrast, the relationship between clinical RFs and the causal chain of events underlying SCD is unclear, but were there a strong causal chain, RFs should have similar predictive capacity to electrophysiological approaches in assessing patients' risk of SCD.

The purpose of this paper is to scrutinize models for the prediction of SCD, their clinical consequences, and to explain intelligibly the statistical reasoning and how they translate to informing clinical decisions. We conclude that these methods, which are based on regression between clinical RFs and risk, have predictive capacities that are at the lower limit of acceptability and that new approaches through

* Corresponding author. Tel: +44 1223 863489. E-mail address: rcsaumarez@btinternet.com

© The Author(s) 2023. Published by Oxford University Press on behalf of the European Society of Cardiology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

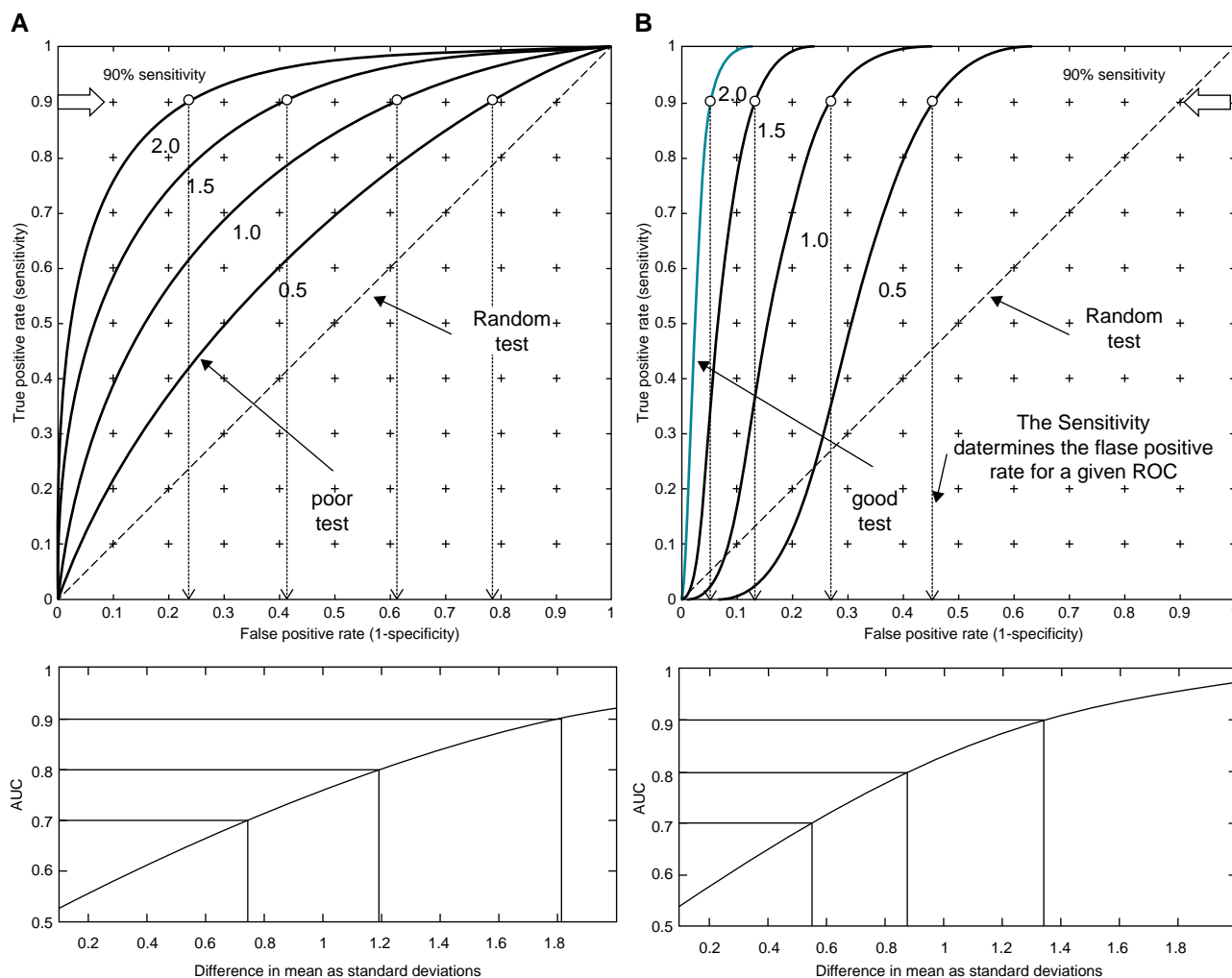


Figure 1 ROCs from two normal distributions whose means are separated by 0.5, 1.0, 1.5, and 2.0 SDs of the lowest valued distribution. (A) The SDs are the same and (B) the SD of the higher valued distribution is reduced to 0.3, thus decreasing the overlap of the distributions and improving the predictive value. The "poor test" curve marked in (A) is an example of poor predictive capacity (AUC = 0.63), while the "good test" in (B) is an example of high predictive capacity (AUC = 0.97). The points of 90% sensitivity are marked on the curves, and their corresponding false positive rates are shown by the arrows projected on the abscissa. The lower curves show how the AUC varies with the difference in means.

electrophysiology may be more promising and warrant further development. The RF studies analysed have enough primary data to allow statistical analysis and are regarded as exemplars of the RF approach. One is the first multiple RF study for prediction of SCD,⁵ another in which a multiple regression model³ was further tested and validated⁶ and a model combining late gadolinium enhancement (LGE) on MRI combined with RFs.⁷ A further study using multiple RFs in paediatric patients with a similar methodology to O'Mahony et al. and O'Mahony et al.^{3,6} is analysed.⁸

These RF studies are compared to an electrophysiological method, 'paced electrogram fractionation analysis (PEFA)', described in multiple prior studies^{9,10,11,12} and was designed to expose explicitly the electrophysiological (EP) effects of disarray that could form one component of the re-entrant EP substrate: delayed intramyocardial conduction as a result of slowed or tortuous conduction and/or temporary block, which allows repolarization of refractory tissue before being re-excited by the advancing re-entrant wavefront. Although disarray had been speculatively associated with a re-entrant substrate,^{13–16} the challenge was to demonstrate its effects. The method involves detection of small,

fractionated potentials at the end of the paced electrogram that appear and are delayed as the coupling interval of the extrastimuli is reduced (see Saumarez et al.⁹ and Figure 3). This fractionation is not seen during sinus rhythm and shows the capacity for dynamic delayed conduction as the myocardium is perturbed. The PEFA technique quantifies these effects by using two parameters, the increase in electrogram duration and the S1S2 coupling interval at which delay increases, and so is a physiological, rather than a regression, model. This was developed using 143 patients and resulted in a discriminant criterion with 18 VF patients (8 prospective) and 13 non-VF patients being test positive and 2 VF patients and 110 non-VF patients being test negative. (Subsequent data showed that prospective and retrospective PEFA values in VF patients were from the same distribution.) This criterion was tested prospectively in a study of 179 patients designed to determine if PEFA had superior predictive capacity to RFs by measuring both in the same patients,¹⁷ and 2 further patients who had events following publication are included in the analysis below. These studies were designed to use the minimum number of patients (since they are invasive) to achieve statistically significant results. However, they form a large electrophysiological series, and they

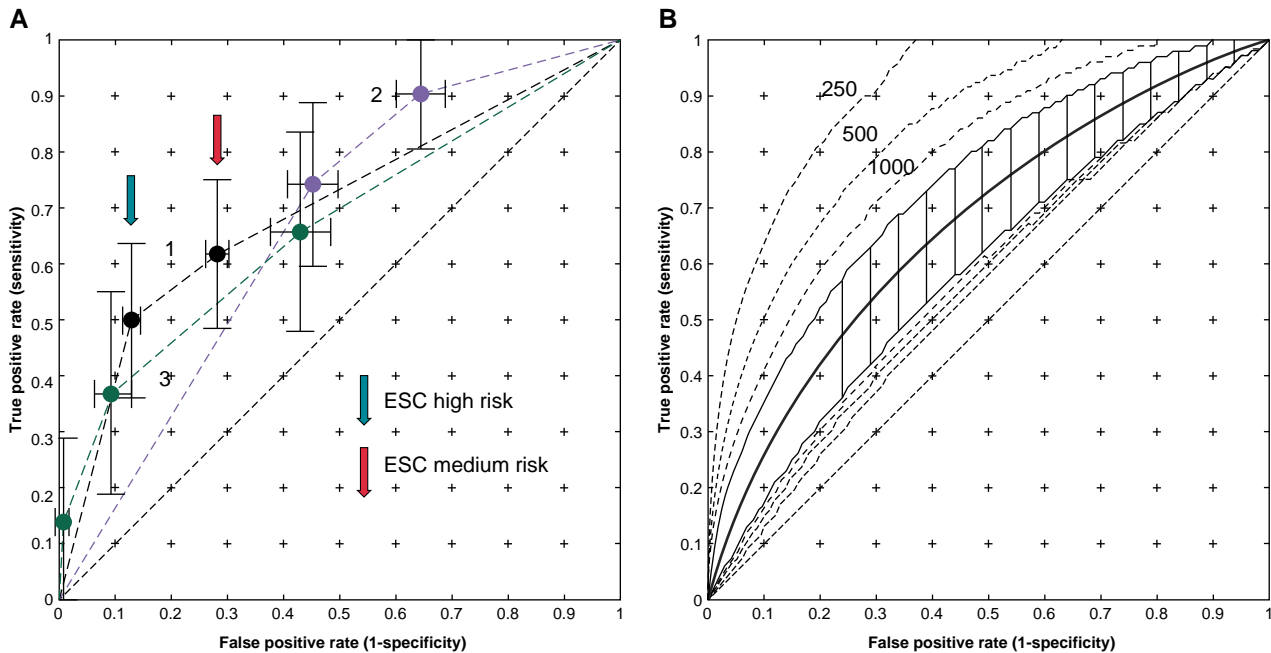


Figure 2 Mean convex hull ROCs for non-electrophysiological risk prediction together with the 2.5% and 97.5% confidence limit. (A) Curve 1 — O'Mahony 'Model A', Curve 2 Monte Carlo analysis using estimated 5-year survival for Elliott et al.,⁵ and Curve 3 from Norrish.⁸ (B) O'Mahony (2017) 'Model B' with $\pm 2.5\%$ limits for 2147 patients. The dashed lines show the limits when calculated for 250, 500, and 1000 patients.

suggest that electrophysiological methods with further development may give substantially better risk prediction. The purpose of this paper is to stimulate research into, and to provoke discussion over, the merits of risk factor and EP-based risk prediction.

Prediction

'Prediction is very difficult, especially if it's about the future' (Niels Bohr). Prediction of SCD risk and ICD implantation revolve around the statistical models that have been used to determine the relationship between the risk and the variables measured to infer that risk. Thus, ICD prescription is related to the performance of the statistical models, and so the methods used in this paper are explained so as to be intelligible to the general reader.

A useful tool for analysing this problem is the 'receiver operating characteristic' (ROC) curve, which is a method of quantifying the overlap of two sets of data. A predictor, which varies from, say, zero to 100, is calculated for a population. For each value of the predictor, the number of events and non-events below that value is determined, which allows the sensitivity (true positive rate) to be calculated as well as the false positive rate (1-specificity). This yields a curve that shows what fraction of patients will be truly and falsely identified for any value of the predictor. Examples of these curves are shown in Figure 1. The critical measurement from the point of SCD is the number of false positives for a given true positive rate. If one is trying to detect 90% of events, then the false positive rate can be determined from the intersection of the ROC (Figure 1) with 90% sensitivity projected on the abscissa. There is a difficulty with ROCs when the data are categorical rather than continuous. While on a continuous curve, one can determine the false positive rate for any sensitivity, when using an ROC based on categories, one can only determine the sensitivity and false positive for each category, and it is meaningless to ask, if one is using the number of clinical RFs as categories, 'what is the sensitivity of one and a half RFs'?

Thus, the ROCs for categorical data are plotted as discrete points rather than a continuous curve.

The performance of prediction is measured by the AUC. 'In general, an AUC of 0.5 suggests no discrimination (i.e. ability to diagnose patients with and without the disease or condition based on the test), 0.7 to 0.8 is considered acceptable, 0.8 to 0.9 is considered excellent, and more than 0.9 is considered outstanding.'¹⁸ While it is accepted that an AUC of 0.9 represents an 'outstanding' test, this lacks intuitive meaning. As an example, UK adult height data¹⁹ were used to calculate the ROC for distinguishing men from women on the basis of height. The AUC is 0.90 and so (astonishingly) the majority of medical tests are substantially less accurate than distinguishing men from women on the basis of their height! However, this is age dependent and comparing the heights of 16-year-old girls with those of 14-year-old boys, the AUC is 0.5, 0.7 for 15-year-old boys, and 0.83 for 16-year-old boys, and these comparisons should be borne in mind when considering the studies described below. For normal distributions with equal variances, an AUC = 0.7 implies a difference in the means of $\sim 0.75 \sigma$, and an AUC = 0.9 corresponds to a difference in means of 1.8σ . If the variance of the SCD group is less than the non-SCD group, the AUC will be increased for a given difference between the means of groups as shown in Figure 1, and this effect is important in enhancing predictive accuracy.

Given that an ROC has been calculated from a single set of data, how can its errors be estimated? There are two approaches, which are widely used in statistics, diagnostics, and artificial intelligence: the first is the 'bootstrap',²⁰ which is a method of making a generalization about the population from which a single group is drawn and involves generating a new data set by random sampling of the data. The second is to generate a sample of the population based on random sampling of its underlying statistics and then apply the risk model to the sample ('Monte Carlo' method). By repeating these methods many times, the limits of the ROC can be established.

A problem with using data taken from publications is that it is rarely complete. One major issue is that each patient has been followed for

different lengths of time, 'censored', and this requires correction. The EP and RF data in Saumarez et al.¹⁷ are not censored, each patient was followed for 5 years, and the ROC can be constructed directly. In censored data (where the length of follow-up of each patient is not specified), an indirect approach is used. The mortality of each group in a study at 5 years is quoted,^{5,6,7,9} having been calculated through Kaplan–Meier (KM) statistics, which is specifically designed to calculate mortality rates while compensating for censoring. Given the number of patients in each group and their mortality, the ROC can be calculated and its limits determined by Monte Carlo simulation (Figure 2). Receiver operating characteristics generated from raw, censored data and those calculated from KM result in difference in an AUC of ~0.03. The AUC for censored data can be calculated and is the concordance or 'C-index'²¹ and is quoted in the studies discussed below.

Methods

Each ROC was calculated with convex hull correction and the confidence limits determined via a bootstrap or from Monte Carlo simulation with 10 000 simulations. For categorical ROCs, the results of simulation were checked analytically. The *t*-bootstrap was used to determine significance levels of the bootstrapped AUC. The ROC for Elliott et al.⁵ was taken from RF frequencies and SCDs in each group in the text, and the Monte Carlo analysis was performed using the estimated 5-year mortalities of each group. The survival curves in O'Mahony et al.⁶ were used to calculate the survival rates of each group (high, medium, and low) and their proportions of the population (Table 4⁶). The ROC was calculated at 5 years (i.e. non-censored) by simulation of 2147 patients with these proportions ('Model A'). The regression equation in O'Mahony et al. and O'Mahony et al.^{3,6} was used to compute the ROC in Figure 2B, and the data given in Table 2 of O'Mahony et al.⁶ were used for Monte Carlo analysis ('Model B'). The maximum wall thickness, left atrial diameter, and age were assumed to be normally distributed and the incidence of non-sustained ventricular tachycardia (NSVT), a family history of SCD, and the presence of syncope binomially distributed and changed for each simulated ROC to create a wide range of risk profiles for the population. The LV outflow tract gradient was assumed to be from a gamma distribution ($\alpha = 1.75$, $\beta = 7.0$ fitted to conform to the mean and extrema quoted). A sensitivity analysis was performed by varying the incidence of the family history of sudden death (FHSD), NSVT, and syncope from 5 to 35% of the population since these three predictors dominate the regression function. The confidence limits were also calculated for populations of 250 500 and 1000 patients as well as the 2147-patient study population. The data from Norrish⁸ were analysed in a similar way to Model A in O'Mahony et al.⁶ using the initial proportions of the 0–4%, 4–6%, and >6% mortality patient groups.

The LGE data were taken from Chan et al.⁷ and divided in the four groups specified, 0 LGE, 1–9, 10–19, and ≥ 20 , and these classes, together with data in Table 2, were used to form the censored ROC, together with a ROC for the effect of risk factors. The convex hull AUC distribution was determined for RFs with random association with 37 events. A Monte Carlo model, with 1293 subjects, was constructed from the distribution of LGE, which was assumed among the 42% of patients with LGE to be from an exponential distribution ($\lambda = -8.6$, median 6, and quartiles 3 and 12). A sensitivity analysis was performed by perturbations of this distribution. The survival from SCD was obtained by interpolation of the 5-year event rate of patients without RFs (Table 3 in Chan et al.⁷) using a cubic polynomial. The RFs of each subject were computed from their frequencies in Table 1⁷ and assumed to be randomly distributed throughout the population. The increases in hazard due to risk factors were taken from Table 5,⁷ and their potentiation by LGE was applied as a continuous function of 1.5/10% increase in LGE (Table 5⁷) was applied as Cox proportional hazards to the interpolated survival curve, thus enabling the probability of survival of each simulated 'patient' to be computed.

The electrophysiological and risk factor data from Saumarez et al.¹⁷ have been recalculated using similar methods except that the electromagnetic interference in the signals, which varies between different laboratories and at different times in the same laboratory, has been reduced to a standard level using the discrete wavelet transform²² and Bayesian discrimination. During the 5-year follow-up period, two further events occurred in patients with no clinical risk factors and a high degree of fractionation, and these are

included in the analysis. The RF ROC was generated using the method of Elliott et al.,⁵ and its limits were obtained by direct two-dimensional integration of its smoothed distributions, and comparison of the EP and risk factor methods in the same patient was performed via the ROC bootstrap.

Results

Figure 2A shows the ROCs for RFs in the case of number of RFs for Elliott et al.⁵ and survival according to low-, medium-, and high-risk groups for the O'Mahony 'Model A'⁶ and for Norrish.⁸ The sensitivities for the O'Mahony Model A are 50% and 63%, respectively, with false positive rates of 14% and 30% (marked in Figure 2 with "high" and "medium" risk arrows, respectively), which correspond to the ESC guidelines² for implantation and the paediatric study by Norrish,⁸ having sensitivities of 0.75 and 0.9 for false positives of 0.45 and 0.65, respectively. Since these are not continuous variables, no further improvement is possible, and to improve on these sensitivities, the entire population would require an ICD. Figure 2B shows the ROC for O'Mahony 'Model B,' and the dashed lines show the confidence bands obtained when applied to a population of 250, 500, and 1000 patients. The AUC for the model with 2147 patients is 0.67, and the upper limits of the AUCs for 250, 500, and 1000 patients are 0.9, 0.84, and 0.79, respectively. This wide variation of confidence limits explains the results from studies using the method of O'Mahony et al. and O'Mahony et al.^{3,6} with a relatively small number of patients. The AUC of the O'Mahony Model A is 0.71, which is compatible with the C-index of 0.7 quoted in O'Mahony et al.,⁶ the AUC of the Norrish model was 0.67, which is compatible with quoted C-index of 0.69,⁹ and that of the Elliott Monte Carlo model is 0.68. The ROCs are similar, which is due to the methods of O'Mahony et al.³ and O'Mahony et al.⁶ being derived from that developed in Elliott et al.,⁵ and the regression equation in O'Mahony et al. and O'Mahony et al.^{3,6} is dominated by predictors: NSVT, FHSD, LV wall thickness, and syncope, which were used in Elliott et al.⁵ The sensitivity analysis showed a change in an AUC of less than 0.03 for a 5–30% change in prevalence of NSVT, syncope, and FHSD.

Figure 3A shows the ROCs and bootstrapped limits comparing EP and RFs in the same patients. The EP method, which has a continuous variable predictor, has better predictive capacity than risk factors, suggesting that 20% (11–30%) of the population would need implantation to protect 90% of the patients who would suffer events. The difference in convex hull AUCs (0.89 vs. 0.71) is significant at $P < 0.02$ (DeLong) and 0.01 (*t*-bootstrap) for a Type I error of 0.05. The RF curve is indistinguishable from that of Elliott et al. and O'Mahony et al.^{5,6}

Results from Chan et al.⁷ are shown in Figure 4. The raw, censored, ROC curve for LGE (2) is plotted in Figure 4A and has an AUC of 0.67, while the risk factors (1) have no predictive capacity (AUC = 0.521, convex hull AUC of 0.55; Monte Carlo range for random events: 0.5 to 0.58). The ROC using 5-year mortality and LGE with risk factors but no LGE enhancement of risk is shown in Curve 3 (AUC = 0.66), and the result of correction of LGE computed risk by additional interaction between risk factors and LGE is shown in Curve 4, with AUC = 0.7 (the AUC is quoted as 0.74, page 489 in Chan et al.⁷). The AUC of the latter varies from 0.68 to 7.6 when computed with the upper and lower limits of the risk factor correction. Curve 5 is calculated from the data in Figure 2B of Chan et al.⁷

Finally, Figure 5 shows the limits of the AUC estimates. The distribution calculated from the EP study¹⁷ is distinct. The mean AUCs for the RF methods are 0.67–0.72, while the mean AUC for the EP method is 0.89 and at least 0.8.

Discussion

The results of this statistical treatment covering the five studies suggest that prediction methods that use conventional clinical risk factors and MRI have low predictive capacities that will only detect 50–60% of patients at risk with a 15–30% false positive rate (AUC = ~0.7), while the

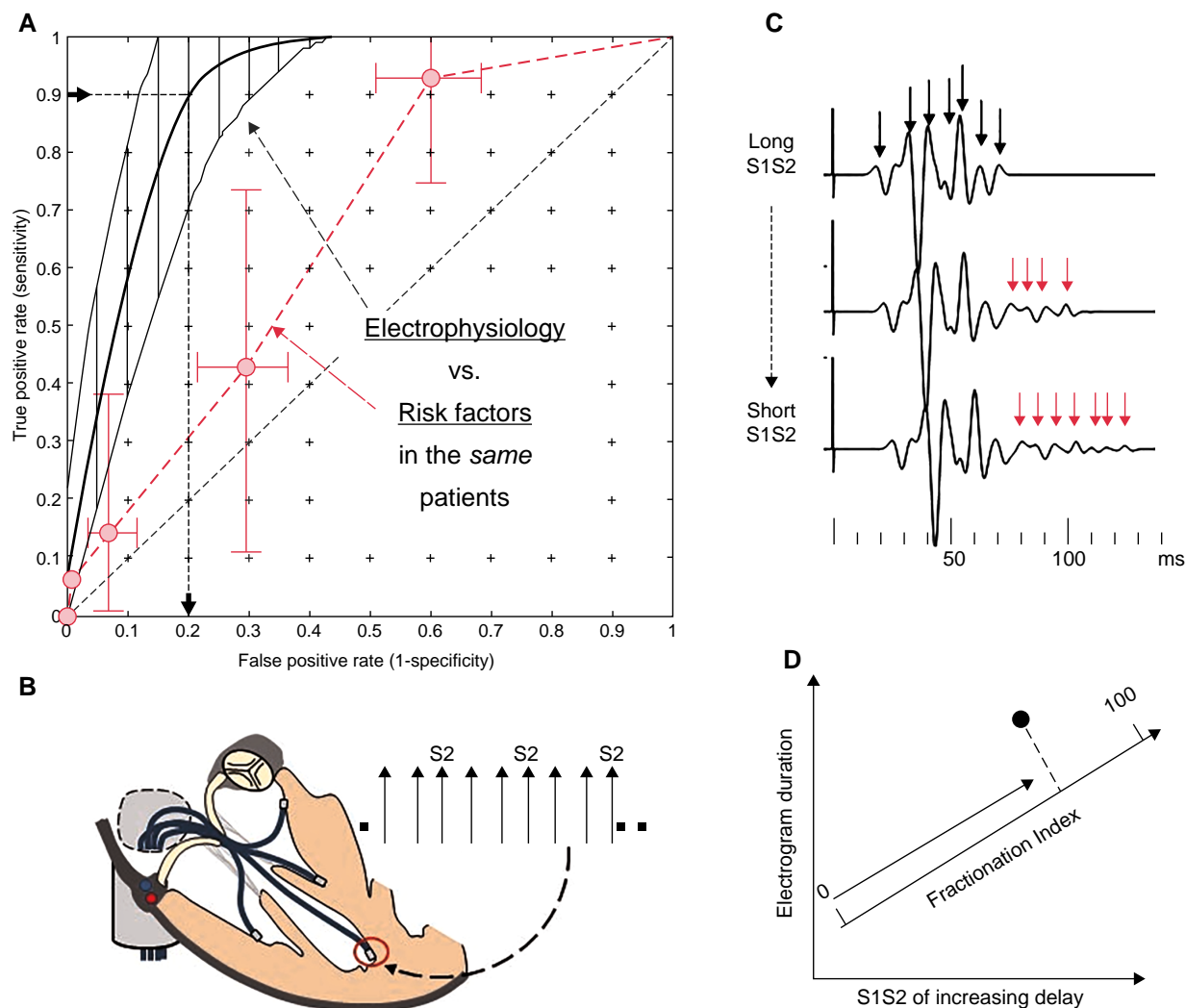


Figure 3 (A) The mean ROC and bootstrapped limits for electrophysiological method (black), which is a continuous variable and so can be plotted as a curve. The non-electrophysiological (RF) method (red) is measured in the same patients and plotted as number of RFs and the limits obtained by integration of the 2D bootstrap. Data from Saumarez et al.¹⁷ The false positive rate for 90% of event detection for EP is shown. (B) Electrodes are placed at four sites in the right ventricle (apex, mid-septum, RVOT, and free wall). Each site is paced in turn with a decelerating sequence with an S2 inserted every third beat. The pacing sequence is repeated at each site with recordings made from the remaining electrodes. (C) The electrograms are processed to identify each potential, and in high-risk patients, the electrograms become prolonged with multiple delayed potentials as the S1S2 interval is decreased (arrowed). These are quantified by determining the S1S2 at which electrogram delay increases and the overall increase in electrogram duration. (D) These are averaged for all the runs to yield a single observation, which is projected on a 'fractionation axis' (formed prospectively by the principal component of the data from the initial studies).

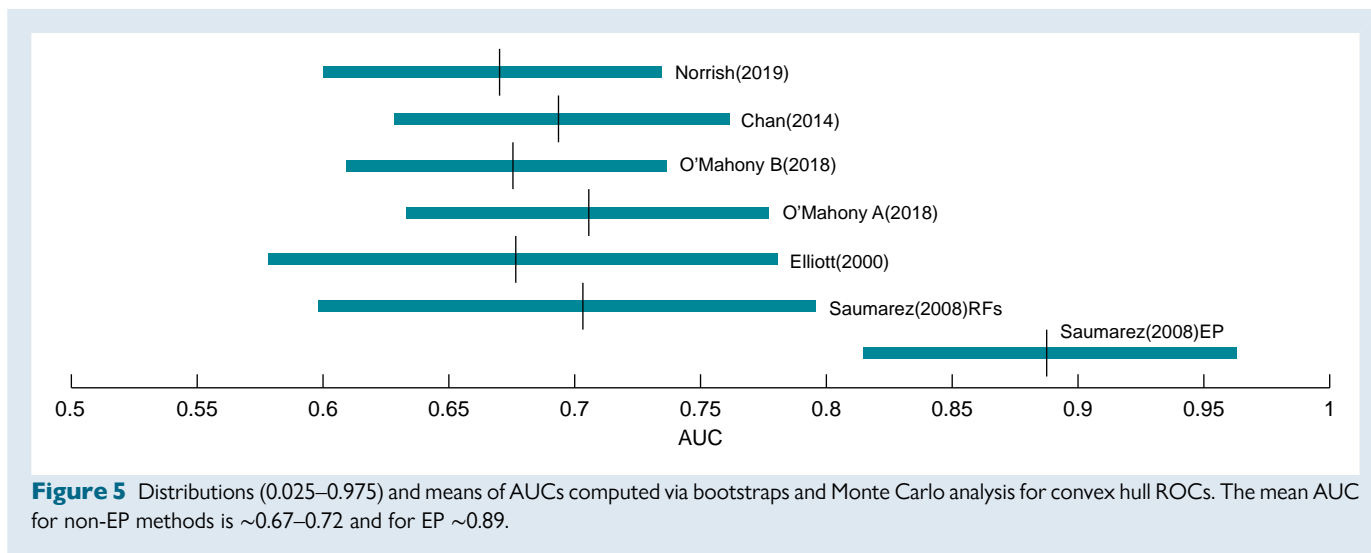
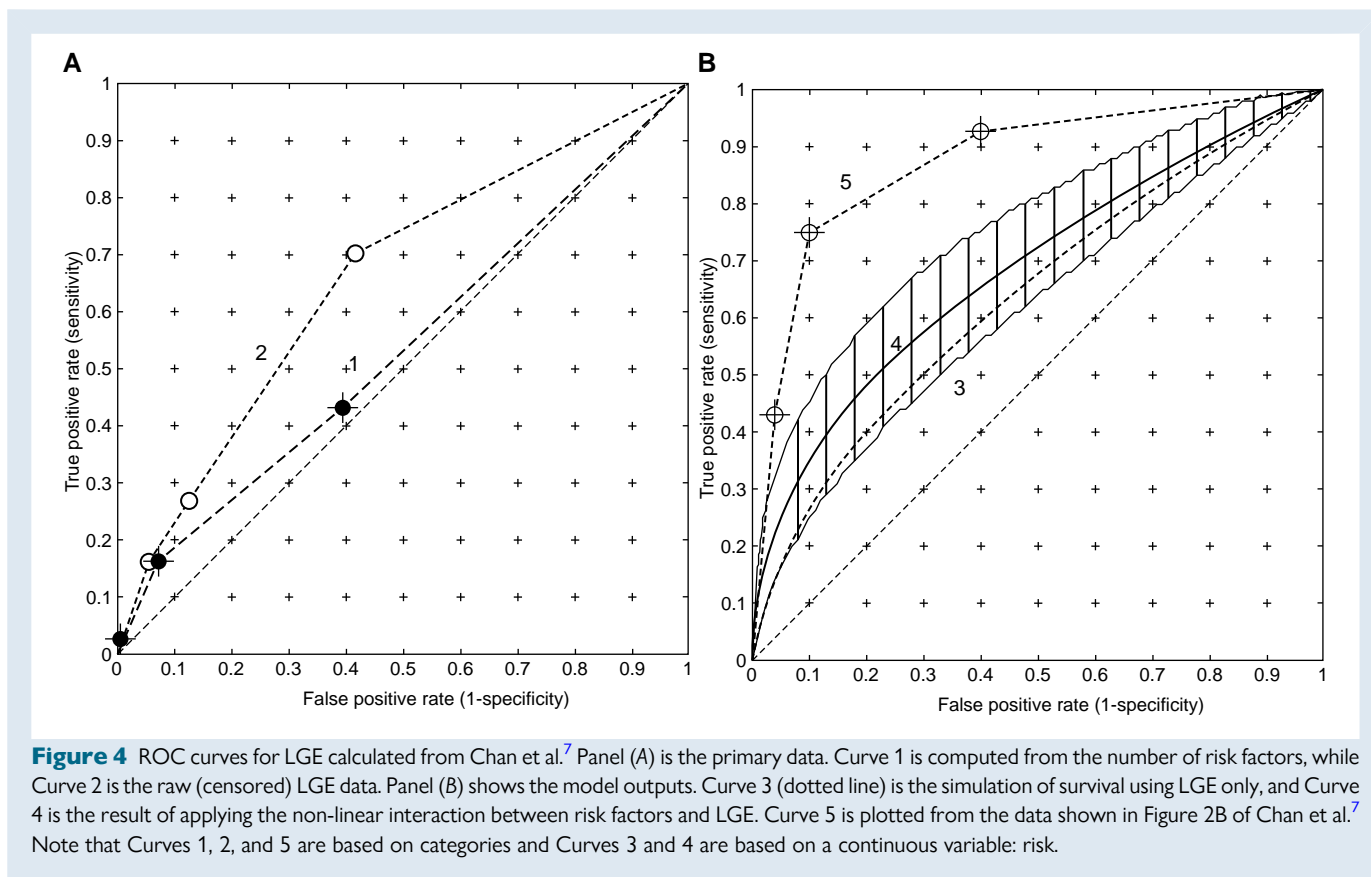
electrophysiological method, evaluated in small but statistically rigorous studies, detects 90% of events with a 20% false positive rate (AUC = ~0.89), justifying the re-examination and development of the electrophysiological methods.

Does identification of a high-risk group necessarily imply prediction?

A common misconception is that if a subgroup of patients can be identified with a high incidence of SCD, this necessarily implies good prediction. Consider 2600 patients in which there is a subgroup of 200 'high-risk' patients with an SCD incidence of 10% at 5 years, a subgroup of 600 with a 4% SCD incidence, while the remainder have an SCD

incidence of 1.5%. This entire group is followed for 5 years, i.e. non-censored. Formal comparison of the survival curves using the Mantel-Haenszel test leads to a highly significant result ($P < 0.001$).

At 5 years, a mean of 27 patients in the low-risk subgroup will have died, 24 in the medium-risk subgroup, as opposed to 20 in the high-risk group, so only 20 out of 71 deaths will have occurred in the high-risk subgroup giving a sensitivity of ~28%. The ROC has an AUC of 0.66, such that if the high- and medium-risk patients received an ICD, 44 patients would be protected and 27 would die at the expense of 736 non-discharging ICDs. If, however, the SCD rate in the low-risk subgroup is lowered to 0.2%, the predictive capacity of the high- and medium-risk groups is substantially increased with an AUC of 0.84. Thus, apparent high predictive ability, on the basis of a high-risk subgroup, is dependent



on the size and mortality of the low-risk subgroup and may be substantially reduced by a large low-risk, but not no-risk, group. Had there been a formal hypothesis in which predictive measures are explicitly stated, for example, the lower limit of the AUC will be greater than a specified value, this study would fail.

Adequacy of statistical power

Construction of regression models require, as a rule of thumb, 10 events for every predictor in the model, and therefore database studies^{3,5,6,7,8} which are used to construct regression models must have a large number of patients to achieve a numerically reliable regression

model irrespective of the predictive capacity of the model in question. Therefore, one should not automatically assume that a study containing a larger number of patients is a 'better' study than one with a smaller number without qualifying the predictive capacity. While the paced fractionation method¹⁷ was designed using substantially less patients (179), Figures 3 and 5 show that it had adequate statistical power because it was measuring large effects.

Differences in recruitment and risk factors

Patients in the PEFA studies¹⁷ were recruited in order of presentation, and this led to a lower proportion of patients with zero risk factors than

database studies, and this may have been influenced by the report that patients with zero or 1 RFs were low risk⁵ although more than half the SCD occurred in these groups.⁵ However, this does not invalidate the physiological relationship between slowed conduction and SCD. The regression equation used in O'Mahony et al., O'Mahony et al., Maron et al., Zegkos et al., Dong et al., Vriesendorp et al., Fernández et al., and Leong et al.^{3,6,23–29} is dominated by four factors, NSVT, syncope, LV wall thickness, and FHSD, and these RFs were measured in the PEFA study¹⁷ although LV thickness was recorded as a categorical variable. The similarity between the bootstrapped ROCs and AUCs for Elliott et al., O'Mahony et al., and Saumarez et al.^{5,6,17} shown in Figure 5 suggests that the results not substantially affected by RF selection.

'All models are wrong—some are useful': what is a useful level of prediction?

The level of prediction by RFs with AUCs ~0.67–0.74 from five studies is at the lower limit of the definition of acceptability¹⁸ and implies that only half of the at-risk patients are identified for a 20% implantation rate. This was reported in a database study,²³ which appears to be within the limits of O'Mahony et al.,⁶ the ACC/AHA and ESC guidelines have poor specificity²⁴ for a reasonable sensitivity, the ESC guidelines had an AUC of 0.68 in a study of 511 Chinese patients,²⁵ and a study comparing the ESC guidelines to the ACC guidelines had a C-index of 0.69.²⁶ However, with an AUC of 0.89, the EP method appears to have useful predictive capacity. Since both the original study on which the ESC guidelines are based³ and the validation study⁶ have quoted C-index (Uno's) of 0.7 and the paediatric study 0.69⁸ (and 0.71 in a validation study²⁷), it is not surprising that there should be a wide variation in results when the guidelines are used predictively. Nevertheless, a study of 502 patients²⁸ reported using the ESC guidelines that the AUC for deaths was 0.92. This high AUC is due to zero mortality in the low-risk group (348 patients), while applying the estimates from O'Mahony et al. and O'Mahony et al.^{3,6} between 3 and 9 deaths or 3 to 31⁵ would be expected, which would have lowered the AUC. Events occurred in low-risk patients in most studies^{3,5,6,7,8,22,23,24} and are in the 2–4% event range, and so the high predictive accuracy in this study²⁸ stems from an unlikely event. The C-index of 0.86 for ESC method in 288 patients²⁹ is within the 2.5% limits of O'Mahony 'Model B' (288 patients, AUC = 0.57–0.87) (as shown in Figure 2B) and has a 50% sensitivity.

To devise a strategy for ICD implantation, the maximum acceptable rate of failed predictions must be defined and viewed against the number of implants that will deliver therapy. If 90% of potentially lethal arrhythmias, aborted by ICD implantation, is a reasonable goal, risk factors cannot achieve this without implanting the entire population. A judgement about what is effective prediction depends on the severity of the effect in question and the potential complications of therapy guided by the predictor as well as the actual SCD rate. For example, given a 3% 5-year SCD rate, the ESC implant guidelines would expect 15 deaths/1000 and 1 discharge for every 10 implantations (12/1000 deaths and 1 discharge/17 ICDs for 'consider implantation'), while the EP method would suggest 3/1000 deaths and 1 discharge/8 ICDs. One might argue that EP is too elaborate and expensive to use in the population, or alternatively one might argue that it identifies more patients with events and prevents life-changing and expensive ICD implantation in young patients who do not need one.

Therefore, the use of EP becomes a policy decision, which may vary between countries (e.g. there is a four-fold difference in total implantations between Germany and the UK), on the costs of preventing an SCD and the costs and the risks of EP testing, as well as those of non-discharging ICDs and their complications, which are roughly the same level as the appropriate discharge rate.^{30,31} In the case of SCD, and in view of potential ICD complications, a strategy based on marginal levels of prediction seems questionable, although ATP therapy for prodromal

VT and increased safety of the Subcutaneous ICD may mitigate complications and improve matching of the device to the patient.^{32,33}

Causality in prediction

Prediction can be developed by regression or causality. Regression^{3,5,6,8} attempts to link a set of variables through a function to an observation (SCD) without any inference about the mechanisms involved, and any assertion about the relationship between risk factors and the mechanisms of SCD is speculative. Causality uses a mechanistic model of the process to infer a result and may be more powerful *as it directs attention towards what variables should be measured* to perform prediction. The basis of the EP method¹⁷ is that lethal arrhythmias are re-entrant, and re-entrant arrhythmias require slowed conduction as the substrate for their initiation, and this could be caused by disarray. Thus, the hypothesis was raised that patients who are at risk of an arrhythmia should have slowed conduction velocities demonstrable in their ventricular myocardium without inducing an arrhythmia. This was tested by measuring conduction delay following paced stimuli at various sites in the ventricles,^{9,10} which raised the hypothesis for formal prospective testing¹⁷ as described in the introduction. However, the delays seen in multiple diseases,¹¹ including dilated cardiomyopathy (DCM), idiopathic VF, the long QT syndrome (LQTS),³⁴ and paroxysmal atrial fibrillation,³⁵ indicated that the mechanism of VF in HCM, assumed to be due to disarray, is a subset of a more general mechanism of arrhythmogenesis and the precise mechanisms in HCM are not fully understood.³⁶ The consistent association between fractionated paced electrograms and arrhythmias from multiple studies suggests that electrophysiology, rather than risk factors, measures something close to the causal chain of events in SCD, implying that substantially better prediction is possible.

A further concept raised by the EP studies is that recordings made in sinus rhythm do not show the effects associated with SCD. This is consistent with the arrhythmia substrate arising dynamically, with the perturbation of the myocardium that leads to an arrhythmia being mimicked by pacing, and it is therefore unlikely that meaningful measurements relating to risk can be made in resting sinus rhythm. This concept is supported by a recent study on the EP mechanisms in the initiation of induced VF, which directs further attention to the key determinants of SCD,³⁷ while the role of PES remains controversial,^{10,38,39} may be non-specific with aggressive stimulation, and does not reveal an arrhythmia mechanism *per se*. A recent study in a small number of patients has shown that ECG-I during and immediately after exercise (but not at rest) discriminates VF and non-VF patients on the basis of abnormal ventricular activation,⁴⁰ which is consistent with PEFA measurements. The relative merits of myocardial perturbation through exercise or pacing and ECG or intracardiac recordings remain an open question in studies designed not to induce an arrhythmia.

Testing prediction

Crucially, predictive models need testing. Clinical predictive models constructed from databases are rapidly evolving,^{41,42} but database use and selection has drawbacks, particularly of inhomogeneity and that key predictive variables may not be included in the database.^{43,44} Ideally patients should be recruited prospectively and the formal predictive hypothesis tested by predicting the patients' risk in advance as was the case in Saumarez et al.¹⁷ The use of databases to test a predictive hypothesis is possible and is easier to perform. However, if the advantage of having a large number of subjects from multiple databases is not reflected in a substantial improvement in prediction statistics, the data may be only distantly related to the effect in question and validation fails. A further important issue in validation and comparison of models is that different models, constructed from the same data, may yield substantially different results, as has recently been described in the computation of the reproduction number ('R') in the recent Covid pandemic.⁴⁵

A number of studies using MRI have suggested that there may be SCD markers, but it is unclear how these markers are related to the fundamental processes involved in arrhythmogenesis. While fibrosis is revealed, this gives no information about the potential for re-entrant pathways or change in tissue refractoriness. The study analysed here reported that LGE was associated with increased mortality in HCM,⁷ when combined with RFs (AUC = 0.74, page 489 in Chan et al.⁷). Despite the RFs in this study having no predictive capacity, *per se*, they apparently enhance the risk associated with LGE, although the predictive capacity remains low. The model that computes the data shown as Curve E in Figure 4 (Figure 2B in Chan et al.⁷) is not sufficiently well described to be analysed, and it is not clear how such enhancement of prediction can be related to the primary data. If MRI has potential for non-invasive risk prediction, it is essential that experiments are designed with prediction in mind, rather than identifying high- and low-risk groups, which can be misleading.

Given modern electrophysiological knowledge, the prediction of an arrhythmia by proxies is an extraordinary claim, and extraordinary claims require extraordinary evidence. This analysis challenges the current approach of using risk factors and imaging and implies they are the wrong predictors to achieve a high and clinically useful risk assessment and that their use to create guidelines for patient management is questionable.² The argument about non-EP proxies also applies to heart failure where the use of left ventricular ejection fraction is used as a predictor. Using data from Curtis' study of the relationship between prognosis and LVEF in 7788 patients,⁴⁶ the calculated AUC for SCD against LVEF is ~0.61, which, despite forming part of national guidelines, implies that it has little value for prediction and appropriate ICD implantation.⁴⁷

Finally, this study implies that the prediction of SCD and ICD implantation needs to be reassessed and refined, using statistical techniques which genuinely reflect prediction,⁴⁸ and that we need to refocus our efforts on the insights gained by cardiac electrophysiology^{17,37} into the mechanisms of arrhythmogenesis to develop rational, physiologically based, models for SCD prediction. The aim of this paper is to open this debate.

Conflict of interest: R.C.S. is a director of a company developing novel electrophysiological equipment for risk prediction in post-MI patients, and N.S.P. is a shareholder in the same company. All remaining authors have declared no conflicts of interest. P.D.B. and E.R.B. are editors of *EP Europace* and were not involved in the peer review process or publication decision.

Data availability

The data is taken from previous publications (Circulation, European Heart Journal).

References

1. Maron BJ. In: Zipes DP, Jalife J and Stevenson WG (eds.), *Ventricular arrhythmias in hypertrophic cardiomyopathy: sudden death, risk stratification, and prevention with implantable defibrillators*. *Cardiac Electrophysiology, from Cell to Bedside*. 7th edition. Philadelphia: Elsevier, 2018. p863–73.
2. Elliott PM, Anastasakis A, Borger MA, Borggrefe M, Cecchi F, Charron P et al. 2014 ESC guidelines on diagnosis and management of hypertrophic cardiomyopathy: the Task Force for the Diagnosis and Management of Hypertrophic Cardiomyopathy of the European Society of Cardiology (ESC). *Eur Heart J* 2014;**35**:2733–79.
3. O'Mahony C, Jichi F, Pavlou M, Monserrat L, Anastasakis A, Rapezzi et al. Hypertrophic cardiomyopathy Outcomes Investigators. A novel clinical risk prediction model for sudden cardiac death in hypertrophic cardiomyopathy (HCM risk-SCD). *Eur Heart J* 2014;**35**:2010–20.
4. Jackson N, Girzuraron S, Masse S, Nanthakumar K. In: Zipes DP, Jalife J and Stevenson WG (eds.), *Mechanisms of human ventricular tachycardia and human ventricular fibrillation*. *Cardiac Electrophysiology, from Cell to Bedside*. 7th edition. Philadelphia: Elsevier, 2018. p457–98.
5. Elliott PM, Poloniecki J, Dickie S, Sharma S, Monserrat L, Varnava A et al. Sudden death in hypertrophic cardiomyopathy: identification of high-risk patients. *J Am Coll Cardiol* 2000;**36**:2212–8.
6. O'Mahony C, Jichi F, Ommen SR, Christiaans I, Arbustini E, Garcia-Pavia P et al. International external validation study of the 2014 European Society of Cardiology guidelines on sudden cardiac death prevention in hypertrophic cardiomyopathy (EVIDENCE-HCM). *Circulation* 2018;**137**:1015–23.
7. Chan RH, Maron BJ, Olivetto I, Pencina MJ, Assenza GE, Haas T et al. Prognostic value of quantitative contrast-enhanced cardiovascular magnetic resonance for the evaluation of sudden death risk in patients with hypertrophic cardiomyopathy. *Circulation* 2014;**130**:484–95.
8. Norrish G, Ding T, Field E, Ziolkowska L, Olivetto I, Limongelli G et al. Development of a novel risk prediction model for sudden cardiac death in childhood hypertrophic cardiomyopathy (HCM Risk-Kids). *JAMA Cardiol* 2019;**4**:918–27.
9. Saumarez RC, Camm AJ, Panagos A, Gill JS, Stewart JT, de Belder MA et al. Ventricular fibrillation in hypertrophic cardiomyopathy is associated with increased fractionation of paced right ventricular electrograms. *Circulation* 1992;**86**:467–74.
10. Saumarez RC, Slade AK, Grace AA, Sadoul N, Camm AJ, McKenna WJ. The significance of paced electrogram fractionation in hypertrophic cardiomyopathy. A prospective study. *Circulation* 1995;**91**:2762–8.
11. Saumarez RC, Grace AA. Paced ventricular electrogram fractionation and sudden death in hypertrophic cardiomyopathy and other non-coronary heart diseases. *Cardiovasc Res* 2000;**47**:11–22.
12. Saumarez RC, Chojnowska L, Derksen R, Pytkowski M, Sterlinski M, Huang CL et al. Sudden death in noncoronary heart disease is associated with delayed paced ventricular activation. *Circulation* 2003;**107**:2595–600.
13. Kawai C, Sakurai T, Fujiwara H, Matsumori A, Yui Y. Hypertrophic obstructive and non-obstructive cardiomyopathy in Japan. Diagnosis of the disease with special reference to endomyocardial catheter biopsy. *Eur Heart J* 1983;**4**:121–5.
14. Segawa I, Suzuki T, Kato M, Tashiro A, Satodate R. Relation between myocardial histological changes and ventricular tachycardia in cardiomyopathy: a study by 24-hour ECG-monitoring and endomyocardial biopsy. *Heart Vessels Suppl* 1990;**5**:37–40.
15. Watson RM, Schwartz JL, Maron BJ, Tucker E, Rosing DR, Josephson ME. Inducible polymorphic ventricular tachycardia and ventricular fibrillation in a subgroup of patients with hypertrophic cardiomyopathy at high risk for sudden death. *J Am Coll Cardiol* 1987;**10**:761–74.
16. Corrado D, Carraro R, Buja G, Fasoli G, Thiene G. Cardiomiopatia ipertrofica e morte improvvisa giovanile. Caso clinico e rivalutazione del problema [Hypertrophic cardiomyopathy and juvenile sudden death. A clinical case and re-evaluation of the problem]. *G Ital Cardiol* 1986;**16**:1002–8. Italian. PMID: 3556936.
17. Saumarez RC, Pytkowski M, Sterlinski M, Bourke JP, Clague JR, Cobbe SM et al. Paced ventricular electrogram fractionation predicts sudden cardiac death in hypertrophic cardiomyopathy. *Eur Heart J* 2008;**29**:1653–61.
18. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. 2nd ed. New York: Wiley; 2000. p156–64.
19. WHO Growth standards.
20. Efron B. *The Jackknife, the Bootstrap and Other Resampling Plans*. Philadelphia: SIAM; 1982.
21. Longato E, Vettoretti M, Di Camillo B. A practical perspective on the concordance index for the evaluation and selection of prognostic time-to-event models. *J Biomed Inform* 2020;**108**:103496.
22. Coifman RR, Donoho DL. In: Antoniadis A and Oppenheim G (eds.), *Translation-invariant de-noising, in Wavelets and Statistics*. San Diego, CA: Springer-Verlag; 1995. p125–150.
23. Maron BJ, Casey SA, Chan RH, Garberich RF, Rowin EJ, Maron MS. Independent assessment of the European Society of Cardiology sudden death risk model for hypertrophic cardiomyopathy. *Am J Cardiol* 2015;**116**:757–64.
24. Zegkos T, Tziomalos G, Parcharidou D, Ntelios D, Papanastasiou CA, Karagiannidis E et al. Validation of the new American College of Cardiology/American Heart Association guidelines for the risk stratification of sudden cardiac death in a large Mediterranean cohort with hypertrophic cardiomyopathy. *Hellenic J Cardiol* 2022;**63**:15–21.
25. Dong Y, Yang W, Chen C, Ji J, Zheng W, Zhang F et al. Validation of the 2020 AHA/ACC risk stratification for sudden cardiac death in Chinese patients with hypertrophic cardiomyopathy. *Front Cardiovasc Med* 2021;**8**:691653.
26. Vriesendorp PA, Schinkel AF, Liebrechts M, Theuns DA, van Cleemput J, Ten Cate FJ et al. Validation of the 2014 European Society of Cardiology guidelines risk prediction model for the primary prevention of sudden cardiac death in hypertrophic cardiomyopathy. *Circ Arrhythm Electrophysiol* 2015;**8**:829–35.
27. Norrish G, Qu C, Field E, Cervi E, Khraiche D, Klaassen S et al. External validation of the HCM Risk-Kids model for predicting sudden cardiac death in childhood hypertrophic cardiomyopathy. *Eur J Prev Cardiol* 2022;**29**:678–86.
28. Fernández A, Quiroga A, Ochoa JP, Mysuta M, Casabé JH, Biagetti M et al. Validation of the 2014 European Society of Cardiology sudden cardiac death risk prediction model in hypertrophic cardiomyopathy in a reference center in South America. *Am J Cardiol* 2016;**118**:121–6.

29. Leong KMW, Chow JJ, Ng FS, Falaschetti E, Qureshi N *et al*. Comparison of the prognostic usefulness of the European Society of Cardiology and American Heart Association/American College of Cardiology Foundation risk stratification systems for patients with hypertrophic cardiomyopathy. *Am J Cardiol* 2018;**121**:349–55.
30. Wang N, Xie A, Tjahjono R, Tian DH, Phan S, Yan TD *et al*. Implantable cardioverter defibrillator therapy in hypertrophic cardiomyopathy: an updated systematic review and meta-analysis of outcomes and complications. *Ann Cardiothorac Surg* 2017;**6**: 298–306.
31. Vriesendorp PA, Schinkel AF, Van Cleemput J, Willems R, Jordaens LJ, Theuns DA *et al*. Implantable cardioverter-defibrillators in hypertrophic cardiomyopathy: patient outcomes, rate of appropriate and inappropriate interventions, and complications. *Am Heart J* 2013;**166**:496–502.
32. Lambiase PD, Theuns DA, Murgatroyd F, Barr C, Eckardt L, Neuzil P *et al*. Subcutaneous implantable cardioverter-defibrillators: long-term results of the EFFORTLESS study. *Eur Heart J* 2022;**43**:2037–50.
33. Francia P, Olivetto I, Lambiase PD, Autore C. Implantable cardioverter-defibrillators for hypertrophic cardiomyopathy: the times they are a-changin'. *Europace* 2022;**24**: 1384–94.
34. Saumarez RC, Pytkowski M, Sterlinski M, Hauer RN, Derksen R, Lowe MD *et al*. Delayed paced ventricular activation in the long QT syndrome is associated with ventricular fibrillation. *Heart Rhythm* 2006;**3**:771–8.
35. Pytkowski M, Jankowska A, Maciag A, Kowalik I, Sterlinski M, Szwed H *et al*. Paroxysmal atrial fibrillation is associated with increased intra-atrial conduction delay. *Europace* 2008;**10**:1415–20.
36. Finocchiaro G, Sheikh N, Leone O, Westaby J, Mazzarotto F, Pantazis A *et al*. Arrhythmogenic potential of myocardial disarray in hypertrophic cardiomyopathy: genetic basis, functional consequences and relation to sudden cardiac death. *Europace* 2021;**23**:985–95.
37. Haissaguerre M, Cheniti G, Hocini M, Sacher F, Ramirez FD, Cochet H *et al*. Purkinje network and myocardial substrate at the onset of human ventricular fibrillation: implications for catheter ablation. *Eur Heart J* 2022;**43**:1234–47.
38. Gatzoulis KA, Georgopoulos S, Antoniou CK, Anastasakis A, Dilaveris P, Arsenos P *et al*. Programmed ventricular stimulation predicts arrhythmic events and survival in hypertrophic cardiomyopathy. *Int J Cardiol* 2018;**254**:175–81.
39. Fananapazir L, Tracy CM, Leon MB, Winkler JB, Cannon RO III, Bonow RO *et al*. Electrophysiologic abnormalities in patients with hypertrophic cardiomyopathy. A consecutive analysis in 155 patients. *Circulation* 1989;**80**:1259–68.
40. Leong KMW, Ng FS, Shun-Shin MJ, Koa-Wing M, Qureshi N, Whinnett ZI *et al*. Non-invasive detection of exercise-induced cardiac conduction abnormalities in sudden cardiac death survivors in the inherited cardiac conditions. *Europace* 2021;**23**:305–12.
41. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 2014;**35**:1925–31.
42. Chen L. Overview of clinical prediction models. *Ann Transl Med* 2020;**8**:71.
43. Madigan D, Ryan PB, Schuemie M, Stang PE, Overhage JM, Hartzema AG *et al*. Evaluating the impact of database heterogeneity on observational study results. *Am J Epidemiol* 2013;**178**:645–51.
44. Rostami R, Nahm M, Pieper CF. What can we learn from a decade of database audits? The Duke Clinical Research Institute experience, 1997–2006. *Clin Trials* 2009;**6**:141–50.
45. Wagenmakers EJ, Sarafoglou A, Aczel B. One statistical analysis must not rule them all. *Nature* 2022;**605**:423–5.
46. Curtis JP, Sokol SI, Wang Y, Rathore SS, Ko DT, Jadbabaie F *et al*. The association of left ventricular ejection fraction, mortality, and cause of death in stable outpatients with heart failure. *J Am Coll Cardiol* 2003;**42**:736–42.
47. NICE Technology Appraisal Guidance (TAR 314) <https://www.nice.org.uk/guidance/tpg603>
48. Alfonso F, Torp-Pedersen C, Carter RE, Crea F. European Heart Journal quality standards. *Eur Heart J* 2021;**42**:2729–36.