nature portfolio

Corresponding author(s): Natalia A. Trayanova

Last updated by author(s): Feb 10, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	MRI data in this manuscript were collected using default/pre-installed software proprietary to the device manufacturers, including Signal, GE Medical Systems, Waukesha, Wisconsin; Avanto, Siemens, Erlangen, Germany. The exact software versions for the devices cannot be precisely retroactively ascertained given the very broad nature of the study (100+ centers spanning several years). The data were provided to the authors in DICOM format and processed with the Python package pydicom 1.2.2. All other clinical covariates were provided in .csv format and processed using Python pandas 0.24.2 and numpy 1.16.2.
Data analysis	The entire model development was performed using open-source software in Python 3.7 (Tensorflow 1.15, Keras 2.2.4, numpy 1.16.2, scipy 1.2.1, openCV 3.4.2, hyperopt 0.1.2, lifelines 0.25.5) and can be replicated using details described in Methods. Custom code for the data processing and the Tensorflow/Keras models used in this manuscript can be found at https://gitlab.com/natalia-trayanova/sscar.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

De-identified CMR images and patient clinical data can be provided by the authors pending approval by Johns Hopkins University Institutional Review Board and

Brigham and Women's Hospital Institutional Review Board and a completed material transfer agreement. Requests for these data should be made to N.A.T. and/or C.M.A.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For the internal cohort, all patients in the consecutive clinical study with adequate LGE-CMR data for analysis were included. The external cohort was a matched 4-1 case-control, where all cases with adequate LGE-CMR data were included.
Data exclusions	For the internal cohort, patients were included if they had contraindications to CMR, New York Heart Association (NYHA) functional class IV, acute myocarditis, acute sarcoidosis, infiltrative disorders (e.g., amyloidosis), congenital heart disease, hypertrophic cardiomyopathy, or renal insufficiency (creatinine clearance <30 mL/minute after July 2006 or < 60 mL/minute after February 2007). For the external cohort, exclusion criteria included a history of cardiac arrest not associated with acute MI, current or planned ICD, or life expectancy < 6 months.
Replication	Replication entailed cross validation of 100 train/test data splits on the development set and testing in an independent, external cohort.
Randomization	All data splits in the cross-validation processed were assigned randomly.
Blinding	Blinding was not feasible, given that the model presented in the manuscript is a supervised deep neural network which requires target data during training and for performance evaluation purposes.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☐	☒ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	ClinicalTrials.gov ID NCT01076660 (internal cohort), ClinicalTrials.gov ID NCT01114269 and ClinicalTrials.gov ID NCT00487279 (external cohort)
Study protocol	Study protocols can be found at ClinicalTrials.gov.
Data collection	For the internal cohort, patients satisfying clinical criteria for ICD therapy for SCDA (LVEF <= 35%) were enrolled at 3 sites: Johns Hopkins Medical Institutions (Baltimore, MD), Christiana Care Health System (Newark, DE), and the University of Maryland (Baltimore, MD). A total of 382 patients were enrolled between November 2003 and April 2015. For the external cohort, study populations are multicenter prospective cohort studies comprised of patients with coronary disease on angiography or documented history of myocardial infarction (MI). The PRE-DETERMINE study enrolled 5764 patients with documented MI and/or mild to moderate LV dysfunction (LVEF between 35-50%) who did not fulfill consensus guideline criteria for ICD implantation on the basis of LVEF and NYHA class (i.e., LVEF > 35% or LVEF between 30% - 35% with NYHA Class I HF) at study entry. Exclusion criteria included a history of cardiac arrest not associated with acute MI, current or planned ICD, or life expectancy <6 months. The accompanying DETERMINE Registry included 192 participants screened for enrollment in PREDETERMINE who did not fulfill entry criteria on the basis of having an LVEF < 30% (n=99), LVEF between 30% - 35% with NYHA Class II-IV heart failure (n=19), or an ICD (n=31) or were

unwilling to participate in the biomarker component of PREDETERMINE (n=43). Within these cohorts, 809 participants had LGE-CMR imaging performed. Within this subset of patients, 23 cases of SCD occurred and were matched to 4 controls on age, sex, race, LVEF and follow-up time using risk set sampling. Out of the resulting 115 patients, the current study focused on 113 patients with adequate LGE-CMR images for analysis. Finally, covariate data for this cohort were minimally harmonized with the internal cohort, by retaining common covariates only. All patients provided informed consent to be part of the original studies without compensation.

Outcomes

Composite SCD outcomes [Time Frame: Every 6 months for 5 years] (external cohort) and sudden and/or arrhythmic cardiac death or resuscitated ventricular fibrillation [Time Frame: Median follow-up time estimated to be 8 years] (internal cohort). The primary endpoint for LVSPSCD was SCDA defined as therapy from the ICD for rapid ventricular fibrillation or tachycardia, or a ventricular arrhythmia not corrected by the ICD. For the PRE-DETERMINE studies, the primary end point was sudden and/or arrhythmic death. Deaths were classified according to both timing (sudden versus non-sudden) and mechanism (arrhythmic versus non-arrhythmic). Unexpected deaths due to cardiac or unknown causes that occurred within 1 hour of symptom onset or within 24 hours of being last witnessed to be symptom free were considered sudden cardiac deaths. Deaths preceded by an abrupt spontaneous collapse of circulation without antecedent circulatory or neurological impairment were considered arrhythmic in accordance with the criteria outlined by Hinkle and Thaler (2018). Deaths that were classified as non-arrhythmic were excluded from the endpoint regardless of timing. Out-of-hospital cardiac arrests due to ventricular fibrillation that were successfully resuscitated with external electrical defibrillation were considered aborted arrhythmic deaths and included in the primary endpoint.