

Supplementary Material: Measurement definition, reference availability and testability shaped the verdicts on classical ECG rules in routine hospital data

Supplement to the preprint (version 1) of 17 October 2026. Leo Probst, Johannes Kepler University Linz, Medical Faculty, Altenberger Strasse 69, 4040 Linz, Austria; leo.probst.research@proton.me. Reference numbers [n] refer to the reference list of the main text; references that appear only here continue that numbering at the end of this document.

Contents: S1 Specification binding, amendment history, access history and the second test pass; S2 Methods in detail; S3 Candidate rules and selection; S4 Cohort description and availability; S5 Result tables and Figure S1; S6 Disclosures and deviations; S7 Reporting checklist.

S1. Specification binding, amendment history, access history and the second test pass

S1a. What was fixed, and when

The analysis was fixed in a written specification (version 3.2, bound on 15 October 2026) and an analysis script. The random seed is the first 8 hexadecimal characters of the specification's SHA-256 as an integer (425,498,204). Times are UTC; local time was UTC+2.

Step	Time	Object	SHA-256
Specification v3.2 (bound)	15 Oct 2026	Pre-analysis specification	195c965c57f808a6b077f6563ad1164ea28a38d35f9256d40da195f00455d683
First freeze (superseded)	15 Oct 19:23:47Z	Analysis script v1	654543515e15ebbb7f97bca67476bce6f4bfa2e028fe15f53519a967bbe457b6
Second freeze (superseded)	15 Oct 19:44:21Z	Analysis script v2	b2253e58fe11db555d73c641fbfd859f960ee04b31f68b4cb539c6aa8874e366
Amendment 1 to the specification	first version 15 Oct 20:58Z; final text before the final freeze	Amendment text (final version)	bd98cbf91e739c844d79f12303c785e22fc1d1aaa4f7953b6dbdf6159a7e88d3
Final freeze (binding)	15 Oct 21:28:08Z	Analysis script v3.2 (original)	1d40d782c65ceeffd4be9c4380e4eef822cbc98aaa710a4d7bf340699bf773e8
Published copy of the script	repository release v1.2.0	k1_auswertung.py, cleaned of local paths and user names (11 path and 8 name replacements)	67aaaf6acad88d53246320bf95f31b747d76c2042df9489d29ea160abed76943
Input manifest (22 files with SHA-256)	15 Oct	Inputs, helper files, script	0b78c49b530728fdb9c4531b074236b72dae77f5d38c0cef4a9a83c774c0915f
Development run	15 Oct 21:29Z to 22:59Z	Development result files	a96607ed278dd4531bcd5a145aae0108d6688b42c1c7439824edd02ec05f9cf8
Test run (single pass)	16 Oct 05:19:35Z to 05:28:32Z	Test result files	09abdddf59be80ec72eebf8744ab30915a97ff921cb6cda82c51423e8afff232c
CODE-15% reproduction	16 Oct 05:30Z	Result file	not hashed
Second test pass: rules file (renewed version)	saved 16 Oct 21:43:42 local (19:43:42Z); states 21:46 local	Rules for the second test pass	not recorded in this supplement
Second test pass: driver script	16 Oct, before 21:45 local	k1_testergaenzung_treiber.py	8116e86e28a49dafdb1c67f5192d2777961bbcb61716a6e732705946ce7b5d3e
Second test pass: tests of the driver (77 checks)	same	test_k1_testergaenzung_treiber.py	ec1111c3c06a6c4d84f588e860289fcd00cf436c247f0cbd65c95b4bd0e07638
Second test pass: manifest M' (103 entries: manifest M3 plus driver and extraction code)	same	Inputs and code	61efafb2007cdc02cdb549375b6693c65a9aaf481b607eb0ef042c1e39527325
Second test pass: register start	16 Oct 19:45:17Z (21:45:17 local)	Start of the pass	-
Second test pass: register end	16 Oct 23:36:09Z (01:36 local on 17 Oct)	End of the pass	-
Second test pass: extracted measurement file of the test partition	after the start	31,776 index ECGs	e52d76b1296b9838288eb886a1a8f1ae2b35fbfc4a658bff1989e62e3a095881

Step	Time	Object	SHA-256
Second test pass: result files (six claims)	16 Oct 19:45Z to 23:36Z	Test result files, second pass	0abbf8a8e1cc1a3158bf3e1b5319536d439cabfe0ca2189775658ff87a906140
Extraction software state (code hash of the measurement worker)		Identical to that of the development extraction	c166bd00403dcde9f154d6c50f15ae16791029d9b507c4a4094e6f96215789b9
Cohort table script (descriptive, development data only)	after the claim results were known	k1_kohortentafel.py	b671b4ab7988b0cdb40984069e7313438fa4c945f24bb483bdec1c8a0c450984

The development run started at 23:29 local time on 15 October and ended at 00:59 local time on 16 October. The script and the amendment text did not change between the final freeze and the end of the last run (hashes checked at the end). Hashes of the other published scripts (SHA-256 of the original and of the published file; the three helper files that the script checks at start are published byte-identical):

File in pub13_classical_rules/	SHA-256 (original; published identical unless stated)
regeln.py (rule functions)	1e04fc99a5381512ade60654974b7a252548beee804b83225ef0ff50c9a58e11
k1_geruest.py (framework functions)	7736a0a7eece05f80f8f3c2e6239b63d4f5a1a244607c706d6828d321d44bd62
adapter_cache.py (adapter from the feature table to the rule functions)	f4f651467e742169b844bdc65456bd365c1a99557dbd430bbae6c6a9e32c075d
k1_codelisten.py (code-list builder)	3245f2931fbddac091ccab8e217b60bfc9f258d993e377f9eb01e142708d7aba
codelists/codelisten_k1.csv (explicit ICD-9 and ICD-10 lists)	14cfc4fbc71d37607cee05cd51c96ee99e0b819ca887d84017b42a416573d516
k1_bericht.py (report writer, not part of the frozen script; published copy 14bdcd1c16692a6fc32afc4e77d769c12f191c986bc7746f6807bc088bda8470)	55e6954aeb9f29244be8153353a639d0bb642c24da9af1246b7fe2cca51f208b
test_k1_auswertung.py (61 tests on synthetic data)	c44ee1724f3ef30ecbc0aa9cb52f7c987983fa4844df5191ae68a9f574670cf2

The repository is <https://github.com/leoprobst/ecg-auswertung-code> (release v1.2.0, folder pub13_classical_rules). The specification file itself is not in the repository; its hash above allows the check performed by the script at start. Release v1.2.0 includes the driver script of the second test pass, its tests and the cohort-table script.

S1b. What had been seen before the final freeze

1. *Earlier statistics.* Rule-versus-outcome statistics (counts, sensitivity, specificity, AUC) for Sokolow-Lyon, Cornell, Peguero-Lo Presti, QTc above 500 ms and Q waves had been computed in an earlier automated run of the project and may have been seen; these rules are reported as previously examined. For the QRS threshold of 120 ms, an AUC had been read in earlier exploratory notes. For low voltage, right bundle branch block, PR above 200 ms, the sex-specific QRS claim and normal ECG, statistics had been computed earlier but were not known to have been seen.

2. *Positivity rates.* The unweighted positivity rates of several rules in the development ECGs (Sokolow-Lyon 3.4%, Cornell 9.8%, Peguero-Lo Presti 15.5%, low voltage 6.7%) were known before the margins were set.
3. *Partial results of early script versions.* The first development run with script v1 completed two claims (right bundle branch block and PR above 200 ms) before it was stopped, and their results were seen (hazard ratios for death 1.127 and 0.901, for atrial fibrillation 1.532 for PR above 200 ms). A test run with 20 and 10 bootstrap repetitions over all claims was also seen; it was used only to find errors. After script v2 (fixes for a Cox fitting failure and a convergence rule), the claims for right bundle branch block, PR above 200 ms and part of the Sokolow-Lyon claim were completed and seen. None of these results entered the reported analysis: all claims were recomputed with script v3.2 in a new output folder.
4. *CODE-15%.* While checking column names of the exam table, four data rows (age, sex and label columns) became visible by chance; nothing was analysed or noted.
5. *Test partition.* See S1e.

S1c. Amendments after partial results

Amendment 1 was written after the early results in S1b were seen and rests on independent reviews (a second-opinion review of the specification by a separate model and a code review of the script). It did not change any margin (0.90, 0.98, 0.80, 1.20, 10 ms and so on). It changed the following.

- **QTc claims.** The rule was split into two claims judged separately, ventricular tachycardia, fibrillation or arrest and death within 30 days, instead of one combined claim. The criterion for “supported” (lower one-sided bound above 1.0) was not changed. The criterion for “not supported” was changed from an upper bound below 1.25 to an upper bound of 1.0 or lower, which makes it harder to reach; the test run with few bootstrap repetitions had shown the rule as supported. The ventricular outcome was declared a concurrent association with no prognostic claim.
- Reference states for low voltage with pericardial effusion and amyloidosis, and for Q waves, were redefined (positive, negative, unknown). The population of the right bundle branch block and PR claims was renamed “no documented prior disease”. Confidence-bound, bootstrap-failure and replication rules were tightened, and technical errors were fixed.
- The knots of the age spline were set at the 5th, 35th, 65th and 95th percentiles of age in the development population; the test partition used the same knots.

Changes after the amendment were limited to the findings of three further code reviews (script v3.1 and v3.2); they did not touch rules, margins or claims.

S1d. The gap in the first test pass and the second test pass

In the first test pass the feature table from which the voltage, Q-wave and low-voltage rules are computed did not contain the test ECGs; voltage and Q-wave measurements had not been extracted for them. Six claims (Sokolow-Lyon, Cornell, Peguero-Lo Presti, low voltage with effusion, low voltage with amyloidosis, Q waves) could not be computed there, and nothing was substituted. This was recorded before the first test run.

Second pass, specified after the other seven test results were known. The rules file of the pass states this openly. Known at that time were the test results of the QTc claims (supported), of the normal-ECG, right bundle branch block, PR and QRS claims (inconclusive) and the development results of all 13 claims. The expectation noted before the pass (voltage-based claims supported again, based on development results) is recorded in S2.

What the pass did. (1) The amplitudes and Q-wave flags of the 31,776 test index ECGs were extracted with the same measurement worker that produced the development feature table (identical code hash, c166bd00...). The worker received only waveform, age, sex, device intervals and device report text, and read no labels, endpoints, echocardiographic, laboratory or follow-up data. (2) A driver script called the unchanged functions of the analysis script (version 3.2; hash unchanged before and after) and supplied them with the extracted measurements instead of the feature table that does not cover the test ECGs; it forced exactly the six claims and rejected any other unit. (3) Only 19 columns of the rule table that had been empty in the first pass were allowed to change (Sokolow-Lyon and Cornell variants and fixed-offset versions, Peguero-Lo Presti variants, three low-voltage definitions, the Q-wave flag and the voltage hypertrophy flag); all other columns, the schema and the keys had to be exactly equal to the first pass, and the columns for right bundle branch block, PR and QRS duration were reset to the first-pass state, so those three claims were not recomputed. (4) The extraction passed an error limit of 1% (0 extraction errors; 186 ECGs (0.59%) failed signal quality and 46 (0.14%) were excluded for lead placement, as in the development feature table; these 232 ECGs enter the rules as “not evaluable”). (5) The run used the development design unchanged (age spline knots, standardisation) and $B = 4,000$ bootstrap repetitions (1,000 for sensitivities) with unchanged seeds.

Gates before the start (development data only). Selftest on synthetic data with negative tests (77 checks); the rule table of the development partitions produced by the driver was exactly equal to the existing one (49 columns, 128,405 rows, no tolerance); the six development claims run through the driver path gave identical estimates, intervals, classes and verdicts; the extraction path reproduced all 4,800 cells of 32 columns for 150 development ECGs.

Code reviews. The driver was reviewed three times by a separate language model working only from the files (no data, no execution). The first two reviews were not releasable: they asked for a stricter error limit across all extraction parts, an explicit list of changeable columns with an exact comparison including empty columns, manifest verification before every test access, forcing of exactly six claims, a pre-check at the final step, and an exact value comparison without floating-

point conversion. The third review found no new errors in the corrected parts and released the driver.

Time of the rules file. The rules file of the second pass states “renewed on 16 October 2026, 21:46 local”. The file system shows that it was saved at 21:43:42; the pass was registered at 21:45:17. The file was therefore saved before the start, and the time stated in it was an estimate about 2 minutes too late. Earlier versions of the file (21:25 and 21:40) also carry estimated times. Hashes and order of events are not affected.

Limits of the pass. The driver is not part of the 61 tests of the analysis script; its coverage is the selftest and the gates above. The equivalence on development data shows the shared path but not the loader of the test measurements; this loader was covered by the selftest and a probe with the rule table built from the measurement file. The extraction was a separate computation with the bound worker, not the project’s regular extraction stage (which requires a freeze of the feature run that does not exist). The test partition was not unseen for this pass; its results are reported as post hoc.

S1e. Access history of the test partition

Four statements apply, in this order. (1) The test partition belongs to the same database and had previously been used for a confirmatory analysis of a different marker in a related study (a single bound test run on 12 October 2026), for a larger screen of marker pairs with a further bound test run on the same day, and for exploratory post hoc analyses of the project on the same day;[32] incidental sightings of test aggregates in unrelated analyses are recorded in the project’s access log. It is therefore not unseen data. The related manuscripts [29,32-34] use the same database and the same person-disjoint partitions, and some of the same reference labels (for example, echocardiographic ejection fraction), but not the ECG rules tested here. (2) For this study, a row count of the test table without a split filter and without outcome information was made before the first test pass and recorded in the access log. (3) The first test pass (16 October 2026, 05:19:35Z to 05:28:32Z) computed seven claims. (4) The second pass (S1d) computed the six voltage-based claims after the results of the first pass were known. A single run does not restore the unopened status of a partition; the replication is internal and, for the six claims, post hoc.

S1f. Analyses specified after results were known

The cohort table (Table 2 of the main text and S4c), the analysis by heart-rate stratum and the analysis of the thresholds 480, 500 and 520 ms (Table 5 and S5h) were specified after the results of the claims were known, on development data only (the test partition was not read for them). They are descriptive; the cohort-table script was bound by its own rules file and checked against the cells of the QTc claims (the 500 ms cells equal those of the analysis exactly). The strata (below 60, 60 to 100, above 100 per minute) and the thresholds (480, 500, 520 ms) were specified after the results were known.

S2. Methods in detail

Cohort. Adults (18 years or older) with a 12-lead ECG with usable device metadata and signal quality in MIMIC-IV-ECG, linked to MIMIC-IV. The development data are the union of the training and validation partitions (129,224 persons with 640,119 ECGs in the source list, person-disjoint folds); the test partition is a separate person-disjoint partition of the same database. One ECG per person: the first eligible adult ECG in the partition, chosen by ECG time without knowledge of references or rule results. Machine-reported pacing and atrial fibrillation (including uncertain statements) excluded a person from the QTc analysis; pacing and left bundle branch block (device statements) excluded voltage and Q-wave rules; atrial fibrillation also excluded the PR claim.

Measurements. QT, RR, PR and QRS durations are the device measurements held in the feature table of the analysis pipeline; amplitudes (voltage and Q-wave criteria) were measured by the author's software LoopLens (research snapshot in the repository) and entered the rule functions through an adapter. The recording devices were from various manufacturers; the algorithm and version are not recorded for each ECG. The exact derivation of the QT interval in the feature table from the device fiducial points is not documented in the result reports.

Rule definitions. The rule functions follow the primary descriptions (Table 1 of the main text); where a source contradicts itself, the normative statement was bound (for example, the recommendation box rather than the text for the QT threshold in men). Cornell voltage: R in aVL + S in V3 above 2.8 mV in men and 2.0 mV in women after Casale 1987.[1] The earlier Casale 1985 paper gives 35 and 25 mm for a related criterion;[16] we used the 1987 values.

Margins and their origin. The margins were set by the author before the final runs and are listed with their orientation in Table 1 of the main text. In addition: Sokolow-Lyon specificity 0.90, the upper end of the range of 85% to 90% quoted by the standardization statement[25] and below the autopsy value of 100%;[1] Peguero-Lo Presti specificity 0.90 (original validation 89% to 90%);[17] low voltage and effusion: specificity 0.98 and sensitivity 0.10 rounded from the original Table 3 (29 positives, 126 negatives);[18] low voltage and amyloidosis: sensitivity 0.60 (13% to 60% by definition);[19] Q waves: specificity 0.80; normal ECG: 1% oriented on a negative predictive value of 99% and one third from the guideline wording;[26] hazard ratio 1.20 as a usual non-inferiority margin; QRS: 10 ms, with guideline simulations quoting 9 to 13 ms.[24]

Informative targets. For Sokolow-Lyon and Cornell, a positive likelihood ratio of at least 5 and a sensitivity of at most 0.50 (orientation: a threshold for a large change in probability[31] and the autopsy sensitivities[1]) were bound in advance as non-binding, descriptive targets with one-sided limits at 0.025 without correction. For QTc, the question whether a doubling of the risk can be excluded was bound as a secondary question (upper bound below 2). A ratio below 2 is excluded only through the lower bound (S5a).

Expectations noted before the analysis.

Claim	Expectation	Outcome
Peguero-Lo Presti, specificity at least 0.90	Not supported (positivity in a public data set was 22% to 23%, in the feature table 15.5%)	Supported under the primary definition in both partitions (0.923 and 0.929); claim as a whole inconclusive because of selection in both
Normal ECG, one third of NSTEMI infarcts	Not supported (strict definition of normal)	Primary definition not supported (3.5%); alternative definition 38%; inconclusive
Right bundle branch block and PR above 200 ms, hazard ratio below 1.20	Not supported (excess risk as in community cohorts)	Right bundle branch block inconclusive; PR prolongation not supported for atrial fibrillation under the primary definition, supported for death under the primary definition; inconclusive as a unit
QTc above 500 ms	Not supported was possible (small adjusted odds ratio in an intensive care cohort)	Supported for both outcomes in both partitions
Sokolow-Lyon, Cornell and Q waves, test partition (noted before the second pass, from the development results)	Supported again	Supported in the test partition (Sokolow-Lyon 0.973, Cornell 0.963, Q waves 0.930); the point estimates could not have missed the margins (S5f)
QTc, can a risk ratio of 2 or more be excluded (upper bound below 2)?	Reported as secondary	Not met: upper bounds 3.1 to 6.6, so a ratio of 2 or more is not excluded. Conversely, a ratio below 2 is excluded where the lower bound exceeds 2 (six of eight estimates, S5a)

Reference standards (items of STARD). *Echocardiographic hypertrophy:* left ventricular mass index above 115 g/m² (men) or 95 g/m² (women); the reported index if present, otherwise the mass from septum, posterior wall and end-diastolic diameter (Devereux formula, $0.8 \times 1.04 \times [(\text{septum} + \text{diameter} + \text{posterior wall})^3 - \text{diameter}^3] + 0.6 \text{ g}$)[30] divided by the body surface area of the echocardiogram report; the index column itself was empty in the data. *Echocardiogram:* nearest transthoracic study within 30 days (earlier study in case of ties); transoesophageal and stress studies and windows of 7 days or 24 hours only as sensitivity analyses. Of 122,880 patients with an evaluable Sokolow-Lyon or Cornell rule, 20,150 (16%) had a computable reference; echocardiograms are mostly clinically indicated, and blinding to the ECG cannot be assumed. *Pericardial effusion:* moderate, moderate-to-large or large effusion field against documented very small, small, trivial or anterior fat pad; an empty field and “small-moderate” were unknown. *Cardiac amyloidosis:* amyloidosis code (ICD-10 E85, ICD-9 277.3) with a cardiomyopathy code in the same stay or a septum of at least 12 mm; the systemic code alone was a sensitivity analysis. *Earlier infarct:* infarct code (I25.2, 412, or acute infarct) in an earlier stay; negative: an earlier stay with a diagnosis list and none of these codes and no coronary intervention; the code is sometimes assigned because of

Q waves, which overstates agreement. *Infarction in chest pain*: project infarct label in emergency visits with chest pain as presenting complaint and a troponin measurement; NSTEMI infarcts by diagnosis codes.

Cohort table, heart-rate strata and thresholds. Population: the 120,260 development patients with an evaluable QTc and no device-reported atrial fibrillation. Care setting: intensive care if the ECG fell into an intensive care stay, otherwise emergency department if an emergency visit was linked, otherwise ward if a hospital stay was linked, otherwise other (no linked stay). Heart failure before the index ECG is reported only for patients with earlier diagnosis codes, because patients without an earlier stay have no recorded history and the share would be underestimated; QT-prolonging medication (prescription or reported home medication) only for patients with medication data, because without data it counted as absent in the analysis. Heart rate: the device value, otherwise from the RR interval, limited to 30 to 220 per minute. Counts rounded to 10; counts of 1 to 9 as “<10”; when one cell of a variable was masked, all cells of that variable were rounded to 100. Risk ratios with Wald intervals as for the QTc claims, unadjusted.

Weights and selection scenarios. Inverse-probability weights for having a computable reference (logistic model with an age spline, rule result and interactions; stabilised; truncated at the 1st and 99th percentile; ridge penalty only for separation); diagnostics: effective sample size at least 50% of verified patients, standardised differences after weighting below 0.1. Selection scenarios: odds of the reference among patients without a reference 0.5, 1 or 2 times those of comparable patients with one, separately for rule-positive and rule-negative patients (3 by 3 grid); a verdict counted as stable only if it held over the whole grid.

Specificity floor, chance and sensitivity ceiling. With T+ a positive rule result, D- absence and D+ presence of the reference condition, all in the weighted target population: floor = $1 - P(T+)/P(D-)$, the lowest specificity possible if every positive result were a false positive; chance expectation = $1 - P(T+)$, the specificity of a rule whose result is unrelated to the reference; ceiling = $P(T+)/P(D+)$, the highest sensitivity possible. When the floor is at or above the margin, the point estimate of the specificity cannot fall below the margin at the observed marginal distributions; when the ceiling is at or below a sensitivity margin, the point estimate cannot exceed it. The verdict additionally depends on the confidence bound, the alternative definition and the selection scenarios.

QTc exposure. Device QT and RR intervals. Bazett QTc = $QT / RR^{(1/2)}$ (primary); Fridericia QTc = $QT / RR^{(1/3)}$ (alternative); linear (Framingham) QTc = $QT + 154 \times (1 - RR)$ ms with RR in seconds (sensitivity).[44] QTc categories for the dose-response analysis, with limits after published thresholds:[23] 0 (below 450 ms in men and 460 ms in women), 1 (up to 480 ms), 2 (above 480 up to 500 ms), 3 (above 500 ms); the 500 ms threshold follows hospital guidance.[9] We could not check the category limits against the full text of the standardization statement.

QTc outcomes. Ventricular tachycardia, fibrillation or cardiac arrest: positive if the project label for ventricular tachycardia or fibrillation (ICD codes or event) or the label for cardiac arrest was

positive at hospital-stay level; negative if both were negative; otherwise missing (hence the smaller analysis population among patients with an admission record). Timing relative to the ECG cannot be established, so this is a concurrent association. Death within 30 days: the project label for 30-day mortality, ordered in time after the index ECG.

Estimation. Unadjusted risk ratio from the 2 by 2 table with a Wald interval on the log scale (equivalent to a Poisson model with robust variance when unadjusted). Adjusted ratios: Poisson regression with robust (HC0) variance and the covariates age (natural cubic spline), sex, admission type, heart rate; fully adjusted additionally QT-prolonging medication, potassium, magnesium, total calcium and heart failure. Laboratory values were the nearest value within 24 hours of the ECG (the plan said the same calendar day), and potassium was not filtered for haemolysis. Missing covariates were handled by an indicator variable and the median; there was no multiple imputation. Adjusted models used the same patients as the primary analysis. The adjusted models contain no comorbidity index, vital signs other than heart rate, or intensive care admission.

Unstable minimally adjusted model. The adjusted ratios come from an unpenalised Poisson regression with robust variance and no convergence safeguard. In the test partition, the minimally adjusted ventricular model (10,200 patients, fewer than 50 events among those with QTc above 500 ms, because the cell is rounded to 100 in the report) returned a finite point estimate (2.39) but limits of 0 and infinity, which the code produces when the robust standard error is not finite. The result reports do not record the cause. The fully adjusted model on the same patients converged (2.22; 1.53 to 3.22). We treat the minimal model as unstable in the test partition and do not interpret it; the development estimate (2.87; 2.42 to 3.42) is shown for completeness.

Decision rule. The two QTc claims and the sex-specific QRS claim belong to a family of three hypotheses (one-sided limits at $0.025/3 = 0.00833$, Bonferroni); the other families have 8 (voltage, Q-wave and low-voltage claims), 2 (normal ECG) and 4 (right bundle branch block and PR, death and atrial fibrillation each) hypotheses. Supported: lower bound above 1.0 (QTc). Not supported: upper bound of 1.0 or lower. Otherwise inconclusive. A claim verdict required the primary and the alternative definition to agree. Replicated: supported in both partitions; the test-to-development ratio of risk ratios is reported (S5b).

Other claims. Cox models with Breslow ties and robust variance, adjusted for age (spline), sex, admission type, earlier diagnoses, AV-nodal and diuretic drugs and heart rate (right bundle branch block, PR); atrial fibrillation as a cause-specific hazard with death as a competing event, censored at the last documented contact; the population of the claim had no documented prior heart or lung disease (patients without an earlier stay counted as without documentation, which does not mean healthy). The sex-specific QRS claim used logistic models of the risk of an ejection fraction of 35% or lower per QRS duration (natural spline, knots at 90, 110, 130, 150 and 170 ms), separately by sex; the difference in QRS duration at which the risk in women equals the risk in men at 150 ms was read on a 1-ms grid, and results with no unique crossing point are reported as not testable.

Data versions. MIMIC-IV-ECG version 1.0 (device measurements and report text). MIMIC-IV version 3.1 is documented for the diagnosis dictionary against which the code lists were confirmed; the label tables were built by the project's earlier pipeline, and its result reports do not name their MIMIC-IV version. Label definitions: version 2.9.0.

CODE-15%. The open subset of the CODE cohort (345,779 exams, 233,770 patients; 233,600 patients with positive follow-up time; 8,341 deaths). Only death is available (first exam per patient). Exposure: the database label for right bundle branch block or for first-degree AV block, without the other labelled abnormalities; comparison group: none of the six labels (a sensitivity analysis used normal ECG). Cox model with age (spline) and sex, robust variance, Breslow ties; bounds at 0.025/2. The labels combine the cardiologist's report with the output of the Glasgow ECG analysis program (right bundle branch block rejected below a QRS duration of 115 ms, first-degree AV block below a PR interval of 190 ms),[36,37] so they are neither the device definitions nor an independent reading. This is a reproduction in an open subset of a cohort that has been published on, not an independent confirmation.

S3. Candidate rules and selection

Twenty-five candidate rules were compiled before the analysis from textbooks, guidelines, standardization statements and primary studies; their wording is paraphrased. Ten rules (11 candidates, as low voltage covers two) were analysed.

Candidate	Rule	Decision and reason
C01	A normal 12-lead ECG makes systolic dysfunction unlikely	Not analysed here; covered by a companion study on QRS-T coupling and ejection fraction[32]
C02	Sokolow-Lyon voltage indicates hypertrophy	Analysed
C03	Cornell voltage indicates hypertrophy	Analysed
C04	Peguero-Lo Presti criterion is more sensitive than Cornell	Analysed
C05	Romhilt-Estes score of 5 or more indicates hypertrophy	Not analysed; kept as a possible sensitivity analysis
C06	Low voltage suggests pericardial effusion	Analysed
C07	Low voltage suggests amyloidosis	Analysed
C08	QTc above 500 ms is dangerous	Analysed
C09	Isolated right bundle branch block is harmless	Analysed
C10	New left bundle branch block in chest pain is an infarct equivalent	Excluded: our amplitude flags abstain in left bundle branch block
C11	Sgarbossa criteria for infarction in left bundle branch block	Excluded: same reason; a new implementation would be needed
C12	S1Q3T3 and related signs suggest pulmonary embolism	Excluded: no imaging-based reference in the data
C13	Peaked T waves indicate hyperkalaemia	Excluded: no threshold in the source and no haemolysis filter
C14	Pathological Q waves indicate an old infarct	Analysed
C15	Poor R-wave progression suggests anterior infarction	Excluded: implementation not audited
C16	Right axis and R/S above 1 in V1 suggest right ventricular hypertrophy	Excluded: studied for the same data type by others[10]
C17	P-wave criteria indicate left atrial enlargement	Excluded: the measurement export would need a new full run
C18	PR above 200 ms is benign	Analysed
C19	Fixed QRS thresholds have the same meaning in women and men	Analysed
C20	A normal ECG does not exclude acute coronary syndrome	Analysed
C21	T-wave inversion beyond V1 is pathological	Excluded: reference rarely coded
C22	Wellens pattern indicates critical stenosis	Excluded: reference and screen not reliable
C23	Digitalis scooping indicates digoxin exposure	Excluded: operationalisation failed an earlier check
C24	Regular narrow tachycardia near 150 per minute is flutter	Excluded: flutter not separable from fibrillation in codes
C25	Bifascicular block predicts high-grade AV block	Excluded: in-hospital follow-up too short

S4. Cohort description

S4a. Numbers by claim

The numbers of ECGs at each step are in Figure 1 of the main text. Rule-evaluable and analysed numbers for the right bundle branch block and PR claims: 88,400 and 114,900 evaluable ECGs; 87,400 and 109,800 in the population without documented prior disease; death and atrial fibrillation analyses 81,100 and 69,900 (right bundle branch block) and 74,700 and 65,000 (PR). Patients without any follow-up contact were excluded from the atrial fibrillation analyses (15,800 and 13,800).

S4b. Availability of QTc outcome labels by exposure (development; rounded counts)

Group	Evaluable QTc	With 30-day mortality label	With admission record and complete ventricular labels
QTc above 500 ms	2,230	2,100 (94%)	1,460 (65%)
QTc 500 ms or less	118,030	101,940 (86%)	39,500 (33%)

In the test partition, 610 patients had QTc above 500 ms in the death claim, 570 (93%) of them with a mortality label, and 600 in the ventricular claim, 400 (67%) with ventricular labels; among about 29,100 others in each claim, 25,040 (86%) and 9,800 (34%) had them (test counts partly rounded to 100). Patients with QTc above 500 ms thus more often had an admission record, which suggests more inpatients among the exposed.

S4c. Characteristics of the cohort

Table 2 of the main text gives the characteristics of the 120,260 development patients with an evaluable QTc by QTc group (descriptive; specified after the results were known; development data only). Availability of information must be read with it: earlier diagnosis codes were available for 11.3% of all patients (9.4% of those with QTc above 500 ms), medication data for 32.6% (57.4% of the exposed), so heart failure (2.2% of patients with earlier codes) and QT-prolonging medication (29.9% of patients with medication data) describe subgroups. Heart rate was available for all patients. Female share in analysed populations: 55% in the right bundle branch block death analysis (45,000 of 81,100) and 54% in the PR death analysis (40,700 of 74,700). The test partition was not described.

S4d. Availability of references and outcomes by analysis (rounded counts)

Counts as in the aggregate result files (rounded to 10 or 100; fewer than 50 shown as such). The test partition has no cohort table; the comparison by QTc group (Table 2) exists for the development data only. Percentages are shares of the analysed population.

Analysis	Partition	Index ECGs	Analysed population	With reference or outcome label	Reference-positive or exposed
Sokolow-Lyon and Cornell (echocardiographic hypertrophy)	Development	128,400	122,880	20,150 (16%)	5,570
	Test	31,780	30,460	4,940 (16%)	1,330
Peguero-Lo Presti	Development	128,400	116,100	18,440 (16%)	4,960
	Test	31,780	28,740	4,530 (16%)	1,190
Low voltage, effusion	Development	128,400	126,820	4,190 (3%)	230
	Test	31,800	31,400	1,000 (3%)	100
Low voltage, amyloidosis	Development	128,400	200	100	fewer than 10 true positives
	Test	31,800	fewer than 50	fewer than 50	fewer than 50
Q waves (earlier infarct code)	Development	128,400	11,660	11,660	160
	Test	31,800	2,800	2,800	fewer than 50
QTc above 500 ms, death within 30 days	Development	128,400	120,260	104,040 (86%)	2,230 exposed; 280 deaths among 2,100 exposed with a label
	Test	31,780	29,710	about 25,610 (86%)	610 exposed; 570 with a label
QTc above 500 ms, ventricular events	Development	128,400	120,260	40,960 (34%)	2,230 exposed; 1,460 with labels
	Test	31,800	29,700	about 10,200 (34%)	600 exposed; 400 with labels
Right bundle branch block (no documented prior disease)	Development	128,400	87,400 (88,400 evaluable)	death 81,100; atrial fibrillation 69,900	2,800 exposed
	Test	31,800	21,600 (21,900 evaluable)	not tabulated	700 exposed
PR above 200 ms (no documented prior disease)	Development	128,400	109,800 (114,900 evaluable)	death 74,700; atrial fibrillation 65,000	3,100 exposed
	Test	31,800	27,200 (28,400 evaluable)	not tabulated	1,800 exposed
Normal ECG (emergency visits with chest pain and troponin)	Development	128,400	11,100	11,100	1,300 with a normal ECG
	Test	31,800	2,700	2,700	400 with a normal ECG

The QRS threshold by sex (LVEF within 30 days) is not tabulated here; its numbers are in Figure 1 of the main text. Echocardiographic references covered about 16% of evaluable patients for the voltage rules and about 3% for the effusion claim; the weights and the selection grid (S2) address

this only under assumptions. Mortality labels covered 94% of exposed and 86% of unexposed patients, ventricular labels 65% and 33% (S4b).

S5. Result tables and Figure S1

S5a. QTc above 500 ms: all analyses (risk ratio, RR)

Counts in the underlying tables are rounded to 10 (to 100 in the test partition for the ventricular outcome); the interval for the minimally adjusted ventricular estimate in the test partition was unbounded. “Lower one-sided bound” is the Bonferroni-adjusted bound used for the verdict. A risk ratio below 2 can be excluded only if the lower bound exceeds 2; an upper bound above 2 shows only that a ratio above 2 is not excluded. By the lower one-sided bound a ratio below 2 is excluded for death under both definitions in both partitions (4.40, 3.39, 3.10, 2.17) and for ventricular events with Bazett (3.01, 2.03), but not with Fridericia (1.37, 1.81).

Death within 30 days

Analysis	Development RR (95% CI)	Lower one-sided bound	Test RR (95% CI)	Lower one-sided bound
Primary: Bazett, unadjusted	5.07 (4.52 to 5.68)	4.40	4.19 (3.28 to 5.37)	3.10
Concordance: Fridericia, unadjusted	4.31 (3.54 to 5.25)	3.39	3.64 (2.38 to 5.55)	2.17
Linear correction (Framingham)	4.33 (3.52 to 5.33)	n/a	3.44 (2.17 to 5.44)	n/a
QRS below 120 ms only	5.35 (4.55 to 6.29)	n/a	3.78 (2.59 to 5.51)	n/a
Sinus rhythm only	5.56 (4.90 to 6.31)	n/a	3.86 (2.89 to 5.16)	n/a
ECG on the admission day only	not defined for this outcome	n/a	not defined for this outcome	n/a
Minimally adjusted (age, sex, admission type, heart rate)	2.02 (1.78 to 2.29)	n/a	1.53 (1.17 to 2.00)	n/a
Fully adjusted	1.79 (1.59 to 2.02)	n/a	1.52 (1.18 to 1.96)	n/a
Per QTc category step (0 to 3)	2.05 (1.99 to 2.12)	n/a	1.98 (1.86 to 2.10)	n/a

VT, VF or arrest, same admission

Analysis	Development RR (95% CI)	Lower one-sided bound	Test RR (95% CI)	Lower one-sided bound
Primary: Bazett, unadjusted	3.69 (3.12 to 4.35)	3.01	3.18 (2.20 to 4.58)	2.03
Concordance: Fridericia, unadjusted	2.07 (1.48 to 2.89)	1.37	3.46 (2.03 to 5.89)	1.81
Linear correction (Framingham)	2.11 (1.48 to 3.02)	n/a	not shown (fewer than 10 exposed events)	n/a
QRS below 120 ms only	3.35 (2.61 to 4.30)	n/a	2.45 (1.39 to 4.33)	n/a
Sinus rhythm only	3.51 (2.89 to 4.27)	n/a	3.23 (2.15 to 4.85)	n/a
ECG on the admission day only	3.92 (3.27 to 4.70)	n/a	3.34 (2.20 to 5.08)	n/a
Minimally adjusted (age, sex, admission type, heart rate)	2.87 (2.42 to 3.42)	n/a	not estimable	n/a
Fully adjusted	2.52 (2.12 to 3.00)	n/a	2.22 (1.53 to 3.22)	n/a
Per QTc category step (0 to 3)	1.79 (1.70 to 1.87)	n/a	1.61 (1.44 to 1.79)	n/a

S5b. Absolute risks, test-to-development ratios and the share of deaths with QTc above 500 ms (derived from rounded counts; all percentages are approximations)

Denominators: risks are within the analysable cohort of each outcome (death: patients with a 30-day mortality label; ventricular events: patients with an admission record and complete labels). “QTc 500 ms or less” means all values up to and including 500 ms. The share of deaths describes the cohort and is not a diagnostic performance claim. The two test claims have slightly different cohort sizes (31,780 and 31,800 index ECGs; 29,710 and 29,700 evaluable) because of masking of rounded cells.

Quantity	Development	Test partition
Index ECGs; with evaluable QTc; QTc above 500 ms (Bazett)	128,400; 120,260; 2,230 (1.9%)	31,800; 29,700; 600 (2.0%)
Death within 30 days, QTc above 500 ms	about 13% (280 of 2,100)	10.5% (60 of 570)
Death within 30 days, QTc 500 ms or less	2.7% (2,710 of 101,940)	2.6% (650 of 25,040)
Ventricular tachycardia, fibrillation or arrest, QTc above 500 ms	9.6% (140 of 1,460)	not shown (cell masked by rounding)
Ventricular tachycardia, fibrillation or arrest, QTc 500 ms or less	2.7% (1,060 of 39,500)	about 2.0% (200 of 9,800)
Share of deaths within 30 days with index QTc above 500 ms	about 9% (280 of 2,990)	about 8% (60 of 710)
Share of ventricular events with index QTc above 500 ms	about 12% (140 of 1,200)	not shown
Patients flagged by Fridericia and by Bazett (death analysis)	740 and 2,100	200 and 570
Ratio of risk ratios, test over development (95% CI)	death 0.83 (0.63 to 1.09); ventricular 0.86 (0.58 to 1.29)	same

Fewer patients flagged under the Fridericia definition does not show better classification or better clinical decisions; it is reported as a concordance result only.

S5c. Verdicts of all 13 claims (10 rules)

Totals: development 6 supported, 0 not supported, 7 inconclusive; test partition 5 supported, 0 not supported, 8 inconclusive; replicated 5 (Sokolow-Lyon, Cornell, Q waves, QTc ventricular, QTc death), inconclusive in both 7 (Peguero-Lo Presti, low voltage with effusion, right bundle branch block, PR, normal ECG twice, QRS by sex), divergent 1 (low voltage with amyloidosis). Families (letter, number of hypotheses): E voltage, Q-wave and low-voltage claims; A normal ECG; P right bundle branch block and PR (death and atrial fibrillation each); S QTc claims and the QRS threshold by sex; the one-sided limit is 0.025 divided by the family size. Reference-based claims (E and A) used inverse-probability weights for having a reference and the 3 by 3 selection grid (S2); outcome claims (P, S) did not. Terms: a claim is one numeric criterion; a rule can contribute several claims (low voltage 2, normal ECG 2, QTc above 500 ms 2); a replicated claim is supported in both partitions. Classes: supported; not supported (by that definition); inconclusive. “Primary” and “alternative” are the two definitions of a rule; the final verdict needs agreement and, for specificity and sensitivity claims, stability over the selection grid.

Claim	Family (size)	Primary (development)	Alternative (development)	Final, development	Test partition	Note
Sokolow-Lyon	E (8)	supported	supported	supported	supported in both	positivity rate 3.6% (test 3.8%); floor 0.954 (0.951); likelihood ratio 2.43 (2.70)
Cornell voltage	E (8)	supported	supported	supported	supported in both	positivity rate 6.8% (6.5%); floor 0.911 (0.915); likelihood ratio 4.25 (4.25)
Peguero-Lo Presti	E (8)	supported	supported	inconclusive	inconclusive (both partitions)	1 of 9 selection scenarios flips in development; test: not stable either
Low voltage, effusion	E (8)	supported	not supported	inconclusive	inconclusive (both partitions)	2 of 9 scenarios flip; limited identifiability; effusion reference is a documentation proxy; test: primary inconclusive (specificity lower bound below 0.98, sensitivity upper bound above 0.10), alternative not supported, reference in about 1,000 patients

Claim	Family (size)	Primary (development)	Alternative (development)	Final, development	Test partition	Note
Low voltage, amyloidosis	E (8)	supported	supported	supported	inconclusive (divergent replication)	development: about 200 patients, penalised weight model, 2 of 21 covariates with standardised difference above 0.1; test: fewer than 50 patients, sensitivity and its interval not shown (fewer than 50 patients; one-sided bounds 0.000 to 0.670)
Pathological Q waves	E (8)	supported	supported	supported	supported in both	reference may itself rest on ECG findings; positivity rate 7.3% (6.9%); floor 0.926 (0.930); test: fewer than 50 reference-positive patients
Normal ECG, risk 1% or higher	A (2)	inconclusive	inconclusive	inconclusive	inconclusive	1.7% (95% CI 1.0 to 2.6) in development; below 5% (fewer than 10 events) in test
Normal ECG, one third of infarcts	A (2)	not supported	inconclusive	inconclusive	inconclusive	primary 3.5%, alternative 38% (34 to 44); test: primary below 5% (fewer than 10 events), alternative 50%
Right bundle branch block	P (4)	inconclusive	inconclusive	inconclusive	inconclusive	alternative definition not computed in the test partition (columns reset in the second pass)
PR above 200 ms	P (4)	not supported (atrial fibrillation)	inconclusive	inconclusive	inconclusive	primary not supported for fibrillation, supported for death
QTc above 500 ms, ventricular	S (3)	supported	supported	supported	supported in both	concurrent association
QTc above 500 ms, death	S (3)	supported	supported	supported	supported in both	
QRS threshold by sex	S (3)	inconclusive	inconclusive	inconclusive	inconclusive, not testable	4,000 repetitions: 1.6% without a unique crossing point in development, 11% in test; alternative definition not computed in the test partition

S5d. Right bundle branch block and PR above 200 ms (Cox models, hazard ratio, 95% CI)

Claim and endpoint	Development primary	Development alternative	Test primary
Right bundle branch block, death	1.12 (0.98 to 1.27)	1.10 (0.96 to 1.26)	1.23 (0.97 to 1.57)
Right bundle branch block, new atrial fibrillation	1.29 (1.14 to 1.47)	1.38 (1.21 to 1.57)	1.38 (1.08 to 1.76)
PR above 200 ms, death	0.78 (0.66 to 0.91)	0.95 (0.87 to 1.04)	1.07 (0.91 to 1.26)
PR above 200 ms, new atrial fibrillation	1.48 (1.32 to 1.67)	1.22 (1.11 to 1.34)	1.56 (1.36 to 1.79)

Test-to-development ratio of hazard ratios for death with PR above 200 ms: 1.38 (1.10 to 1.73); the cause is not established (the exposure definition differs between the device statement and the alternative, and the populations differ). Absolute 365-day risks (Kaplan-Meier), exposed vs unexposed, development: right bundle branch block, death 10.9% vs 6.1%, atrial fibrillation 14.4% vs 7.0%; PR above 200 ms, death 6.2% vs 6.2%, atrial fibrillation 15.6% vs 6.5%. Patients without any follow-up contact (risk time zero) were excluded from the atrial fibrillation analysis (15,800 and 13,800), which is the recording bias discussed in the Limitations.

S5e. CODE-15% (death only; 95% CI; bounds at 0.0125)

Rule (database label)	Comparison	HR (95% CI)	n; exposed; deaths (exposed)	Verdict
Right bundle branch block	none of six labels	1.14 (1.02 to 1.28)	215,487; 4,650; 6,624 (344)	inconclusive, increase shown
Right bundle branch block	normal ECG	1.34 (1.17 to 1.53)	96,787; 4,650; 1,701 (344)	inconclusive, increase shown
First-degree AV block	none of six labels	1.19 (1.04 to 1.37)	213,378; 2,541; 6,485 (205)	inconclusive, increase shown
First-degree AV block	normal ECG	1.35 (1.14 to 1.59)	94,678; 2,541; 1,562 (205)	inconclusive, increase shown

The claim “hazard ratio below 1.20” is supported only if the upper bound is below 1.20, and not supported only if the lower bound is above 1.20. Neither happened.

S5f. Specificity and sensitivity claims: floor, chance, ceiling and informative targets

Claim	Positivity rate $P(T+)$ (weighted)	Weighted prevalence of the reference	Floor 1 - $P(T+)/P(D-)$	Chance 1 - $P(T+)$	Ceiling $P(T+)/P(D+)$	Margin	Observed	Sensitivity
Sokolow-Lyon, specificity	3.6%	0.233	0.954	0.964		at least 0.90	0.973	0.065
Sokolow-Lyon, specificity (test)	3.8%	0.237	0.951	0.962		at least 0.90	0.973	0.072
Cornell, specificity	6.8%	0.229	0.911	0.932		at least 0.90	0.961	0.166
Cornell, specificity (test)	6.5%	0.234	0.915	0.935		at least 0.90	0.963	0.157
Peguero-Lo Presti, specificity	11.3%	0.219	0.855	0.887		at least 0.90	0.923	0.243
Peguero-Lo Presti, specificity (test)	11.0%	0.224	0.858	0.890		at least 0.90	0.929	0.245
Low voltage (effusion), specificity	1.7%	0.046	0.982	0.983		at least 0.98	0.985	0.052
Low voltage (effusion), specificity (test)	1.8%	0.037	0.982	0.982		at least 0.98	0.983	not shown (fewer than 10 true positives)
Low voltage (effusion), sensitivity	1.7%	0.046			0.36	at most 0.10	0.052	
Low voltage (effusion), sensitivity (test)	1.8%	0.037			about 0.5	at most 0.10	not shown (fewer than 10 true positives)	
Q waves, specificity	7.3%	0.014	0.926	0.927		at least 0.80	0.928	0.158
Q waves, specificity (test)	6.9%	0.010	0.930	0.931		at least 0.80	0.930	not shown (fewer than 50 reference positives)
Low voltage (amyloidosis), sensitivity	about 4%	about 0.65			0.07	at most 0.60	0.07	
Low voltage (amyloidosis), sensitivity (test)	not shown (fewer than 50 patients)	not shown (fewer than 50 patients)			not shown (fewer than 50 patients)	at most 0.60	not shown (fewer than 50 patients)	

Floors are as reported by the analysis (three decimals); chance expectations are 1 minus the weighted positivity rate. Floors are at or above the margin for Sokolow-Lyon, Cornell, Q waves and the primary specificity of the effusion claim, and the ceiling is below the margin for amyloidosis, so the point estimates could not have missed these margins at the observed marginal distributions. This does not decide the verdicts: the confidence bound, the alternative definition and the selection scenarios enter separately (main text, Table 4). For effusion, the floor of the alterna-

tive definition (chest-lead low voltage) was 0.935 in development and 0.929 in the test partition, below the margin of 0.98, and its specificity was 0.943 and 0.942; in the test partition the lower bound of the primary specificity was 0.979. In the test partition the floor equals the observed Q-wave specificity (0.930): the positive results were almost all false positives against the earlier infarct codes (sensitivity not shown, fewer than 50 reference positives). For amyloidosis the observed sensitivity equals the ceiling in both partitions (about 0.07 in development; the test value is not shown, fewer than 50 patients) because no positive result was a false positive (specificity 1.000); the test claim was inconclusive because fewer than 50 patients contributed and the upper bound (0.670) lies above the margin of 0.60. Informative targets for Sokolow-Lyon and Cornell: positive likelihood ratio at least 5 (development 2.43, 95% CI 2.11 to 2.82, and 4.25, 3.88 to 4.67; test 2.70, 2.01 to 3.74, and 4.25, 3.46 to 5.21; not reached, with 5 inside the test interval for Cornell) and sensitivity at most 0.50 (development 0.065, 0.059 to 0.070, and 0.166, 0.158 to 0.175; test 0.072, 0.061 to 0.085, and 0.157, 0.141 to 0.173; reached). Other likelihood ratios reported: Q waves 2.21 in development (test not shown, fewer than 50 reference positives), low voltage for effusion 3.43, Peguero-Lo Presti 3.17. A systematic review of 21 studies in hypertensive patients found median negative likelihood ratios of 0.85 to 0.91 across six indexes (including the Sokolow-Lyon and Cornell voltage indices) and advised against using them to rule out hypertrophy.[2]

S5g. Second test pass: six voltage-based claims in the test partition (post hoc; rounded counts)

Claim	Evaluable patients	With a reference	Positivity P(T+)	Specificity (95% CI)	Sensitivity (95% CI)	LR+ (95% CI)	Bounds for the verdict (one-sided, alpha 0.00313)	Verdict
Sokolow-Lyon	30,460	4,940	3.8%	0.973 (0.968 to 0.978)	0.072 (0.061 to 0.085)	2.70 (2.01 to 3.74)	0.967 to 0.980	supported; alternative supported; selection grid stable
Cornell	30,460	4,940	6.5%	0.963 (0.958 to 0.968)	0.157 (0.141 to 0.173)	4.25 (3.46 to 5.21)	0.956 to 0.970	supported; alternative supported; stable
Peguero-Lo Presti	28,740	4,530	11.0%	0.929 (0.922 to 0.935)	0.245 (0.223 to 0.267)	3.42 (2.93 to 4.00)	specificity 0.919 to 0.938; difference to Cornell 0.056 to 0.111	both parts supported; alternative supported; selection grid not stable; inconclusive
Low voltage, effusion	31,400	about 1,000	1.8%	0.983 (0.980 to 0.986)	not shown (fewer than 10 true positives)	not shown (fewer than 10 true positives)	specificity 0.979 to 0.987; sensitivity 0.003 to 0.102	primary inconclusive (bounds cross the margins); alternative (chest wall) not supported (specificity 0.942); inconclusive

Claim	Evaluable patients	With a reference	Positivity P(T+)	Specificity (95% CI)	Sensitivity (95% CI)	LR+ (95% CI)	Bounds for the verdict (one-sided, alpha 0.00313)	Verdict
Low voltage, amyloidosis	fewer than 50	fewer than 50	not shown (fewer than 50 patients)	n/a	not shown (fewer than 50 patients)	n/a	0.000 to 0.670	inconclusive; alternative supported; penalised weight model, 3 covariates with standardised difference above 0.1
Q waves	2,800	2,800	6.9%	0.930 (0.920 to 0.940)	not shown (fewer than 50 reference positives)	not shown (fewer than 50 reference positives)	0.916 to 0.943	supported; alternative supported

Cells marked “not shown” are withheld because, together with other shown values, they would allow a patient count below 10 to be recalculated. Extraction statistics: 31,776 of 31,776 index ECGs processed; 31,544 “ok”, 186 “signal quality failed” (0.59%), 46 “lead placement excluded” (0.14%), 0 extraction errors. The development numbers of the same claims are in S5c, S5d and the main text. Test claims for right bundle branch block, PR and the QRS threshold were not recomputed in the second pass.

S5h. Heart-rate strata and thresholds (descriptive; development; full table)

The table is Table 5 of the main text. In addition: patients with a heart rate below 60, 60 to 100 and above 100 per minute numbered 18,180, 89,960 and 12,120; the share with QTc above 500 ms was 0.6%, 1.7% and 5.0%. At heart rates below 60 per minute fewer than 10 ventricular events occurred among the 100 exposed patients, so the ratio for ventricular events is not estimable there. Heart rate was available for all evaluable patients. The 500 ms cells are identical to those of the primary analysis (S5a); the thresholds 480 and 520 ms were run on the same population. Ratios are unadjusted and the groups differ in age, setting and illness (Table 2). The risk of death among patients with QTc 500 ms or less was 1.4%, 2.3% and 7.4% in the three strata, and among those with QTc above 500 ms 11.1%, 12.4% and 16.1% (from rounded counts), so the absolute differences were about 9.7, 10.1 and 8.7 percentage points; the fall of the ratio goes with a rising risk in the comparison group, and the data do not identify its cause.

Figure S1

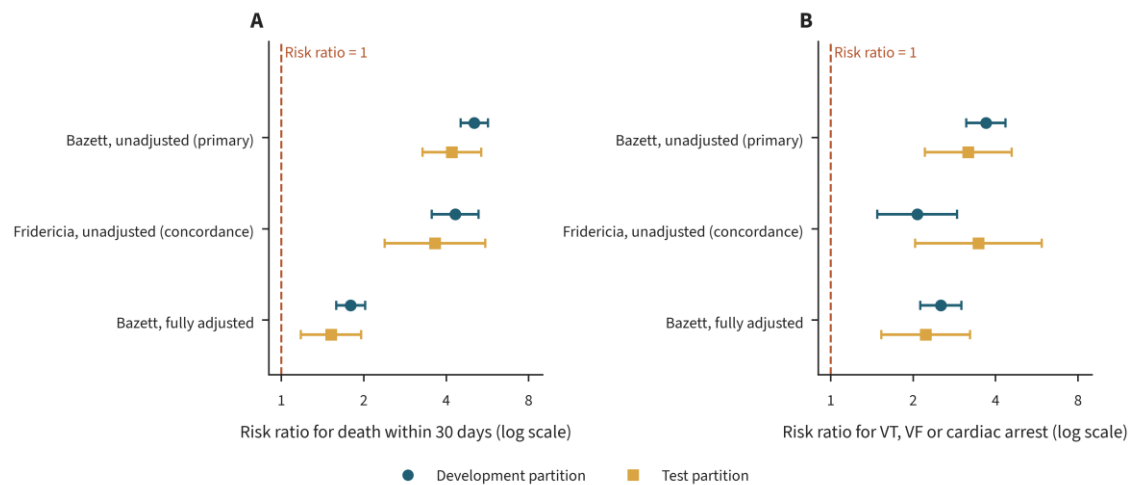


Figure S1. Risk ratios for a corrected QT interval (QTc) above 500 ms versus 500 ms or less, with 95% confidence intervals on a logarithmic scale, in the development partition (circles) and the internal test partition (squares) of MIMIC-IV-ECG. Panel A: death within 30 days. Panel B: ventricular tachycardia (VT), ventricular fibrillation (VF) or cardiac arrest during the same admission (concurrent association with unknown timing). Rows: Bazett correction without adjustment (primary definition); Fridericia correction without adjustment (alternative definition); Bazett correction with full adjustment (age, sex, admission type, heart rate, QT-prolonging drugs, potassium, magnesium, calcium, heart failure; Poisson regression with robust variance). Intervals are Wald intervals for the unadjusted ratios and robust intervals for the adjusted ratios, not the one-sided Bonferroni-adjusted bounds used for the verdicts. The dashed line marks a risk ratio of 1. The analysable cohorts differ by outcome (Table S5b); the numerical values are in Table S5a.

S6. Disclosures and deviations

1. The specification and script were revised after partial results (S1b, S1c). A change that made refutation harder was made for the QTc claims after a test run with few bootstrap repetitions showed them as supported.
2. The test partition had previously been used for a confirmatory analysis of a different marker in a related study and for further analyses (S1e). It is an internal replication of the same hospital, not unseen data and not external validation. The related manuscripts share the database and partitions.
3. The test results of six voltage-based claims come from a second pass that was specified after the results of the other seven test claims were known (S1d). It changed no rule, margin or decision rule, but it is post hoc. The time stated in its rules file was an estimate; the file was saved before the pass was registered (S1d).
4. In the second pass, the Q-wave claim rested on fewer than 50 reference-positive patients, the amyloidosis claim on fewer than 50 patients and the effusion claim on about 1,000 patients with a reference; the test sensitivities are not interpretable and are not shown.
5. For Sokolow-Lyon, Cornell and Q waves the specificity criterion was met in both partitions, but the point estimates could not have missed the margins at the positivity rates of

the rules, so the diagnostic information of the criterion is limited; the likelihood ratio target of 5 was not reached for Sokolow-Lyon and not demonstrated for Cornell (S5f).

6. For the QTc analyses the laboratory window was 24 hours instead of the planned calendar day, and potassium was not filtered for haemolysis. The ventricular outcome is a concurrent association. The adjusted models contain no comorbidity index.
7. Process deviations without effect on results: the resource reservation for the development run was registered a few minutes after the workers started; the test-run log shows two end entries with identical hashes; the first test-run entry was written in the same call as a resource check that reported the cap as exceeded, and the run was started after a new check. In the second pass, four “continuation” lines in the register come from the script’s own continuation function (preparation and three unit processes), without technical restart; a numerical warning (“all-NaN slice”) occurred in a secondary analysis of the amyloidosis claim by sex (cells too small) and does not affect the verdict.
8. The two QTc claims and the sex-specific QRS claim share one correction family (alpha 0.025/3). The CODE-15% reproduction used alpha 0.025/2.
9. Margins were set when positivity rates of several rules were already known (S1b), and the QTc statistics had been computed earlier in the project.
10. The cohort table, the heart-rate strata and the thresholds were specified after results were known and are descriptive (S1f).

S7. Reporting checklist

Reporting guideline: RECORD (extension of STROBE for studies using routinely collected health data).[38] The 13 RECORD items (wording paraphrased) and where each is addressed. “Partly” or “open” marks a gap that is not hidden. STROBE items not specific to routine data are covered by the main text and S2.

RECORD item	Requirement	Where and status
1.1	Type of data and database names in title or abstract	Abstract: MIMIC-IV-ECG linked to MIMIC-IV (electronic health record and device-measured ECG). Reported.
1.2	Region and time frame in title or abstract	Abstract and Methods: Boston, 2008 to 2019. Reported.
1.3	Linkage stated in title or abstract	Abstract: “linked”. Reported.
6.1	Methods of study population selection, codes or algorithms listed	Methods 2.1 and S2; explicit code lists in the repository (code-lists/codelist_k1.csv). Reported.
6.2	Validation studies of the codes or algorithms used to select the population	No validation study of the outcome or reference codes exists for this study; code-based references are labelled as documentation proxies. Not done, stated.
6.3	Flow diagram if databases were linked	Figure 1 shows the numbers at each step per claim; a single linkage count of ECGs to hospital records is not given. Partly.
7.1	Complete list of codes and algorithms for exposures, outcomes, confounders, effect modifiers	Definitions in S2; explicit ICD-9 and ICD-10 lists in the repository. Reported.
12.1	Extent of access to the database population	Credentialed access to the full MIMIC-IV modules; open access to MIMIC-IV-ECG. Declarations and S2. Reported.
12.2	Data cleaning methods	Exclusion rules (usable device metadata, signal quality) and a plausibility check of the inputs (amplitude minimum, QRS and PR medians) in S2. Partly: no row-level cleaning log is published.
12.3	Linkage methods and quality	ECGs were linked to hospital records through the database identifiers; linkage quality was not evaluated. ECG date-times come from device clocks that were not synchronised with an external time source in most cases, so timing relative to admissions carries unquantified error. Stated, not quantified.
13.1	Selection of persons, including filtering for data quality, availability and linkage	Figure 1, S4a and S4b. Reported.
19.1	Implications of data not created for the research question (misclassification, unmeasured confounding, missing data)	Limitations in the main text and S6. Reported.
22.1	Access to protocol, raw data, programming code	Specification and script hashed (S1a); code at release v1.2.0, including the driver of the second test pass and the cohort-table script; patient-level data cannot be shared (data availability statement). Partly.

Additional references (continuing the numbering of the main text where an entry also appears there):

1. Casale PN, Devereux RB, Alonso DR, Campo E, Kligfield P. Improved sex-specific criteria of left ventricular hypertrophy for clinical and computer interpretation of electrocardiograms: validation with autopsy findings. *Circulation*. 1987;75:565-72. doi:10.1161/01.cir.75.3.565
2. Pewsner D, Juni P, Egger M, Battaglia M, Sundstrom J, Bachmann LM. Accuracy of electrocardiography in diagnosis of left ventricular hypertrophy in arterial hypertension: systematic review. *BMJ*. 2007;335:711. doi:10.1136/bmj.39276.636354.AE
3. Drew BJ, Ackerman MJ, Funk M, Gibler WB, Kligfield P, Menon V, et al. Prevention of torsade de pointes in hospital settings: a scientific statement from the American Heart Association and the American College of Cardiology Foundation. *J Am Coll Cardiol*. 2010;55:934-47. doi:10.1016/j.jacc.2010.01.001
4. Fuchigami Y. Machine-reported electrocardiographic right axis deviation or right ventricular hypertrophy and echocardiographic pulmonary hypertension phenotypes and right-heart abnormalities. *Echocardiography*. 2026;43:e70539. doi:10.1111/echo.70539
5. Casale PN, Devereux RB, Kligfield P, Eisenberg RR, Miller DH, Chaudhary BS, et al. Electrocardiographic detection of left ventricular hypertrophy: development and prospective validation of improved criteria. *J Am Coll Cardiol*. 1985;6:572-80. doi:10.1016/s0735-1097(85)80115-7
6. Peguero JG, Lo Presti S, Perez J, Issa O, Brenes JC, Tolentino A. Electrocardiographic criteria for the diagnosis of left ventricular hypertrophy. *J Am Coll Cardiol*. 2017;69:1694-703. doi:10.1016/j.jacc.2017.01.037
7. Eisenberg MJ, de Romeral LM, Heidenreich PA, Schiller NB, Evans GT Jr. The diagnosis of pericardial effusion and cardiac tamponade by 12-lead ECG: a technology assessment. *Chest*. 1996;110:318-24. doi:10.1378/chest.110.2.318
8. Cyrille NB, Goldsmith J, Alvarez J, Maurer MS. Prevalence and prognostic significance of low QRS voltage among the three main types of cardiac amyloidosis. *Am J Cardiol*. 2014;114:1089-93. doi:10.1016/j.amjcard.2014.07.026
9. Rautaharju PM, Surawicz B, Gettes LS, Bailey JJ, Childers R, Deal BJ, et al. AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram: part IV: the ST segment, T and U waves, and the QT interval. *J Am Coll Cardiol*. 2009;53:982-91. doi:10.1016/j.jacc.2008.12.014
10. Glikson M, Nielsen JC, Kronborg MB, Michowitz Y, Auricchio A, Barbash IM, et al. 2021 ESC guidelines on cardiac pacing and cardiac resynchronization therapy. *Eur Heart J*. 2021;42:3427-520. doi:10.1093/eurheartj/ehab364
11. Hancock EW, Deal BJ, Mirvis DM, Okin P, Kligfield P, Gettes LS, et al. AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram:

- part V: electrocardiogram changes associated with cardiac chamber hypertrophy. *Circulation*. 2009;119:e251-61. doi:10.1161/CIRCULATIONAHA.108.191097
12. Byrne RA, Rossello X, Coughlan JJ, Barbato E, Berry C, Chieffo A, et al. 2023 ESC guidelines for the management of acute coronary syndromes. *Eur Heart J*. 2023;44:3720-826. doi:10.1093/eurheartj/ehad191
 13. Probst L. Repeat variation and definition dependence of ECG/VCG markers: an exploratory audit of one implementation (LoopLens 10.1). Preprint submitted to Research Square, October 2026.
 14. Devereux RB, Alonso DR, Lutas EM, Gottlieb GJ, Campo E, Sachs I, et al. Echocardiographic assessment of left ventricular hypertrophy: comparison to necropsy findings. *Am J Cardiol*. 1986;57:450-8. doi:10.1016/0002-9149(86)90771-x
 15. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? *JAMA*. 1994;271:703-7. doi:10.1001/jama.271.9.703
 16. Probst L. Does QRS-T loop coupling add to the QRS-T angle? Frechet distance and reduced ejection fraction in hospital ECGs, with named measurement-error scenarios. Preprint submitted to Research Square, October 2026.
 17. Probst L. Heart-rate-based clock alignment in intensive-care ECGs, serial-ECG matching and administrative follow-up in MIMIC-IV. Preprint submitted to Research Square, October 2026.
 18. Probst L. Same name, different threshold call: QRS and QTc classification across ECG programs and public datasets. Manuscript in preparation. 2026.
 19. Lima EM, Ribeiro AH, Paixao GMM, Ribeiro MH, Pinto-Filho MM, Gomes PR, et al. Deep neural network-estimated electrocardiographic age as a mortality predictor. *Nat Commun*. 2021;12:5117. doi:10.1038/s41467-021-25351-7
 20. Ribeiro AH, Ribeiro MH, Paixao GMM, Oliveira DM, Gomes PR, Canazart JA, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nat Commun*. 2020;11:1760. doi:10.1038/s41467-020-15432-4
 21. Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med*. 2015;12:e1001885. doi:10.1371/journal.pmed.1001885
 22. Sagie A, Larson MG, Goldberg RJ, Bengtson JR, Levy D. An improved method for adjusting the QT interval for heart rate (the Framingham Heart Study). *Am J Cardiol*. 1992;70:797-801. doi:10.1016/0002-9149(92)90562-d