

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

Leo Probst

leo.probst.research@proton.me

Johannes Kepler University of Linz <https://orcid.org/0009-0003-4130-7981>

Research Article

Keywords: electrocardiography, QT interval, left ventricular hypertrophy, hospital patients, routine data, diagnostic criteria, MIMIC-IV-ECG

Posted Date: October 20th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11410506/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Abstract

Background. Simple electrocardiogram (ECG) rules, such as voltage criteria, Q waves and a corrected QT interval (QTc) above 500 ms, are used daily. We found no study that tested several together against numeric criteria in routine hospital data.

Methods. Retrospective analysis of MIMIC-IV-ECG: 128,400 adults for development and 31,800 (rounded) in an internal test partition. Thirteen claims about 10 rules were judged against criteria written down before the final runs, but after partial results of earlier script versions. Six claims were tested in a second pass specified after the other test results were known; the replication is internal.

Results. Six claims were supported in development and 5 in test; none was not supported and 8 were inconclusive in test. For Sokolow-Lyon voltage, Cornell voltage and Q waves, the specificity criterion was met in both partitions; diagnostic information is limited at these marginal distributions (rules positive in 4% to 7%). Sokolow-Lyon and Cornell kept measurable information (positive likelihood ratio 2.70 and 4.25), but the target of 5 was not reached for Sokolow-Lyon and not demonstrated for Cornell (95% confidence interval [CI] 3.46 to 5.21). Q waves rested on fewer than 50 positive patients, too few for a stable sensitivity or likelihood ratio. QTc above 500 ms was associated with death within 30 days (risk ratio 5.07, CI 4.52 to 5.68) and with ventricular events (3.69), as already known, but only about 9% of deaths had a QTc above 500 ms. In a descriptive analysis added after the results were known, the risk ratio for death was 9.5 below 60 and 2.2 above 100 per minute while the comparison group's risk rose from 1.4% to 7.4%; Bazett overcorrection is one possible explanation, not tested.

Conclusions. Measurement definition, reference availability and testability shaped most verdicts, and no rule was refuted.

Key findings.

- Thirteen claims about ten rules were tested against numeric criteria written down before the final runs, with partial earlier results and several positivity rates already known: 6 supported in development, 5 in the test partition, none not supported.
- The specificity criteria for Sokolow-Lyon, Cornell and Q waves were met in both partitions; their diagnostic information is limited at these marginal distributions. Sokolow-Lyon and Cornell keep measurable information (positive likelihood ratio 2.70 and 4.25 in the test partition; target of 5 not reached and not demonstrated), whereas the Q-wave result rests on fewer than 50 reference-positive patients, too few for a stable estimate of sensitivity or likelihood ratio.
- QTc above 500 ms marked patients with about five times the 30-day mortality, a known association, but identified only about 9% of deaths, and adjusted ratios were 1.5 to 2.5.
- The risk ratio for death with QTc above 500 ms fell from 9.5 at heart rates below 60 per minute to 2.2 above 100 (descriptive, after the results were known) while the risk of the comparison group rose from 1.4% to 7.4%; Bazett overcorrection is one possible explanation, not tested here, and no recommendation follows.

Evidence status. Retrospective single-centre analysis. Development data plus an internal test partition of the same database that was used in earlier analyses; no external validation and no blinded validation of measurements or references. The criteria were written down before the final runs, but partial results of earlier script versions and several positivity rates were already known; amendments followed partial results, and the second test pass for six claims, the cohort table, and the heart-rate and threshold analyses were specified after earlier results were known (Supplement S1).

1. Introduction

Cardiologists use ECG rules every day: voltage criteria for left ventricular hypertrophy, pathological Q waves, a PR interval above 200 ms, right bundle branch block, and a QTc above 500 ms. Voltage criteria are specific but insensitive: in an autopsy series the Sokolow-Lyon criterion had a sensitivity of 22% and a specificity of 100%, and the Cornell criterion 42% and 96%;[1] a systematic review in hypertension advised against using them to rule out hypertrophy.[2] A prolonged PR interval and right bundle branch block carry prognostic information in community cohorts.[3, 4] In hospital patients a QTc of 500 ms or more is associated with higher short-term mortality,[5, 6] although not in every cohort.[7, 8] Guidance calls a QTc above 500 ms highly abnormal and links it to a two- to threefold higher risk of torsade de pointes, based on congenital long QT data.[9]

The nearest work in the same database tested one rule at a time: machine-reported right axis deviation against echocardiographic pulmonary hypertension,[10] and conventional hypertrophy criteria against a learned algorithm in 42,016 patients.[11] These findings

come from different populations, definitions and thresholds. We found no earlier study that tested several classical ECG rules together, against numeric criteria written down before the final runs (Section 2.7 states what was known by then), in routine hospital data with device-reported intervals and automatically measured amplitudes. We searched PubMed and Europe PMC up to 16 October 2026. Each rule has been studied separately, and the association of a long QTc with mortality is already established. We therefore did not ask whether these rules carry information; we asked what a test with criteria written down before the final runs can show in routine data: which claims can be supported, which can fail at all, and which cannot be tested because a measurement or reference is missing.

2. Methods

2.1 Data and cohort

We used MIMIC-IV-ECG (version 1.0) linked to MIMIC-IV (Beth Israel Deaconess Medical Center, Boston; ECGs from 2008 to 2019, recorded with devices from various manufacturers).[12–14] Adults (18 years or older) with usable device metadata and signal quality contributed one index ECG each, the first eligible ECG of the person within its partition. The development data are the training and validation partitions (129,224 adults, 640,119 ECGs, 128,400 eligible index ECGs). The test partition is a person-disjoint partition of the same database (31,800 index ECGs). Rule-specific exclusions removed ECGs with device-reported pacing, left bundle branch block (voltage and Q-wave rules) or atrial fibrillation (QTc and PR rules).

2.2 Rules, claims and margins

Candidate rules were compiled before the analysis from guidelines, standardization statements and primary studies (25 candidates). Ten rules were analysed, as 13 claims, each a testable statement about one rule with a numeric criterion (Table 1); the other candidates were excluded because the measurement was not available in the data, the reference could not be constructed, or the rule was already covered by other work (Supplement S3). Rule definitions follow the primary descriptions: Sokolow-Lyon,[1, 15] Cornell,[1, 16] Peguero-Lo Presti,[17] low voltage,[18, 19] Q waves,[20] bundle branch block,[21] PR interval,[22] QTc,[9, 23] and QRS duration thresholds.[24] The numeric margins were set by us (Table 1; origin of each margin in Supplement S2). When we set them, the unweighted positivity rates of several rules in the development ECGs were already known (Sokolow-Lyon 3.4%, Cornell 9.8%, Peguero-Lo Presti 15.5%, low voltage 6.7%). Where we noted an expectation before the analysis, we report it (Supplement S2).

Table 1

The ten rules, their definitions, the claims tested and the origin of the margins. Definitions follow the cited primary descriptions; margins were set by the author. LVH, left ventricular hypertrophy; echo, echocardiographic; LVEF, left ventricular ejection fraction; AF, atrial fibrillation; NSTEMI, non-ST-elevation; UDMI, Universal Definition of Myocardial Infarction.

Rule	Definition used	Claim tested (criterion)	Origin of the margin
Sokolow-Lyon voltage [1, 15]	SV1 + larger of RV5 and RV6 above 3.5 mV	Specific for echo LVH: specificity at least 0.90	Set by the author; upper end of the 85% to 90% range quoted in the standardization statement [25]
Cornell voltage [1]	RaVL + SV3 above 2.8 mV (men) or 2.0 mV (women)	Specific for echo LVH: specificity at least 0.90	Set by the author; same range as for Sokolow-Lyon
Peguero-Lo Presti [17]	Deepest S wave + SV4 at least 2.8 mV (men) or 2.3 mV (women)	Specificity at least 0.90 and more sensitive than Cornell	Specificity of 90% in the original validation; difference above 0
Low voltage, effusion [18]	All six limb leads below 0.5 mV	Specificity at least 0.98 and sensitivity at most 0.10 for moderate or large pericardial effusion	Rounded from the original series (small numbers: 29 positives, 126 negatives)
Low voltage, amyloidosis [19]	All limb leads at most 0.5 mV	Sensitivity at most 0.60 (an insensitive sign)	Prevalence 13% to 60% by definition in the original series
Pathological Q waves [20]	Q-wave criteria of the Universal Definition, as implemented in our software	Specificity at least 0.80 for earlier infarct codes	Set by the author
Normal ECG in chest pain [26]	Device statement “normal ECG” without qualifier	(1) Infarct risk after a normal ECG at least 1%; (2) normal ECG in at least one third of NSTEMI infarcts	(1) Set by the author, oriented on a negative predictive value of 99%; (2) guideline wording “more than one third”
Right bundle branch block [4, 21]	Device statement	Hazard ratio below 1.20 for death and for newly documented AF (no excess risk of 20% or more)	Set by the author (usual non-inferiority margin); before the analysis we expected the claim not to be supported (Supplement S2)
PR above 200 ms [3, 22]	Device PR interval, sinus rhythm	Hazard ratio below 1.20 for death and for newly documented AF	As above
QTc above 500 ms [9]	Bazett correction of device QT and RR	Risk ratio above 1 for death within 30 days and for ventricular events	Direction claim (“marker”)
QRS threshold by sex [24, 27, 28]	Device QRS duration (130 or 150 ms thresholds in guidelines)	Same risk of LVEF 35% or lower at QRS durations differing by less than 10 ms between women and men	Set by the author; guideline simulations quote 9 to 13 ms

2.3 Measurements

Intervals (RR, PR, QRS, QT) were taken from the device measurements of MIMIC-IV-ECG, as stored in the feature table of our analysis pipeline. Amplitudes (voltage and Q-wave criteria) were measured by our own software (LoopLens) and passed to the rule calculations through an interface; this software is the subject of related work.[29] For the test ECGs the amplitudes were extracted in a second pass (Section 2.7). Each claim also had an alternative definition of the same rule (for example, Fridericia instead of Bazett correction, a database label instead of a machine statement, or our own PR measurement).

2.4 References and outcomes

Echocardiographic hypertrophy was a left ventricular mass index above 115 g/m² in men or 95 g/m² in women, reported or computed with the Devereux formula,[30] from the transthoracic echocardiogram nearest to the ECG within 30 days. Further references were a moderate or large pericardial effusion against a small or trivial one, cardiac amyloidosis (amyloidosis code with a cardiomyopathy code or a septum of 12 mm or more), an earlier infarct (infarct code from an earlier stay), and infarction in emergency visits with chest pain and a troponin measurement (definitions in Supplement S2). Outcomes were death within 30 and 365 days, newly documented atrial fibrillation within 365 days (cause-specific hazard, death as competing event), and ventricular tachycardia, fibrillation or cardiac arrest in the same admission. Only a minority of patients had an echocardiogram near the ECG and these were mostly clinically indicated, so reference-based claims used inverse-probability weights for having a reference and a 3 by 3 grid of selection scenarios

(odds of the reference among patients without one 0.5, 1 or 2 times those with one, separately for positive and negative rule results); a verdict counted as stable only if it held over the whole grid.

2.5 Verdicts

For each claim, one-sided Bonferroni-adjusted bounds were computed within families of related claims (family sizes 2 to 8; one-sided α 0.025 divided by the family size) from a person-clustered bootstrap (4,000 repetitions), or from Wald intervals for the QTc claims. A claim was “criterion supported” if the bound met the margin under the primary and the alternative definition, “criterion not supported” if the bound excluded the margin under both, and otherwise “inconclusive”. A claim counted as replicated if it was supported in both partitions; this says nothing about how informative the claim is. Two informative targets were fixed for Sokolow-Lyon and Cornell voltage: a positive likelihood ratio of at least 5 (a conventional threshold for a large change in probability)[31] and a sensitivity of at most 0.50.

To judge whether a criterion could fail at all, we computed from the weighted positivity rate $P(T+)$ and the prevalence of the reference the lowest specificity possible if every positive result were a false positive (floor, $1 - P(T+)/P(D-)$), the specificity of a rule unrelated to the reference (chance, $1 - P(T+)$), and for sensitivity claims the highest sensitivity possible (ceiling, $P(T+)/P(D+)$). If the floor is at or above the margin, the point estimate cannot miss the margin at these marginal distributions; the verdict also depends on the confidence bound, the alternative definition and the selection grid (Section 3.3).

2.6 QTc claims

Exposure was a Bazett QTc above 500 ms (alternative definition: Fridericia), excluding device-reported atrial fibrillation. The primary measure was the unadjusted risk ratio, because the claim is that QTc marks risk. Adjusted models (covariates in Supplement S2) were a sensitivity analysis, together with linear heart-rate correction, sinus rhythm only, QRS below 120 ms, ECGs from the admission day only, and QTc categories. A cohort table (Table 2), an analysis by heart-rate stratum (below 60, 60 to 100, above 100 per minute) and an analysis of the thresholds 480, 500 and 520 ms were added after the results were known; they are descriptive, unadjusted and carry no verdicts.

2.7 Specification, amendments and access history

The specification, the margins and the analysis script (version 3.2) were fixed and hashed before the final runs; the hashes and the amendment history are in Supplement S1. Partial results of earlier script versions had been seen by then, and the specification was amended afterwards in several points (Supplement S1c); no margin was changed. Statistics for the QTc rule against these outcomes had been computed earlier in the project. The test partition had previously been used for a confirmatory analysis of a different marker in a related study, for further bound test runs and for exploratory post hoc analyses of the project;[32] the related manuscripts use the same database, the same partitions and some of the same reference labels, but not these rules.[29, 32–34] For this study it was first analysed in one pass, after a row count without outcomes; seven claims could be computed, because amplitude measurements had not been extracted for the test ECGs.

We then ran a second test pass for the six voltage-based claims, specified after the results of the other seven test claims were known. It changed no rule, margin, population or decision rule. The amplitudes of 31,776 test ECGs (the exact number of index ECGs with amplitudes; test counts elsewhere are rounded to 100) were extracted with the same software version and without access to labels. The unchanged analysis script (version 3.2) ran through a driver script that only supplies these measurements and had reproduced all development results exactly. Three code reviews by a separate model preceded the run. Technical details, hashes, times and a two-minute discrepancy between the time stated in its rules file and the file-system time are in Supplement S1d. The test partition was not unseen for this pass; the replication is internal and post hoc.

2.8 CODE-15%

For right bundle branch block and first-degree atrioventricular block we reproduced the death component in the open CODE-15% subset (233,600 patients with positive follow-up time, 8,341 deaths; Cox model with age spline and sex; bounds at 0.025 divided by 2).[35, 36] Its labels combine the cardiologist’s report with the Glasgow analysis program (right bundle branch block was rejected below a QRS duration of 115 ms, first-degree block below a PR interval of 190 ms),[37] so they are neither our machine definitions nor an independent reading.

2.9 Reporting, ethics and use of AI

Reporting followed STROBE and RECORD (Supplement S7).[38] Counts are rounded to 10 (to 100 for some test cells), counts of 1 to 9 are shown as < 10, and percentages from rounded counts are approximate. Rates, intervals and likelihood ratios that would allow a count below 10 to be recalculated are not shown (marked “not shown”). The author used generative AI tools (Anthropic Claude and OpenAI Codex) to assist with analysis code, literature searches, data analysis workflows, drafting and editing of text, and figure preparation. AI-assisted output was reviewed, corrected where necessary and approved by the author, who takes full responsibility for the content. Patient-level analyses were run locally. AI tools are not listed as authors.

3. Results

3.1 Cohort and flow

Of 129,224 adults in the development partitions, 128,400 had an eligible index ECG; the test partition had 31,800 (Fig. 1). Patients with QTc above 500 ms (Table 2) were older (median 68 vs 57 years), more often seen in the emergency or intensive care unit, and had a faster heart rate (87 vs 73 per minute) and a wider QRS complex (120 ms or more in 48% vs 7.1%). Earlier diagnosis codes were available for 11.3% of all patients (9.4% of those with QTc above 500 ms) and medication data for 32.6% (57.4%), so the shares of heart failure and QT-prolonging drugs describe subgroups. Availability of outcomes and references depended on the analysis (Supplement S4d): a mortality label existed for about 94% of patients with QTc above 500 ms and 86% of the others, ventricular labels for about 65% and 33%, and an echocardiographic reference for about 16% of patients with an evaluable voltage rule. No cohort table exists for the test partition.

Table 2

Characteristics of the development patients with an evaluable QTc, by QTc group (Bazett, device QT and RR; development partitions only). Descriptive, specified after the results were known; counts rounded to 10, percentages from rounded counts. IQR, interquartile range. 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).

Characteristic	QTc above 500 ms	QTc 500 ms or less	All evaluable
Patients	2,230	118,030	120,260
Age, years, median (IQR)	68 (55 to 80)	57 (42 to 70)	57 (42 to 70)
Women	1,120 (50.2%)	62,700 (53.1%)	63,820 (53.1%)
Intensive care unit	140 (6.3%)	3,680 (3.1%)	3,820 (3.2%)
Emergency department (not intensive care)	1,250 (56.1%)	37,790 (32.0%)	39,040 (32.5%)
Ward	270 (12.1%)	13,900 (11.8%)	14,160 (11.8%)
Other (no linked stay)	580 (26.0%)	62,660 (53.1%)	63,240 (52.6%)
Earlier diagnosis codes available (share of all patients)	210 (9.4%)	13,360 (11.3%)	13,570 (11.3%)
Heart failure before the index ECG (share of patients with earlier codes)	30 (14.3%)	270 (2.0%)	300 (2.2%)
Medication data available (share of all patients)	1,280 (57.4%)	37,880 (32.1%)	39,150 (32.6%)
QT-prolonging medication (share of patients with medication data)	320 (25.0%)	11,380 (30.0%)	11,710 (29.9%)
Heart rate, per minute, median (IQR)	87 (74 to 102)	73 (64 to 86)	74 (64 to 86)
QRS duration (device), ms, median (IQR)	116 (94 to 150)	90 (84 to 98)	90 (84 to 99)
QRS duration 120 ms or more	1,070 (48.0%)	8,360 (7.1%)	9,420 (7.8%)

3.2 Overall verdicts

In the development data 6 of 13 claims were supported, none was not supported and 7 were inconclusive (Table 3). In the test partition 5 claims were supported (Sokolow-Lyon, Cornell, Q waves and the two QTc claims), none was not supported and 8 were inconclusive. Five claims were supported in both partitions; 7 were inconclusive in both; the amyloidosis claim was supported in development and inconclusive in the test partition, where fewer than 50 patients had an amyloidosis code. The five claims are not equivalent confirmations: three are specificity criteria whose diagnostic information is limited at these marginal distributions (Section 3.3), and two are direction claims for QTc that set a low bar (Section 3.4). The test results of six voltage-based claims came from the second pass (Section 2.7).

Table 3

Thirteen claims, the clinical question behind each, results and whether the point estimate could miss the margin. Values are estimates with 95% confidence intervals; verdicts rest on one-sided Bonferroni-adjusted bounds (Methods, Section 2.5). Development $n = 128,400$ and test partition $n = 31,800$ index ECGs (analysed numbers in Fig. 1). The test results of the six voltage-based claims come from a second pass specified after the other test results were known (Section 2.7). LR+, positive likelihood ratio. Floor, ceiling and chance are defined in Section 2.5; the last column concerns the point estimate at the observed marginal distributions only, and Table 4 shows how bounds, alternative definitions and selection scenarios can still decide differently. Values that would allow a count below 10 to be recalculated are not shown.

Claim	Clinical question	Development	Test partition	Could the point estimate miss the margin?
Sokolow-Lyon: specificity at least 0.90	Does a positive result reliably show hypertrophy?	Supported. Specificity 0.973 (0.971 to 0.976); sensitivity 0.065; LR + 2.43 (2.11 to 2.82), below the target of 5	Supported in both. 0.973 (0.968 to 0.978); sensitivity 0.072; LR + 2.70 (2.01 to 3.74), target of 5 not reached	No: floor 0.954 (test 0.951)
Cornell: specificity at least 0.90	Does a positive result reliably show hypertrophy?	Supported. 0.961 (0.959 to 0.963); sensitivity 0.166; LR + 4.25 (3.88 to 4.67), target of 5 not reached	Supported in both. 0.963 (0.958 to 0.968); sensitivity 0.157; LR + 4.25 (3.46 to 5.21), target not demonstrated (interval includes 5)	No: floor 0.911 (test 0.915)
Peguero-Lo Presti: specificity at least 0.90 and more sensitive than Cornell	Does it find more hypertrophy than Cornell, with a specificity of at least 0.90?	Inconclusive. Specificity 0.923 (Cornell 0.961); sensitivity 0.243, 0.069 above Cornell; unstable in 1 of 9 selection scenarios	Inconclusive. 0.929 (0.922 to 0.935; Cornell 0.963); sensitivity 0.245, 0.083 above Cornell (0.065 to 0.103); unstable in the selection scenarios	Yes: floor 0.855 (test 0.858)
Low voltage, effusion: specificity at least 0.98 and sensitivity at most 0.10	Does low voltage point to a large effusion?	Inconclusive. Specificity 0.985; sensitivity 0.052; alternative definition specificity 0.943	Inconclusive. 0.983 (0.980 to 0.986); sensitivity not shown (fewer than 10 true positives); alternative definition 0.942, not supported; reference in about 1,000 patients	Primary specificity: no (floor 0.982), but the alternative definition could (floor 0.935) and was not supported; sensitivity: yes (ceiling 0.36; test about 0.5)
Low voltage, amyloidosis: sensitivity at most 0.60	Is low voltage an insensitive sign of cardiac amyloidosis?	Supported. Sensitivity about 0.07 (about 200 patients)	Inconclusive (divergent). Sensitivity not shown (fewer than 50 patients)	No: ceiling about 0.07; test value not shown (fewer than 50 patients), yet inconclusive in test (upper bound 0.670)
Q waves: specificity at least 0.80	Do Q waves point to an earlier infarct?	Supported. Specificity 0.928 (0.924 to 0.933); sensitivity 0.158	Supported in both. 0.930 (0.920 to 0.940); sensitivity and LR + not shown (fewer than 50 reference positives)	No: floor 0.926 (test 0.930)
Normal ECG in chest pain: infarct risk at least 1%	Does a normal ECG exclude infarction?	Inconclusive. 1.7% (1.0 to 2.6)	Inconclusive. Below 5% (fewer than 10 events)	Yes
Normal ECG: at least one third of NSTEMI infarcts	How often is the ECG normal in NSTEMI infarction?	Inconclusive. 3.5% (primary, not supported) or 38% (alternative)	Inconclusive. Primary below 5% (fewer than 10 events) or 50% (alternative)	Yes
Right bundle branch block: hazard ratio below 1.20	Is the excess risk with right bundle branch block below 20%?	Inconclusive. Death 1.12 (0.98 to 1.27); AF 1.29 (1.14 to 1.47)	Inconclusive. 1.23 (0.97 to 1.57); 1.38 (1.08 to 1.76)	Yes
PR above 200 ms: hazard ratio below 1.20	Is the excess risk with a long PR interval below 20%?	Inconclusive. Death 0.78 (0.66 to 0.91); AF 1.48 (1.32 to 1.67; not supported under the primary definition only)	Inconclusive. 1.07 (0.91 to 1.26); 1.56 (1.36 to 1.79)	Yes
QTc above 500 ms: risk ratio above 1, ventricular events	Does it mark patients with arrhythmia or arrest in the same stay?	Supported. 3.69 (3.12 to 4.35)	Supported in both. 3.18 (2.20 to 4.58)	In principle yes; low bar

Claim	Clinical question	Development	Test partition	Could the point estimate miss the margin?
QTc above 500 ms: risk ratio above 1, death within 30 days	Does it mark patients who die soon?	Supported. 5.07 (4.52 to 5.68)	Supported in both. 4.19 (3.28 to 5.37)	In principle yes; low bar
QRS threshold by sex: difference below 10 ms	Do women need another QRS threshold?	Inconclusive. 10.7 ms (-6.5 to 22.5)	Inconclusive. 16.4 ms; not testable (no unique crossing point in 11% of repetitions)	Yes

3.3 Could the criteria fail?

Given the observed positivity rate and the prevalence of the reference, the point estimate of a specificity cannot fall below the floor. For Sokolow-Lyon (floor 0.954 in development, 0.951 in the test partition; margin 0.90), Cornell (0.911, 0.915) and Q waves (0.926, 0.930; margin 0.80) the floor was at or above the margin, so these point estimates could not have missed their margins (Table 4); the same holds for the highest possible sensitivity of low voltage for amyloidosis (about 0.07 in development against the margin of 0.60; in the test partition the value is not shown, fewer than 50 patients, but also lay far below the margin). Observed specificities exceeded chance only slightly (Table 4). In the test partition the Q-wave specificity equalled its floor (0.930): fewer than 50 patients had an earlier infarct code (sensitivity not shown), and almost every positive result was a false positive against that reference. Where no positive result is a false positive, as for amyloidosis, the sensitivity equals the ceiling.

This holds for the point estimate at fixed marginal distributions. A verdict also rests on a one-sided confidence bound, the alternative definition and the selection grid, which can decide differently (Table 4). Low voltage for effusion had a floor (0.982) above its margin (0.98), yet was inconclusive: the alternative definition (chest-lead low voltage; floor 0.935) was not supported (0.943 and 0.942), and in the test partition the lower bound of the primary specificity (0.979) fell below 0.98. Amyloidosis in the test partition could not exceed a point estimate far below the margin of 0.60 (value not shown, fewer than 50 patients), yet its upper bound (0.670) lay above the margin because fewer than 50 patients contributed. For Sokolow-Lyon, Cornell and Q waves, the bounds, alternative definitions and selection grids all met the margins. Peguero-Lo Presti could have failed on specificity (floor 0.855 and 0.858) and met it, but was not stable across selection scenarios.

Not all informative targets were met. The positive likelihood ratio of Sokolow-Lyon was 2.43 (95% confidence interval 2.11 to 2.82) in development and 2.70 (2.01 to 3.74) in the test partition: the target of 5 was not reached. For Cornell it was 4.25 (3.88 to 4.67) and 4.25 (3.46 to 5.21): not reached in development, and not demonstrated in the test partition, whose interval includes 5. The sensitivity target of at most 0.50 was reached (0.065 and 0.166; test 0.072 and 0.157). A positive result thus raised the odds of hypertrophy roughly two- to fourfold, which is modest but measurable information, and most hypertrophy was missed. For Q waves the test sensitivity and positive likelihood ratio are not shown (fewer than 50 reference positives; development likelihood ratio 2.21). Of the expectations noted before the analysis, the Peguero-Lo Presti specificity staying below 0.90 was wrong in both partitions (0.923 and 0.929; Supplement S2).

Table 4

Floor, chance and ceiling against the margin, and the bound and verdict they did not decide. P(T+), weighted positivity rate; prevalence, weighted prevalence of the reference condition; floor, lowest possible specificity if every positive were a false positive; chance, specificity of a rule unrelated to the reference; ceiling, highest possible sensitivity. Floor and ceiling bound the point estimate at the observed marginal distributions; the verdict uses the one-sided bound shown (Bonferroni-adjusted), the alternative definition and the selection grid. Where no positive result is a false positive (amyloidosis), the sensitivity equals the ceiling. Amyloidosis values are rounded or not shown because they rest on fewer than 10 true positives (development) or fewer than 50 patients (test); the same applies to the test sensitivity for effusion. D, development partitions; T, test partition.

Claim	Part.	P(T+)	Prevalence of reference	Floor	Chance	Ceiling	Margin	Observed	Bound used	Verdict
Sokolow-Lyon, specificity	D	3.6%	23.3%	0.954	0.964		at least 0.90	0.973	lower 0.970	supported
	T	3.8%	23.7%	0.951	0.962		at least 0.90	0.973	lower 0.967	supported
Cornell, specificity	D	6.8%	22.9%	0.911	0.932		at least 0.90	0.961	lower 0.957	supported
	T	6.5%	23.4%	0.915	0.935		at least 0.90	0.963	lower 0.956	supported
Peguero-Lo Presti, specificity	D	11.3%	21.9%	0.855	0.887		at least 0.90	0.923	lower 0.919	inconclusive (selection grid)
	T	11.0%	22.4%	0.858	0.890		at least 0.90	0.929	lower 0.919	inconclusive (selection grid)
Low voltage (effusion), specificity	D	1.7%	4.6%	0.982	0.983		at least 0.98	0.985	lower 0.983	inconclusive (primary met; alternative definition not supported)
	T	1.8%	3.7%	0.982	0.982		at least 0.98	0.983	lower 0.979	inconclusive (primary bound below margin; alternative not supported)
Q waves, specificity	D	7.3%	1.4%	0.926	0.927		at least 0.80	0.928	lower 0.922	supported
	T	6.9%	1.0%	0.930	0.931		at least 0.80	0.930	lower 0.916	supported
Low voltage (effusion), sensitivity	D	1.7%	4.6%			0.36	at most 0.10	0.052	upper 0.083	inconclusive (primary met; alternative definition not supported)
	T	1.8%	3.7%			about 0.5	at most 0.10	not shown (fewer than 10 true positives)	upper 0.102	inconclusive (primary upper bound above margin; alternative

Claim	Part.	P(T+)	Prevalence of reference	Floor	Chance	Ceiling	Margin	Observed	Bound used	Verdict
										not supported)
Low voltage (amyloidosis), sensitivity	D	about 4%	about 65%			0.07	at most 0.60	0.07	upper 0.146	supported
	T	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)	upper 0.670	inconclusive (upper bound above margin)

3.4 QTc above 500 ms

QTc exceeded 500 ms in about 1.9% of development patients (2,230 of 120,260). Death within 30 days occurred in about 13% (280 of 2,100) with QTc above 500 ms and 2.7% (2,710 of 101,940) with 500 ms or less (risk ratio 5.07, 4.52 to 5.68). Ventricular tachycardia, fibrillation or arrest was documented in 9.6% (140 of 1,460) against 2.7% (1,060 of 39,500) among patients with an admission record (3.69, 3.12 to 4.35). In the test partition the risk ratios were 4.19 (3.28 to 5.37) for death and 3.18 (2.20 to 4.58) for ventricular events. The claim required only a ratio above 1. A ratio below 2 can be excluded only through the lower limit. The lower one-sided bounds exceeded 2 for death under both definitions in both partitions (Bazett 4.40 in development and 3.10 in the test partition; Fridericia 3.39 and 2.17) and for ventricular events with Bazett (3.01 and 2.03), but not with Fridericia (1.37 and 1.81; lower limits of the 95% intervals 1.48 and 2.03). A ratio below 2 is therefore excluded in six of eight estimates, and for ventricular events under Fridericia it is not (Supplement S5a). Index QTc was above 500 ms in about 9% of the patients who died within 30 days (test about 8%). Fully adjusted ratios were smaller: death 1.79 (1.59 to 2.02) in development and 1.52 (1.18 to 1.96) in the test partition; ventricular events 2.52 (2.12 to 3.00) and 2.22 (1.53 to 3.22; Figure S1).

We added two descriptive analyses after the results were known (Table 5; development data only, unadjusted, no verdicts, no correction for multiple comparisons). By heart rate, the risk ratio for death was 9.5 (95% confidence interval 5.5 to 16.4) below 60 per minute, 5.5 (4.8 to 6.3) at 60 to 100 and 2.2 (1.8 to 2.6) above 100. The risk of death among patients with QTc 500 ms or less rose from 1.4% to 2.3% and 7.4% across these strata, and among those with QTc above 500 ms from 11.1% to 12.4% and 16.1%, so the absolute difference stayed similar (about 9 to 10 percentage points, from rounded counts). For ventricular events the ratio was 3.8 (3.1 to 4.7) at 60 to 100 and 2.9 (2.2 to 3.9) above 100, and not estimable below 60. A QTc above 500 ms was rare at slow rates (0.6% of patients) and common at fast rates (5.0%). By threshold, the risk ratio for death rose from 4.2 (above 480 ms) to 5.1 (500 ms) and 6.4 (520 ms), whereas for ventricular events it was 3.2, 3.7 and 2.5.

Table 5

QTc above 500 ms by heart rate and by threshold (Bazett; development partitions; descriptive, specified after the results were known). Unadjusted risk ratios with Wald 95% confidence intervals, not corrected for multiple comparisons, no verdicts. Counts rounded to 10; events as numbers with the percentage of patients with a known outcome. n/a, not estimable (cell below 10).

Stratum	Patients; with QTc above 500 ms	Outcome	Events with QTc above 500 ms	Events with QTc 500 ms or less	Risk ratio (95% CI)
Heart rate below 60 per minute	18,180; 100 (0.6%)	Ventricular tachycardia, fibrillation or arrest	< 10 of 100	100 (2.1%) of 4,800	n/a
		Death within 30 days	10 (11.1%) of 90	220 (1.4%) of 15,690	9.53 (5.54 to 16.39)
Heart rate 60 to 100 per minute	89,960; 1,530 (1.7%)	Ventricular tachycardia, fibrillation or arrest	90 (9.0%) of 1,000	700 (2.4%) of 28,770	3.82 (3.11 to 4.70)
		Death within 30 days	180 (12.4%) of 1,450	1,740 (2.3%) of 76,060	5.49 (4.76 to 6.34)
Heart rate above 100 per minute	12,120; 600 (5.0%)	Ventricular tachycardia, fibrillation or arrest	50 (12.5%) of 400	240 (4.0%) of 5,960	2.93 (2.19 to 3.92)
		Death within 30 days	90 (16.1%) of 560	750 (7.4%) of 10,200	2.15 (1.76 to 2.63)
Threshold above 480 ms		Ventricular tachycardia, fibrillation or arrest	370 (7.6%) of 4,900	840 (2.3%) of 36,060	3.23 (2.87 to 3.64)
		Death within 30 days	800 (9.7%) of 8,270	2,200 (2.3%) of 95,770	4.19 (3.88 to 4.53)
Threshold above 500 ms (primary)		Ventricular tachycardia, fibrillation or arrest	140 (9.6%) of 1,460	1,060 (2.7%) of 39,500	3.69 (3.12 to 4.35)
		Death within 30 days	280 (13.3%) of 2,100	2,710 (2.7%) of 101,940	5.07 (4.52 to 5.68)
Threshold above 520 ms		Ventricular tachycardia, fibrillation or arrest	40 (8.3%) of 480	1,170 (2.9%) of 40,480	2.51 (1.82 to 3.48)
		Death within 30 days	90 (17.0%) of 530	2,900 (2.8%) of 103,520	6.37 (5.29 to 7.68)

3.5 Other claims

Peguero-Lo Presti had a specificity of 0.923 (test 0.929) against 0.961 (0.963) for Cornell and was more sensitive than Cornell by 0.069 (95% confidence interval 0.060 to 0.078; test 0.083, 0.065 to 0.103), so the gain in sensitivity came with a lower specificity. Low voltage for effusion had a specificity of 0.985 and a sensitivity of 0.052 (test specificity 0.983, reference in about 1,000 patients; test sensitivity not shown, fewer than 10 true positives); for amyloidosis the sensitivity was about 0.07 in about 200 patients (test not shown, fewer than 50 patients). The infarct risk after a normal ECG in chest pain was 1.7% (1.0 to 2.6) in development and below 5% (fewer than 10 events) in the test partition; both left the claim of at least 1% inconclusive. Right bundle branch block was associated with a hazard ratio of 1.12 (0.98 to 1.27) for death and 1.29 (1.14 to 1.47) for newly documented atrial fibrillation; PR above 200 ms, with 0.78 (0.66 to 0.91) and 1.48 (1.32 to 1.67); the lower mortality with PR prolongation did not hold in the test partition (1.07) or with the alternative definition (0.95) and is not interpreted as protective. In CODE-15%, hazard ratios for death were 1.14 (1.02 to 1.28) for right bundle branch block and 1.19 (1.04 to 1.37) for first-degree atrioventricular block, neither supporting nor contradicting a hazard ratio below 1.20. The QRS difference at which women and men have the same risk of an ejection fraction of 35% or lower was 10.7 ms (-6.5 to 22.5), which did not establish equivalence within 10 ms.

4. Discussion

Five claims were supported in both partitions, but they are not five equivalent confirmations. The QTc claims required no more than a risk ratio above 1 for a marker already known to be associated with mortality,[5, 6, 39] and our estimates agree with earlier work: 13% 30-day mortality with a QTc of 500 ms or more in an emergency department,[6] an adjusted hazard ratio of 1.90 in comorbidity-matched hospital patients,[5] and a hazard ratio near 6 with Fridericia correction among sinus-rhythm patients with narrow QRS.[39] Two other cohorts found no such association, one in a small acute medical cohort[7] and one in 257 intensive care patients after

adjustment,[8] so the answer depends on setting and adjustment. For Sokolow-Lyon voltage, Cornell voltage and Q waves, the specificity criterion was met in both partitions, but its diagnostic information is limited at these marginal distributions: the rules were positive in only 4% to 7% of patients, so the point estimate of the specificity could not have fallen below the margin. Sokolow-Lyon and Cornell still carry measurable information (positive likelihood ratios 2.4 to 4.3, sensitivities 0.07 to 0.17), although the target of 5 was not reached or not demonstrated. The Q-wave result rests on fewer than 50 reference-positive patients, too few for a stable estimate of sensitivity or likelihood ratio, and is not informative.

Guidance for hospital patients on a QT-prolonging drug whose QTc exceeds 500 ms or has risen by at least 60 ms is to consider alternative drugs, review interactions and electrolytes, and monitor closely.[9] These data do not change that advice or show that acting on the threshold improves outcomes. Sensitivity for death was low, so the threshold does not identify most patients who die. The heart-rate analysis is descriptive and was added after the results were known. The risk ratio fell with heart rate, but the risk of the comparison group rose from 1.4% to 7.4%, and a ratio falls when its denominator rises even if the absolute excess stays similar. A rising baseline risk may therefore account for much of the fall; the rounded counts do not allow a decomposition. Overcorrection by the Bazett formula at fast rates, which puts more patients with a normal QT interval above 500 ms,[9] is one possible explanation; these data do not test it, and the groups differ in age, setting and illness. We draw no recommendation from this pattern. Examining it would need absolute risks and a comparison of correction formulas with adjustment within the same heart-rate strata. Adjusted ratios stayed above 1; the models contained no comorbidity index or vital signs.

The voltage results agree with the autopsy and review literature: high specificity and low sensitivity,[1, 2] with a Peguero-Lo Presti sensitivity of 0.243 and 0.245 close to a recent registry (0.31 at specificity 0.94 in hypertension).[40] Low voltage for effusion behaved as in the original small series,[18] whereas the sensitivity of low voltage for amyloidosis in development (about 0.07) was far below the prevalence of low limb-lead voltage in an amyloidosis series (34%);[19] we did not resolve the reason. The share of normal ECGs among non-ST-elevation infarcts depended on the definition of normal (3.5% or 38%; 7.9% of first ECGs in a large registry).[41] Right bundle branch block and PR prolongation were associated with a higher hazard of newly documented atrial fibrillation (point estimates 1.29 and 1.48), as in community cohorts,[3, 4] but the claims were inconclusive, and for PR with the primary definition a hazard ratio below 1.20 for atrial fibrillation was not supported. Their hazard ratios for death were lower than in those cohorts; we cannot say whether this reflects selection, exposure measurement or a real difference.

What the study does not show. It does not show that any rule is valid or invalid beyond the single claims and margins tested, that the QTc threshold causes the risk or should be acted on differently, or that the ten rules rank by clinical reliability; eight claims were inconclusive in the test partition. The specificity criteria that were met do not show that the voltage and Q-wave rules work. It does not replace a blinded validation of the measurements and references.

Limitations

All data come from one hospital. Intervals are device measurements of unrecorded algorithm and version,[42] and outcomes and references come from codes and records without blinded validation. Amplitudes come from our software and were not checked against readers, and the earlier infarct code may rest on Q waves itself; validating both would strengthen the voltage and Q-wave claims. Echocardiograms were available for a minority of patients and were clinically indicated; weighting and scenarios address this only under assumptions. The QTc claims are unadjusted markers, and the ventricular outcome is concurrent with the ECG. The test partition belongs to the same hospital and had been used before; no cohort table exists for it. The second pass for six voltage-based claims was specified after the other seven test results were known; the Q-wave and amyloidosis claims rested on fewer than 50 positive patients and the effusion claim on about 1,000 patients with a reference. The QTc rule had been examined earlier and margins were set with the positivity rates known; the cohort table and the heart-rate and threshold analyses were specified after results were known. The label tables derive from an earlier version of the project's pipeline whose database version is not recorded in the run reports (diagnosis dictionary: version 3.1).

5. Conclusions

In routine hospital data, the verdicts on classical ECG rules depended less on the rules than on how they were measured, which reference was available, whether a criterion could fail and how many patients were positive or had a reference. Five claims were supported in both partitions and none was refuted, but they differ in what they show: the QTc claims confirm a known association at a low bar, the specificity criteria for Sokolow-Lyon and Cornell were met with modest remaining information, and the Q-wave result is uninformative. Applying one procedure to all rules shows, for each claim, whether routine data can answer it.

Declarations

Code availability

Analysis scripts are available at <https://github.com/leoprobst/ecg-auswertung-code> (release v1.2.0, folder pub13_classical_rules; MIT licence); the frozen LoopLens snapshot in that repository is under PolyForm Noncommercial 1.0.0. This follows clause 8 of the PhysioNet Credentialed Health Data License 1.5.0. Release v1.2.0 includes the driver script of the second test pass and the script of the cohort table. Aggregate results are provided in the Supplementary Material.

Data availability statement

MIMIC-IV-ECG (version 1.0; Gow et al.; <https://doi.org/10.13026/4nqg-sb35>; ODbL 1.0) is openly available on PhysioNet.[12] The MIMIC-IV clinical modules (version 3.1 for the diagnosis dictionary; Johnson et al. 2024, <https://doi.org/10.13026/kpb9-mt58>; original publication Johnson et al. 2023)[13,14] are available to credentialed users under the PhysioNet Credentialed Health Data License 1.5.0; the label tables were derived from an earlier pipeline version that is not recorded in the run reports. CODE-15% (<https://doi.org/10.5281/zenodo.4916206>; CC BY 4.0) is openly available on Zenodo.[35] The standard citation of the PhysioNet resource is given in reference [43]. Derived patient-level data cannot be shared under the data use agreement. Aggregate results are provided in the Supplementary Material (tables S5a to S5h, Figure S1). No individual signals, identifiers or private joins are redistributed. A table linking each numerical statement to its aggregate source is available from the author.

Ethical Approval and Consent to participate

This is a secondary analysis of de-identified research databases. The Institutional Review Boards of Beth Israel Deaconess Medical Center and the Massachusetts Institute of Technology approved the MIMIC-IV-ECG resource and waived individual consent.[12] The Institutional Review Board of Beth Israel Deaconess Medical Center approved the sharing of the MIMIC-IV research resource and granted a waiver of informed consent.[14] Access to the credentialed MIMIC-IV modules was obtained under the PhysioNet Credentialed Health Data Use Agreement after completion of the required CITI training. CODE-15% is an openly released, anonymised subset of the CODE cohort of the Telehealth Network of Minas Gerais; the CODE study reported approval by the Research Ethics Committee of the Universidade Federal de Minas Gerais, which did not require informed consent for the secondary analysis of anonymised data.[37] No attempt was made to re-identify individuals. The present analysis recruited no participants and collected no new data. Formal ethics review of this secondary analysis was not sought because it used only de-identified or anonymised data that are public or accessed under their data use agreements.

Consent for publication

Not applicable. The manuscript contains no individual patient data, images or identifiers.

Availability of data and materials

MIMIC-IV-ECG (version 1.0; <https://doi.org/10.13026/4nqg-sb35>; ODbL 1.0) is openly available on PhysioNet. The MIMIC-IV clinical modules (<https://doi.org/10.13026/kpb9-mt58>; version 3.1 for the diagnosis dictionary) are available to credentialed users under the PhysioNet Credentialed Health Data License 1.5.0. CODE-15% (<https://doi.org/10.5281/zenodo.4916206>; CC BY 4.0) is openly available on Zenodo. Derived patient-level data cannot be shared under the data use agreement. Aggregate results are in the Supplementary Material. No individual signals, identifiers or private joins are redistributed. Analysis scripts are at <https://github.com/leoprobst/ecg-auswertung-code> (release v1.2.0, folder pub13_classical_rules).

Competing interests

The author is the developer of LoopLens, the software used for the analyses (non-financial interest). The author has no financial interest in LoopLens or in any related company or product, and no other competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The work was self-funded by the author.

Authors' contributions

L.P. conceived the study, developed the software, curated the data, performed the analyses, prepared the figures and wrote the manuscript. L.P. approved the final version.

Acknowledgements

The author thanks the PhysioNet team and the investigators who created and maintain MIMIC-IV, MIMIC-IV-ECG and CODE-15%.

Use of generative AI

The author used generative AI tools (Anthropic Claude and OpenAI Codex) to assist with analysis code, literature searches, data analysis workflows, drafting and editing of text, and figure preparation. AI-assisted output was reviewed, corrected where necessary and approved by the author, who takes full responsibility for the content. Patient-level analyses were run locally. AI tools are not listed as authors.

Registration and related work

The study was not prospectively registered. The specification was fixed and hashed before the final runs (Supplement S1); this is not prospective registration. All claims of the same specification are reported here. The author's related manuscripts use the same database and partitions (references [29,32-34]); this manuscript has not been posted or published elsewhere.

References

1. Casale PN, Devereux RB, Alonso DR, Campo E, Kligfield P (1987) Improved sex-specific criteria of left ventricular hypertrophy for clinical and computer interpretation of electrocardiograms: validation with autopsy findings. *Circulation* 75:565–572. 10.1161/01.cir.75.3.565
2. Pewsner D, Juni P, Egger M, Battaglia M, Sundstrom J, Bachmann LM (2007) Accuracy of electrocardiography in diagnosis of left ventricular hypertrophy in arterial hypertension: systematic review. *BMJ* 335:711. 10.1136/bmj.39276.636354.AE
3. Cheng S, Keyes MJ, Larson MG, McCabe EL, Newton-Cheh C, Levy D et al (2009) Long-term outcomes in individuals with prolonged PR interval or first-degree atrioventricular block. *JAMA* 301:2571–2577. 10.1001/jama.2009.888
4. Bussink BE, Holst AG, Jespersen L, Deckers JW, Jensen GB, Prescott E (2013) Right bundle branch block: prevalence, risk factors, and outcome in the general population: results from the Copenhagen City Heart Study. *Eur Heart J* 34:138–146. 10.1093/eurheartj/ehs291
5. Gibbs C, Thalamus J, Kristoffersen DT, Svendsen MV, Holla OL, Heldal K et al (2019) QT prolongation predicts short-term mortality independent of comorbidity. *Europace* 21:1254–1260. 10.1093/europace/euz058
6. Anderson HN, Bos JM, Haugaa KH, Morlan BW, Tarrell RF, Caraballo PJ et al (2018) Prevalence and outcome of high-risk QT prolongation recorded in the emergency department from an institution-wide QT alert system. *J Emerg Med* 54:8–15. 10.1016/j.jemermed.2017.08.073
7. Mahmud R, Gray A, Nabeebaccus A, Whyte MB (2018) Incidence and outcomes of long QTc in acute medical admissions. *Int J Clin Pract* 72:e13250. 10.1111/ijcp.13250
8. Russell H, Churilov L, Toh L, Eastwood GM, Bellomo R (2019) The incidence, predictors and outcomes of QTc prolongation in critically ill patients. *J Crit Care* 54:244–249. 10.1016/j.jcrc.2019.09.014
9. Drew BJ, Ackerman MJ, Funk M, Gibler WB, Kligfield P, Menon V et al (2010) 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* 55:934–947. 10.1016/j.jacc.2010.01.001
10. Fuchigami Y (2026) Machine-reported electrocardiographic right axis deviation or right ventricular hypertrophy and echocardiographic pulmonary hypertension phenotypes and right-heart abnormalities. *Echocardiography* 43:e70539. 10.1111/echo.70539
11. Huang JT, Tseng CH, Huang WM, Yu WC, Cheng HM, Chao HL et al (2025) Comparison of machine learning and conventional criteria in detecting left ventricular hypertrophy and prognosis with electrocardiography. *Eur Heart J Digit Health* 6:252–260. 10.1093/ehjdh/ztaf003
12. [dataset] Gow B, Pollard T, Nathanson LA, Johnson A, Moody B, Fernandes C et al (2023) MIMIC-IV-ECG: diagnostic electrocardiogram matched subset (version 1.0). PhysioNet. 10.13026/4nqg-sb35

13. [dataset] Johnson A, Bulgarelli L, Pollard T, Gow B, Moody B, Horng S et al (2024) MIMIC-IV (version 3.1). PhysioNet. 10.13026/kpb9-mt58
14. Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S et al (2023) MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data* 10:1. 10.1038/s41597-022-01899-x
15. Sokolow M, Lyon TP (1949) The ventricular complex in left ventricular hypertrophy as obtained by unipolar precordial and limb leads. *Am Heart J* 37:161–186. 10.1016/0002-8703(49)90562-1
16. Casale PN, Devereux RB, Kligfield P, Eisenberg RR, Miller DH, Chaudhary BS et al (1985) Electrocardiographic detection of left ventricular hypertrophy: development and prospective validation of improved criteria. *J Am Coll Cardiol* 6:572–580. 10.1016/s0735-1097(85)80115-7
17. Peguero JG, Lo Presti S, Perez J, Issa O, Brenes JC, Tolentino A (2017) Electrocardiographic criteria for the diagnosis of left ventricular hypertrophy. *J Am Coll Cardiol* 69:1694–1703. 10.1016/j.jacc.2017.01.037
18. Eisenberg MJ, de Romeral LM, Heidenreich PA, Schiller NB, Evans GT Jr (1996) The diagnosis of pericardial effusion and cardiac tamponade by 12-lead ECG: a technology assessment. *Chest* 110:318–324. 10.1378/chest.110.2.318
19. Cyrille NB, Goldsmith J, Alvarez J, Maurer MS (2014) Prevalence and prognostic significance of low QRS voltage among the three main types of cardiac amyloidosis. *Am J Cardiol* 114:1089–1093. 10.1016/j.amjcard.2014.07.026
20. Thygesen K, Alpert JS, Jaffe AS, Chaitman BR, Bax JJ, Morrow DA et al (2018) Fourth universal definition of myocardial infarction (2018). *Circulation* 138:e618–e651. 10.1161/CIR.0000000000000617
21. Surawicz B, Childers R, Deal BJ, Gettes LS, Bailey JJ, Gorgels A et al (2009) AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram: part III: intraventricular conduction disturbances. *J Am Coll Cardiol* 53:976–981. 10.1016/j.jacc.2008.12.013
22. Kusumoto FM, Schoenfeld MH, Barrett C, Edgerton JR, Ellenbogen KA, Gold MR et al (2019) 2018 ACC/AHA/HRS guideline on the evaluation and management of patients with bradycardia and cardiac conduction delay. *Circulation* 140:e382–482. 10.1161/CIR.0000000000000628
23. Rautaharju PM, Surawicz B, Gettes LS, Bailey JJ, Childers R, Deal BJ et al (2009) 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* 53:982–991. 10.1016/j.jacc.2008.12.014
24. Glikson M, Nielsen JC, Kronborg MB, Michowitz Y, Auricchio A, Barbash IM et al (2021) 2021 ESC guidelines on cardiac pacing and cardiac resynchronization therapy. *Eur Heart J* 42:3427–3520. 10.1093/eurheartj/ehab364
25. Hancock EW, Deal BJ, Mirvis DM, Okin P, Kligfield P, Gettes LS et al (2009) AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram: part V: electrocardiogram changes associated with cardiac chamber hypertrophy. *Circulation* 119:e251–e261. 10.1161/CIRCULATIONAHA.108.191097
26. Byrne RA, Rossello X, Coughlan JJ, Barbato E, Berry C, Chieffo A et al (2023) 2023 ESC guidelines for the management of acute coronary syndromes. *Eur Heart J* 44:3720–3826. 10.1093/eurheartj/ehad191
27. Strauss DG, Selvester RH, Wagner GS (2011) Defining left bundle branch block in the era of cardiac resynchronization therapy. *Am J Cardiol* 107:927–934. 10.1016/j.amjcard.2010.11.010
28. Zusterzeel R, Selzman KA, Sanders WE, Canos DA, O'Callaghan KM, Carpenter JL et al (2014) Cardiac resynchronization therapy in women: US Food and Drug Administration meta-analysis of patient-level data. *JAMA Intern Med* 174:1340–1348. 10.1001/jamainternmed.2014.2717
29. Probst L (October 2026) Repeat variation and definition dependence of ECG/VCG markers: an exploratory audit of one implementation (LoopLens 10.1). Preprint submitted to Research Square
30. Devereux RB, Alonso DR, Lutas EM, Gottlieb GJ, Campo E, Sachs I et al (1986) Echocardiographic assessment of left ventricular hypertrophy: comparison to necropsy findings. *Am J Cardiol* 57:450–458. 10.1016/0002-9149(86)90771-x
31. Jaeschke R, Guyatt GH, Sackett DL (1994) 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* 271:703–707. 10.1001/jama.271.9.703
32. Probst L (October 2026) 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
33. Probst L (October 2026) Heart-rate-based clock alignment in intensive-care ECGs, serial-ECG matching and administrative follow-up in MIMIC-IV. Preprint submitted to Research Square

34. Probst L (2026) Same name, different threshold call: QRS and QTc classification across ECG programs and public datasets. Manuscr preparation
35. [dataset] Ribeiro AH, Paixao GMM, Lima EM, Ribeiro MH, Pinto Filho MM, Gomes PR et al (2021) CODE-15%: a large scale annotated dataset of 12-lead ECGs (version 1.0.0). 10.5281/zenodo.4916206. Zenodo
36. Lima EM, Ribeiro AH, Paixao GMM, Ribeiro MH, Pinto-Filho MM, Gomes PR et al (2021) Deep neural network-estimated electrocardiographic age as a mortality predictor. *Nat Commun* 12:5117. 10.1038/s41467-021-25351-7
37. Ribeiro AH, Ribeiro MH, Paixao GMM, Oliveira DM, Gomes PR, Canazart JA et al (2020) Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nat Commun* 11:1760. 10.1038/s41467-020-15432-4
38. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I et al (2015) The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med* 12:e1001885. 10.1371/journal.pmed.1001885
39. Vandenberg B, Vandael E, Robyns T, Vandenberghe J, Garweg C, Foulon V et al (2016) Which QT correction formulae to use for QT monitoring? *J Am Heart Assoc* 5:e003264. 10.1161/JAHA.116.003264
40. Mei DA, Serafini K, Imberti JF, Vitolo M, Pollicino F, Battigaglia E et al (2026) Detection of left ventricular hypertrophy through ECG: diagnostic performance and comparison of ten ECG scores in different hypertrophic conditions. *Eur J Intern Med* 146:106725. 10.1016/j.ejim.2026.106725
41. Welch RD, Zalski RJ, Frederick PD, Malmgren JA, Compton S, Grzybowski M et al (2001) Prognostic value of a normal or nonspecific initial electrocardiogram in acute myocardial infarction. *JAMA* 286:1977–1984. 10.1001/jama.286.16.1977
42. Kligfield P, Badilini F, Denjoy I, Babaeizadeh S, Clark E, De Bie J et al (2018) Comparison of automated interval measurements by widely used algorithms in digital electrocardiographs. *Am Heart J* 200:1–10. 10.1016/j.ahj.2018.02.014
43. Pollard T, Moody B, Lehman LW, Gow B, Fernandes C, Xie C et al (2026) PhysioNet as a global platform for biomedical research. *Nat Health* 1:792–795. 10.1038/s44360-026-00096-z

Figures

	Eligible index ECGs	Analysis population	Reference or outcome available
	Development partitions (blue), test partition (rust)		
Sokolow-Lyon, Cornell	128,400 31,800	122,880 30,460	20,150 (echo) 4,940
Peguerio-Lo Presti	128,400 31,800	116,100 28,740	18,440 (echo) 4,530
Low voltage, effusion	128,400 31,800	126,820 31,400	4,190 (echo) 1,000
Low voltage, amyloidosis	128,400 31,800	200 (code) <50	100 (echo or code) <50
Q waves	13,960 with earlier stay 3,300	11,660 2,800	11,660 (codes) 2,800 (<50 positive)
Normal ECG, chest pain	11,100 ED visits 2,700	1,300 normal ECG 400	900 / 300 (claims 1 / 2) 200 / 100
Right bundle branch block	128,400 31,800	87,400 21,600	81,100 death; 69,900 AF 20,200 death; 17,400 AF
PR above 200 ms	128,400 31,800	109,800 27,200	74,700 death; 65,000 AF 25,600 death; 22,400 AF
QRS threshold by sex	128,400 31,800	126,770 31,740	23,240 (LVEF) 5,730
QTc above 500 ms	128,400 31,800	120,260 29,700	104,040 death; 40,960 ventricular 25,610 death; 10,200 ventricular

Figure 1

Numbers of ECGs analysed for each claim in the development partitions (upper line, blue) and the test partition (lower line, rust). The first column shows eligible index ECGs (one per adult), the second the analysis population after the rule could be evaluated and exclusions were applied, the third the patients with the reference (echocardiogram, code) or outcome. For amyloidosis and Q waves in the test partition, fewer than 50 patients are shown as "<50". ECG, electrocardiogram; ED, emergency department; AF, atrial fibrillation; LVEF, left ventricular ejection fraction. Counts are rounded to 10 or 100 and masked below 10. Development partitions: 129,224 adults, of whom 128,400 had an eligible index ECG.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ProbstclassicalECGrulessupplement.pdf](#)
- [FigS1pub13v3.png](#)