

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

Leo Probst

`leo.probst.research@proton.me`

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

Research Article

Keywords: electrocardiography, vectorcardiography, ventricular dysfunction, QRS-T angle, Frechet distance, measurement error

Posted Date: October 20th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11410133/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

The spatial QRS-T angle is a marker of left ventricular dysfunction. We asked whether a distance between the QRS and T loops (normalized discrete Frechet distance) carries information beyond the measured angle or is a noisy proxy for it.

Methods

Linked MIMIC-IV ECG, clinical and echocardiography data (development folds and a person-disjoint internal assessment partition used before). The primary model related the distance to left ventricular ejection fraction (LVEF) below 40%, adjusted for age, sex, heart rate, QRS duration, QTc and the integral angle. Later analyses were post hoc scenarios with tipping points for attenuation (odds ratio, OR, 1.25) and removal (OR 1.0).

Results

Among 8,371 persons (800 cases) the OR per development standard deviation (SD) was 2.34 (95% CI 2.09–2.62). With the distance in the model, the angle OR fell from 1.79 to 1.01 (training; Spearman correlation 0.75), and regression calibration was uninformative (OR 2.09, 95% CI 0.42–9.55). In development data the OR stayed at 2.28–2.60 across orientation comparators and 2.32–2.61 after excluding ECGs with pacing indicators. Without error correction, a differential marker over-measurement in cases of 0.35–0.38 control SD lowered the OR to 1.25; after correction for classical angle error, itself unstable, 0.15–0.20 SD sufficed at three main anchors. A simulation at true OR 1.0 reproduced an apparent OR of 2.49. Imputed marker values shifted by 1.6–1.8 development SD also reached 1.25. A selection factor B of 1.9–2.0 would attenuate the OR to 1.25; its feasibility depended on model assumptions, and selection could not be excluded in general.

Conclusion

The association with low LVEF remained after adjustment for the measured angle, but we cannot show that the distance carries information independent of true orientation: small differential measurement error would attenuate it. External replication is outstanding and clinical utility is untested.

Key findings

- The Fréchet distance between QRS and T loops was associated with LVEF below 40% after adjustment for the measured QRS-T angle: OR 2.34 per development SD (95% CI 2.09–2.62; 8,371 persons, internal assessment partition).

- With the distance in the model the angle OR fell from 1.79 to 1.01 (Spearman correlation 0.75 between the two). The OR was similar with additional orientation measures (2.39 with all), echo intervals of 1 to 30 days (2.34–2.48) and exclusion of ECGs with pacing indicators (2.32–2.61; development data, pacing spikes in the signal not assessed). The gain in discrimination was small and largely shared with orientation descriptors.
- Measurement error remains a possible explanation. Without error correction, a differential marker over-measurement in cases of 0.35–0.38 control SD (about 30% of the case-control difference) lowers the OR to 1.25; after correction for classical angle error, which is itself unstable, 0.15–0.20 control SD suffices at three main anchors (about 0.25 at an assumed technical-agreement anchor).
- Missing marker values (imputed values shifted by 1.6–1.8 development SD) and echo selection (selection factor B of 1.9–2.0 per marker SD) reduce the OR to 1.25 in separate, assumption-dependent scenarios. Selection could not be excluded in general, and combinations of biases are not bounded.
- Independence from true orientation, external replicability and clinical utility are not established; no diagnostic use or treatment threshold is supported.

Evidence status

The primary association met its specified internal rule in a person-disjoint partition previously used for related hypotheses. Everything else is exploratory: development sensitivities after marker selection, and all assessment and scenario analyses specified after the primary result was known. Scenario analyses give the size of an assumed bias at which the OR would reach 1.25 (attenuation) or 1.0 (removal), not evidence that the bias is absent or present. No independent external replication is available.

Introduction

Is a distance between the QRS and T loops more than a noisy QRS-T angle? Three questions must be kept apart: whether the distance keeps an association with low ejection fraction after adjustment for the measured angle (tested here), whether it measures true orientation better than the angle does (not tested), and whether it adds to clinical decisions (not tested). The clinical background is that an electrocardiographic indication of left ventricular dysfunction could help prioritize echocardiography [1], and that ventricular function enters the assessment for cardiac resynchronization therapy (CRT) together with symptoms and conduction [2]; we make no claim about either.

Spatial QRS-T angles describe the orientation of depolarization relative to repolarization. Associations with ventricular function [3, 4], hypertrophic cardiomyopathy [5], outcomes [6, 7] and reviews [8] predate this work, and mean- and peak-vector definitions and reconstruction matrices change the measurement [7]. Total cosine R-to-T (TCRT) belongs to a wider descriptor family with heterogeneous prognostic evidence [9, 10]. The spatial ventricular gradient keeps magnitude and direction [11, 12]; planarity, T-loop morphology and point-to-point loop variability describe geometry [13–16]. We use “coupling”

descriptively for the geometric similarity of the normalized QRS and T loops; no electrophysiological coupling is implied.

A distance between separately normalized QRS and T trajectories retains orientation and geometry while removing absolute amplitude and the original amplitude ratio from the distance itself (centering each loop at its centroid also removes the displacement between the loops); the delineation that defines both trajectories is not amplitude-invariant. We evaluated representative and beat-median Fréchet distances against reduced LVEF and cardiomyopathy codes and asked whether they kept an association after adjustment for the measured integral angle. We claim neither formula priority nor shape independence. Because both quantities derive from the same loops, measurement error in the angle can create an apparent marker effect. We therefore examine named error scenarios and report where each would tip the result. Neither statistical significance nor incremental area under the receiver-operating-characteristic curve (AUC) establishes clinical usefulness [17–19].

Methods

Design, sources and disclosure

This retrospective cross-sectional study linked MIMIC-IV-ECG version 1.0 to clinical records from MIMIC-IV version 3.1 [20–23] at Beth Israel Deaconess Medical Center. Demographics, encounters, laboratory results, medication orders and diagnosis codes came from the hosp module. LVEF came from the separately supplied MIMIC-IV-ECHO structured measurements (release 1.0) [24], joined to the ECGs by person and time through an echo study index derived from the structured measurements; it was not extracted from report text or ICU bedside measurements (Section S5). The modules were linked by the person identifier (subject_id), the link key given in the documentation of MIMIC-IV-ECG [20] and MIMIC-IV-ECHO [24]; the latter describes one deterministic mapping of medical record number to subject_id across all MIMIC-IV modules. ECGs were linked to a hospital stay or ED stay by subject and by whether the ECG time fell between admission and discharge (or ED arrival and departure), which defined the coded-diagnosis endpoint, device and pacing indicators and the same-stay preference for laboratory values, whereas echocardiograms were linked to ECGs only by subject and the time difference (within ± 30 days for the reference, selected from a ± 90 -day search) without reference to any stay. The documentation does not state explicitly that subject_id is unchanged between MIMIC-IV version 2.2, on which both resources were built, and version 3.1. The ECG resource covers 2008–2019 [20]. Training and internal validation (development) informed the analyses. The assessment partition contained persons absent from development and drawn from the same database; coefficients were fitted anew there, so its AUC values are apparent fit performance.

Computations were run by the author's AI-assisted analysis pipeline; we had not inspected study-related outputs before the association specification (Section S1). The partition had already been used for related hypotheses. This was an analyst-blinded specification after partition opening, with internal timestamps and no external preregistration. Section S1 lists specification times, amendments and the

limits of result knowledge, including the later accesses to the assessment partition for corrected intervals.

Marker selection followed a development screen of 2,410,564 marker-endpoint tests. The primary pair ranked first in the confirmation plan made before the partition was opened; a comparable ranking of both selected pairs within their endpoint families is not documented (Section S1c). The representative marker was selected for LVEF and the beat median for cardiomyopathy codes. The model bootstrap does not capture this selection uncertainty, and repeated internal analyses do not replace independent confirmation.

Participants and references

Adults aged at least 18 years contributed one ECG per endpoint: the first technically successful ECG with an eligible known reference, ordered by acquisition time with a deterministic tie-break. Selection used reference availability, not its value or the marker value. Models used complete cases, without outcome-dependent outlier removal.

LVEF below 40% was compared with LVEF of 40% or more, without rounding; the comparison is named “LVEF $\geq$ 40%”, not “preserved”. The 2021 ESC guideline defines reduced ejection fraction as LVEF of 40% or less [25], unchanged in the 2023 update [26]; our contrast places persons with exactly 40% (276 of 8,371 in the assessment set; Section S15f) among the controls, and the analysis with LVEF $\leq$ 40% (OR 2.29, 1,076 cases) corresponds to the guideline definition. The structured measurement was chosen by priority (three-dimensional, biplane, generic; measurement methods as described in [27]). A matched echo with a timestamp within ± 30 days of the ECG (selected from a ± 90 -day search), a finite physiological LVEF and no intermethod conflict above 10 percentage points were required. A sensitivity analysis restricted the interval to 29 days, avoiding the boundary region affected by hour-scale timestamp offsets (Section S1a). Reader characteristics and validation against reports are unavailable.

Coded cardiomyopathy used ICD-9 prefix 425 or ICD-10 prefixes I42 or I43 in the index hospital encounter. Controls had an assigned encounter without a matching code. A missing diagnosis list counted as absence of the code; this applied to 35 of 70,892 development controls (0.05%; Section S23). Codes refer to the whole encounter, may postdate the ECG and do not verify absence.

Measurements and comparators

LoopLens (version 10.1; fixed release of 7 October 2026), an in-house Python pipeline, reconstructed XYZ with the Kors transformation [28] at 500 Hz. The primary discrete Fréchet distance compared representative QRS and T loops, each sampled at 64 arc-length points and then centered and divided by its own root-mean-square radius, with a minimax recursion [29] and without rotation or mirror alignment. The representative loops come from a template cycle, the median of time-aligned accepted beats. The operational definitions of the marker, the comparator angle and the covariates, including filters, thresholds, template construction, reconstruction and equations, are given in Supplement Section S0a.

For cardiomyopathy, the marker was the median normalized beatwise distance in the primary morphology cluster (at least three eligible beats). The distance is dimensionless.

The comparator was the origin-corrected integral angle between QRS and JT vectors, defined here and entered linearly in degrees. The basic model (B3) contained a fixed natural cubic age basis, recorded sex, heart rate, QRS duration, Fridericia-corrected QT and the angle. A stronger model (B4) added Cornell product [30], Sokolow-Lyon voltage [31], QRS area and peak magnitude, and sine and cosine of the frontal QRS axis. Kors reconstruction (compared with the inverse Dower matrix [32] in the Supplement) does not establish interchangeability with Frank recordings [7, 28, 33]; the agreement of LoopLens with BRAVEHEART [34] has not been tested.

Before assessment, the measurement-quality criterion required a quantitative bound on wave-boundary error in the outcome population. No bound was available, and the author waived the requirement in exchange for a narrower claim. Five limits apply: (1) stored or reproduced Kors-derived values of the fixed release at 500 Hz, hospital complete cases only; (2) no transformation, device or era invariance, no secure pacing exclusion; (3) no individual diagnosis or change claim; (4) B3 controls the scalar integral angle, and amplitude, hypertrophy and axis independence extend only to what B4 tested; (5) unequal repeatability (angle intraclass correlation 0.62 against 0.83–0.84 for the distance) permits residual confounding by the noisier angle.

Statistical analysis

ORs used endpoint-specific development SDs fixed from combined training and validation (LVEF marker 0.765, cardiomyopathy marker 0.659 in dimensionless units). Logistic models used a ridge term of 10^{-6} and person-cluster robust Wald intervals with small-sample correction. No sample-size calculation was done. The primary hypothesis required a positive coefficient with two-sided $P \leq 0.05$; the secondary family used Holm correction, and the two unavailable hour-dependent tests kept $P = 1$.

AUC compared each basis with that basis plus the marker on the same observations, with paired person-bootstrap refits at fixed scaling, 1,000 planned replicates and at least 950 valid fits for percentile intervals. Centering and scaling constants, the age basis, coefficients and thresholds of the models tabulated in Section S25 (basic model with and without marker, development-fixed models, verification model; for the stronger model B4 only centering and SD) are given there.

The bootstrap of the original run had a construction error (Section S18): it required at least 100 distinct persons per class in every replicate instead of in the sample, which rejects typical draws with about 150 cases, and the fitting routine stopped on a criterion per unit of weight whereas acceptance used one per distinct person. Post hoc, with eligibility checked in the sample, 1,000 replicates and a consistent stopping rule, all corrected rows had 1,000 of 1,000 valid replicates; point estimates and Wald intervals did not change. Corrected intervals are marked as such.

Scenario analyses after the primary result

All analyses in this section are post hoc. They are summarized in Table 4 and detailed in Sections S9-S10, S13, S17 and S18-S25. Unless stated, they used development data. Outcome data cannot identify the bias parameters, so each analysis states the value of an assumed parameter at which the OR reaches 1.25 or 1.0 (tipping point). Quantitative bias analysis rests on assigned values of bias parameters [35].

Classical error. Regression calibration (RC), simulation-extrapolation (SIMEX) [36, 37] and a known-truth simulation used error variances from 13,029 clinical repeat pairs (8,966 persons), ECGs within two hours of each other; the intraclass correlation (ICC) is the two-way absolute-agreement ICC(2,1). Repeats include physiological change and do not identify technical error alone [38]; two segments of one recording agree within ± 14 degrees (95% limits) for the QRS-T angle [39], far below the SD of 44 degrees of our clinical repeat differences; shifts of QRS and T boundaries change the measured angle [40]. A set of assumed reliabilities and an error correlation is called an anchor. Three main anchors take all repeat pairs, repeat pairs by LVEF group, or the ICCs of all pairs; a fourth uses the 289 time-aligned repeat pairs; an assumed technical-agreement anchor takes the agreement between the 500 Hz and 250 Hz versions of the same recording as technical error (angle reliability 0.75, marker reliability 0.98). A bivariate RC allowed classical error in both angle and marker with an error correlation. Its first grid (angle reliability 0.59–0.70, marker reliability 0.80–0.98, error correlation 0.26–0.33) spanned the repeat-based estimates and is an assumed scenario set, not a measured bound. The grid was then extended (angle reliability 0.40–0.80, marker reliability 0.60–1.00, error correlation – 0.4 to 0.8) and the anchors re-estimated in 1,000 person-bootstrap replicates. A setting was admissible only if the latent covariance was positive definite.

Differential and persistent error. We compared repeat error by LVEF group and QRS amplitude and modeled (a) a systematic shift of the measured marker or angle in cases, in SD of controls (about 0.93 development SD), (b) a larger error variance in cases (factor κ) and (c) a persistent error share π that repeat differences cannot see. The one quantified mechanism, gain sensitivity of the delineation, served as an anchor. These are scenarios, not a validated error model.

Missing marker values. Incomplete data were more frequent in cases than in controls (development 15.4% and 18.3% of cases against 10.8% and 11.4% of controls had incomplete basic-model data; assessment 17.44% against 9.92%), so missingness may depend on the outcome. We used delta-adjusted multiple imputation (100 imputations), applying the delta-adjustment principle of [41] to missing marker values, and shifted only the imputed marker values by δ development SD, for cases, controls or both in opposite directions, with larger spread, with age- or QRS-dependent shifts, for the angle alone, and for parts of the missing pattern.

Echo verification. LVEF is unobserved in unverified persons, so selection can be examined only through its consequences; selection that depends on marker and outcome biases their association [42]. The selection factor B is the ratio, per marker SD, of the case-to-control ratio λ of verification probabilities. The true OR equals the observed OR divided by B if $\ln \lambda$ is linear in the marker with a common slope $\ln B$

and the covariate function of the main model absorbs the covariate-dependent level of λ (own derivation, Section S22). B is not identifiable. On the index population of the main analysis (first verified ECG for verified persons, first non-excluded ECG otherwise) we calculated the B at which the OR reaches 1.25 or 1.0 and, by linear programming from a fitted verification model, the largest B for which selection probabilities stay between 0 and 1 (positivity [43]) under stated assumptions about the level of λ . The calculation is conditional per person, because odds ratios are non-collapsible [44].

Reference and confounders. We varied the ECG-to-echo interval and its direction; excluded ECGs with a pacing indicator (machine-report text, device codes or procedures) using a rule fixed before inspection; separated controls without any diagnosis record; added orientation comparators on identical persons; and shifted T ends of cases by up to ± 32 ms with QTc recomputed arithmetically (an approximation, Section S17), in a protocol dated and checksummed before sampling. Medication-order indicators and partial extended adjustment are in Section S13.

Reporting follows STROBE, RECORD and SAGER [45–47] and relevant TRIPOD + AI items [48] (Section S7). STARD was not applied because no diagnostic accuracy claim is made. Generative AI tools assisted the work as stated in the Declarations.

Results

Primary association and participant flow

The LVEF analysis retained 8,371 persons (800 cases) and the cardiomyopathy analysis 18,170 persons (433 cases). The primary association met its specified rule: Fréchet OR 2.34 per development SD (95% person-cluster Wald CI 2.09–2.62; $P < 0.001$; Table 2). The specified hypotheses on an acute-heart-failure proxy and on elevated NT-proBNP were not estimable because hour-scale alignment of device and record times could not be established (Section S3). Because the development estimates were already known and the sample was large, the rule was not demanding. For coded cardiomyopathy, the beat-median distance gave OR 2.03 (1.79–2.29; Holm-adjusted $P < 0.001$). The cardiomyopathy endpoint is secondary, shares persons with the first and is treated in Sections S3 and S13e.

Table 1
Post hoc case/control characteristics in the assessment complete-case sets

Characteristic	LVEF < 40%, n = 800	LVEF ≥ 40%, n = 7,571	Coded cardiomyopathy, n = 433	No cardiomyopathy code, n = 17,737
Age, years	70.6 ± 14.2	64.6 ± 16.9	65.7 ± 15.2	61.6 ± 18.2
Heart rate, beats/min	88.5 ± 21.4	80.2 ± 21.1	86.1 ± 23.8	80.3 ± 20.2
QRS duration, ms	134.6 ± 36.1	112.6 ± 21.5	135.1 ± 36.9	112.2 ± 20.7
Fridericia QTc, ms	450.9 ± 43.0	429.6 ± 35.1	454.6 ± 41.8	430.2 ± 35.4
Male, n (%)	557 (69.6)	3,942 (52.1)	288 (66.5)	9,031 (50.9)
Sinus rhythm (machine-report category), n (%)	557 (69.6)	6,322 (83.5)	278 (64.2)	15,627 (88.1)
Left-bundle pattern positive, n (%)	33 (4.1)	73 (1.0)	16 (3.7)	162 (0.9)
Right-bundle pattern positive, n (%)	146 (18.2)	645 (8.5)	66 (15.2)	1,347 (7.6)
Continuous variables are mean ± standard deviation; categorical values are n (%). Descriptive summaries were specified after primary-result inspection. Machine-report rhythm and block patterns are not independently adjudicated diagnoses. Pattern unknowns of the LVEF sets and medians are retained in Section S15a; in the cardiomyopathy sets, negative and unknown patterns are combined as not positive. All model covariates were complete in these sets; pacing, independently adjudicated rhythm, acquisition era, care setting and comorbidity summaries were unavailable. Flow is in Fig. 1. LVEF, left ventricular ejection fraction; QRS, QRS duration; QTc, corrected QT interval.				

Table 2
Specified primary and secondary association results

Reference and marker	Persons / cases / controls	OR per development SD (95% Wald CI)	Raw p	Holm p	Apparent Δ AUC (95% bootstrap CI)
LVEF < 40%; normalized Fréchet	8,371 / 800 / 7,571	2.34 (2.09–2.62)	< 0.001	Not applicable, primary	0.0326 (0.0247–0.0414)
Acute-heart-failure proxy; mirror symmetry	NE	NE	NE	NE	NE
Elevated NT-proBNP; normalized beat-median Fréchet	NE	NE	NE	NE	NE
Coded cardiomyopathy; normalized beat-median Fréchet	18,170 / 433 / 17,737	2.03 (1.79–2.29)	< 0.001	< 0.001	0.0262 (0.0169–0.0367)
The LVEF and cardiomyopathy hypotheses are not independent studies. Time-based stops are explained in Supplement Section S3. ORs use person-cluster robust inference with small-sample correction; AUC intervals use paired person-bootstrap refits. The secondary family retains all three hypotheses. The bootstrap of the original run had a construction error that did not affect these two rows (1,000 valid replicates each); corrected intervals for other rows are in Tables S91 and S92. NE, not estimable; OR, odds ratio; CI, confidence interval; SD, standard deviation; AUC, area under the receiver-operating-characteristic curve; ΔAUC, increment in AUC; LVEF, left ventricular ejection fraction; NT-proBNP, N-terminal pro-B-type natriuretic peptide. The beat median requires at least three eligible beats; mirror symmetry was not estimable in the acute reference window.					

Post hoc flow counts showed that 17.44% of LVEF cases and 9.92% of controls were lost before complete basic-model and marker data were available (cardiomyopathy 15.92% and 8.11%); the Fréchet distance was absent in 96 of 969 cases and 372 of 8,405 controls at the reference stage, overlapping with the incomplete covariate data (Fig. 1, Table 1, Sections S15a, S15e). The endpoints overlap in 7,052 persons of the analysis sets.

Adjustment, comparators and reference

On identical complete sets the OR was 2.49 (2.21–2.82) with the basic model and 2.51 (2.20–2.87) with the stronger voltage-and-axis model (Table 3, Fig. 2); LVEF $\leq$ 40% gave 2.29 (2.06–2.53; 1,076 cases). On identical development persons (33,884 persons, 3,508 cases), adding peak and area angles with a spline, area-vector axes, the ventricular gradient, TCRT with T-morphology dispersion, or planarity moved the OR of the distance from 2.46 (2.31–2.61) to between 2.28 and 2.60, and to 2.39 (2.23–2.60) with all 24 added terms (Section S24). The apparent incremental AUC of the distance fell from 0.033 to 0.014 (0.011–0.016) with all terms (out-of-fold 0.033 to 0.015), so most of it is shared with orientation descriptors; TCRT as implemented here and planarity left it nearly unchanged. In the assessment partition, the basic model's AUC rose from 0.8024 to 0.8351 (increment 0.0326, 0.0247–0.0414).

Table 3
Additional voltage and axis adjustment on identical complete cases

Reference / adjustment	Persons / cases / controls	OR per development SD (95% Wald CI)	Apparent Δ AUC (95% bootstrap CI)
LVEF < 40%, basic model on the stronger-model set	7,840 / 636 / 7,204	2.49 (2.21–2.82)	0.0389 (0.0294–0.0492)
LVEF < 40%, stronger model	7,840 / 636 / 7,204	2.51 (2.20–2.87)	0.0334 (0.0238–0.0428)
Coded cardiomyopathy, basic model on the stronger-model set	17,345 / 334 / 17,011	2.10 (1.83–2.41)	0.0325 (0.0199–0.0463)
Coded cardiomyopathy, stronger model	17,345 / 334 / 17,011	1.98 (1.69–2.33)	0.0219 (0.0114–0.0345)
All entries are descriptive sensitivities. Each Δ AUC compares its named basis with that basis plus the marker. Differences across the two rows are not an independently tested superiority comparison. LVEF, left ventricular ejection fraction; OR, odds ratio; SD, standard deviation; Δ AUC, increment in the area under the receiver-operating-characteristic curve.			

The OR was similar for echocardiograms within 1, 7, 14 and 30 days of the ECG: 2.35, 2.43, 2.45 and 2.48 in training, 2.48, 2.39, 2.42 and 2.39 in validation, and 2.36, 2.42, 2.39 and 2.34 in the assessment partition (7 days: 2.42, 2.14–2.74, 6,337 persons; Section S23). About 13% of echocardiograms preceded the ECG; echo-after-ECG pairs gave the same OR (2.31–2.49), echo-before pairs were too few for stable estimates in validation and assessment. Excluding the 2.6% to 2.7% of ECGs with a machine-report pacing hint changed the OR to 2.53 (training) and 2.44 (validation); also excluding stays with device indicators gave 2.61 and 2.49, and keeping only recordings without any indication 2.52 and 2.32. Half of all recordings had no reliable pacing information and the assessment partition has no pacing variable, so secure pacing exclusion remains unavailable. Among recordings with a report hint, the OR was lower (1.39, 1.08–1.79, training). Of the LVEF values, 97% were generic measurements; the rest were too few to analyze separately.

Classical error and the two null simulations

With the Fréchet distance in the model, development angle ORs (per development SD of the angle, 46.8 degrees) fell to 1.009 (0.946–1.076) and 0.980 (0.872–1.102), from 1.793 (1.699–1.892) and 1.695 (1.542–1.863) without it, while the marker ORs were 2.48 and 2.39. Conditional angle reliability fell to 0.07 (marker-angle Spearman correlation 0.75), and removing relative loop rotation reduced the AUC of the marker for reduced LVEF from 0.805 to 0.559 in a separate internal-validation sample within development (2,609 persons, 624 cases; Sections S10, S13). The marker therefore contains substantial angle information.

Two known-truth simulations of classical angle error gave different answers (Fig. 5). The early null (marker without effect, independent angle error) gave a geometric-mean OR of 1.247 (simulation

quantiles 1.096–1.395) and matched the observed structure poorly: conditional angle reliability 0.499 against about 0.07 observed, marker-angle Spearman correlation 0.430 against 0.75 (Section S13j). A later simulation at the angle-only tipping point (true marker OR 1.0, angle reliability 0.57, error-free marker, strong true angle effect) gave an uncorrected marker OR of 2.49 (2.31–2.67) and an angle OR of 1.005, close to the observed 2.48 and 1.009 (Section S13k, Table S89). Neither simulation represents the full empirical structure, so 1.25 is not an upper bound for artifact, and the outcome data alone cannot exclude the second scenario. It requires a latent angle-marker correlation of 0.97–0.99 (0.87, 0.86–0.88, from cross-repeat correlations), an error-free marker (repeat-based 0.83) and an angle reliability of 0.57 (repeat-based 0.60, interval 0.58–0.61); whether it is excluded depends on whether repeat differences measure the error of interest and on nondifferential error.

In the assessment partition, angle-only RC gave a marker OR of 2.09 (95% bootstrap CI 0.42–9.55), joint RC 8.89 (3.29–48.49, unstable), and the cardiomyopathy point estimate reversed to 0.56 (Fig. 3, Section S15c). SIMEX estimates were 2.33–2.49 (angle-only) and 2.63–2.87 (joint), and were weakly informative because the conditional angle coefficient is near zero. In development data, RC intervals included 1 (Fig. 4).

Bivariate correction and its assumed set. In the first post hoc grid, the corrected OR had a minimum of 2.04 (training) and 2.26 (validation), and the 2.5th percentile of the per-replicate minimum over the report points was 1.80 and 1.53; these limits belong to the report points, not to the fine grid (Section S13k). At the repeat-based anchors the corrected latent covariance was close to singular (partial latent correlation 0.94 to 0.97), and the corrected ORs of 8.7 to 9.3 (25.8 to 27.5 with LVEF-specific error) are singularity effects, not effect estimates. Across 1,000 bootstrap replicates the corrected OR fell below 1.25 in none at the two repeat-based main anchors (0.4% at the anchor from time-aligned pairs), and in the extended grid only at angle reliabilities of 0.54 or less (training) and 0.52 or less (validation), or at 0.57 with an error-free marker, just above the admissibility limit of 0.566 (Section S19).

Differential and persistent error

Repeat error of the angle did not differ by LVEF (SD of differences 44.0 against 42.6 degrees; ratio 1.03, 0.98–1.09; 9,423 pairs from 6,089 persons), whereas the Fréchet error was smaller at low LVEF (ratio 0.93, 0.86–1.00) and the error correlation lower (0.19 against 0.27). Angle error was modestly higher at the lowest QRS amplitude quintile (ratio to the highest 1.07, 1.01–1.13; Section S20). Similar repeat differences do not exclude a persistent bias that depends on LVEF, because repeats cannot see persistent error and include physiological change.

The scenario models locate the vulnerable point (Section S19). Without error correction, a systematic over-measurement of the marker in cases of 0.35 to 0.38 SD of controls (29% to 30% of the observed case-control difference of 1.24 to 1.28 SD) lowered the OR to 1.25, and 0.47 to 0.49 to 1.0. After correction for classical angle error, whose corrected estimate is itself unstable (about 9), the tolerance fell to 0.15 to 0.20 control SD (12% to 16% of the difference) for 1.25 at three main anchors and 0.16 to 0.22 for 1.0; the closeness of these two values shows the sensitivity of the corrected frame. At an

assumed technical-agreement anchor the figures were about 0.25 (20%) and 0.32 to 0.33. A shift of the angle alone did not push the corrected OR below 1.25 up to 2.0 control SD. A persistent error share of only 0.02 to 0.07, or an error variance 1.5 to 1.6 times larger in cases, made the classical correction inadmissible before it reached 1.25: neither a tipping point nor a reassurance, because repeat differences cannot say whether persistent error or physiological change explains it. The one quantified mechanism, gain sensitivity of the delineation, shifted the mean marker by less than 0.005 SD (gain changes of 1% to 5%) and mainly added noise to the angle; it says nothing about devices or LVEF-related delineation error.

Missing marker values and echo selection

In delta-adjusted imputation (6% to 7% of Fréchet values imputed), the pooled OR was 2.44 (training) and 2.38 (validation) under missingness at random; shifts are in development SD. The OR fell to 1.25 when imputed values of cases and controls were shifted in opposite directions by 1.80 and 1.64 SD (1.4 to 1.5 times the observed case-control separation of 1.16 to 1.18 SD) or when only imputed control values were raised by 3.16 and 3.03 SD; for cases shifted down alone it was not reached within 8 SD in training (3.40 SD in validation). Doubling the spread of the imputed values lowered the pooled OR to 1.8 without any shift and moved the opposite-direction tipping point to 1.39 (1.19–1.60) and 1.26 (0.92–1.62) SD, 1.1 to 1.2 times the separation; other variants did not bring the tipping points closer (Section S21). Reducing the OR to 1.0 needed opposite shifts of 2.93 and 2.60 SD and 6.52 and 6.27 SD for controls alone. These are named departures from missingness at random, not estimates, and the observed separation does not bound the values of unmeasured markers.

Outcome-dependent verification can be examined only through its consequences for the observed OR. On the index population of the main analysis, 28.7% (training) and 28.6% (validation) of persons with complete design were verified, and the marker predicted verification (OR 1.53 and 1.57 per SD); the 18% in the weighted verification analysis refers to the first technically successful ECG (128,912 persons; 18.7%, or 18.0% with complete model data), a different population (Section S22). Reducing the OR to 1.25 required a selection factor B of 1.98 (1.86–2.11) in training and 1.91 (1.71–2.14) in validation per marker SD, and to 1.0 required 2.48 and 2.39. The largest B with selection probabilities not exceeding 1 was 1.30 and 1.23 if the level of the ratio did not depend on covariates (1.49 and 1.43 over the central 90% of marker values), at most 1.16 if cases were never verified less often than controls, and 1.78 and 1.76 if the level varied linearly with covariates over all persons; all stay below the tipping values. With log-linearity required only over the central 90% of marker values and a linear level, the maximum was 2.09 and 2.80, which allows an OR of 1.25 in both folds and 1 in validation only. Certain restrictive selection models are therefore incompatible with a complete explanation, but selection could not be excluded in general; these are model-dependent feasibility limits, not empirically identified, because B is not identified. The delayed verification analysis (4,844 persons, 431 cases) gave an unweighted OR of 2.65 (conditional Wald CI 2.28–3.07) and 2.97 (2.52–3.50) with weights; only 777 of 1,000 bootstrap replicates were valid because the fitting routine stopped on a different criterion than the acceptance rule (Section S18), so no bootstrap interval exists.

T-end sensitivity, strata and other post hoc analyses

Shifting all beat T ends in cases (1,500 cases and 1,500 controls, development) lowered the OR from 2.23 to 1.61 (95% bootstrap CI 1.40–1.91) at + 32 ms and 2.01 (1.77–2.33) at -32 ms when QTc was recomputed arithmetically (approximate: the reconstruction also changes QRS duration slightly in a minority of recordings; Section S17); with QTc held at its measured value the figures were 1.51 and 2.32. OR 1.25 was not reached in either direction within 32 ms. For orientation, the 32-ms lengthening lowered the log OR by 0.33, 56% of the 0.58 needed to reach 1.25; a T-end perturbation and a marker shift in control SD are different quantities and cannot be converted directly. Fifty of the 1,500 cases (3.3%), with wider QRS (153 against 137 ms), dropped out of the common sample; including them changed the OR by at most 0.016. The earlier LUDB-predicted T-end correction [49, 50] is inconclusive (Section S14).

Post hoc, the OR was 2.32 (bootstrap 95% CI 2.11–2.59) after multiple imputation of covariates, 2.34 (bootstrap 1.99–2.83; 457 cases) for $QRS \geq 120$ ms, 2.39 (bootstrap 2.05–2.81, Wald 2.04–2.81, as drawn in Fig. 2; 329 cases) for $QRS < 120$ ms without a block pattern and 2.44 (bootstrap 1.91–3.39; 146 cases) in right-bundle pattern. The left-bundle stratum of the assessment partition (106 persons, 33 cases) was not estimable. In development data with Firth penalization, persons with a classical left-bundle pattern (training 318, validation 103; 131 and 32 with LVEF below 40%) showed lower point estimates (OR 1.45, profile-likelihood interval 0.96–2.20, and 1.40, 0.68–2.98; pooled 1.47, 1.02–2.11, an exploratory association) than persons with a negative pattern (2.55 and 2.48); no interaction test was done, and the pattern is a rule-based surrogate, undetermined in 3.2%. Excluding it left the main estimate at 2.48 and 2.42, and a QRS-duration spline changed it by at most 1.2% (Section S26). Seven-day medication-order indicators gave 2.47 (2.31–2.64) and 2.38 (2.14–2.64) in training and validation. Figure 2 and Table 4 give the overview; these internal results do not assess transported clinical benefit.

Table 4
Overview of the analyses after the primary result (all post hoc; data and reading)

Question	Analysis	Data	Result for the OR (reference 2.34 to 2.48)	Reading
Orientation comparators	Five blocks and all 24 terms on identical persons (Section S24)	Development, 33,884 persons	2.28–2.60; 2.39 with all	Association stable; most incremental AUC shared
Echo timing and direction	Windows 1–30 days, before/after ECG (Section S23)	Development and assessment	2.34–2.48	Stable; echo-before pairs too few in validation and assessment
Pacing	Exclusion by report text and device indicators (Section S23)	Development	2.32–2.61	Not lower; about half of recordings uninformative; spikes not assessed
Code-negative controls	Controls without any diagnosis record (Section S23)	Development	2.03 unchanged	0.05% of controls affected
Classical error, early null	Known-truth simulation (Sections S13b, S13j)	Simulated	1.25	Matched the observed structure poorly; not an artifact bound
Classical error, tipping-point simulation	True OR 1.0 at angle reliability 0.57 (Section S13k)	Simulated	2.49	Reproduces the uncorrected pattern; argued against only by repeat-based reliabilities and nondifferential error
Bivariate correction, extended grid	Anchors with population intervals (Section S19)	Development	Below 1.25 in 0 of 1,000 at the main anchors; tips at angle reliability 0.54 or less	Corrected ORs 8.7–9.3 are singularity effects
Differential error in cases	Marker shift, case error variance, persistent share (Section S19)	Development	1.25 at 0.15–0.20 control SD (about 0.25 at an assumed technical anchor); 1.0 at 0.16–0.22	Most vulnerable point
LVEF, left ventricular ejection fraction; OR, odds ratio per development standard deviation; SD, standard deviation; B, selection factor; AUC, area under the receiver-operating-characteristic curve; ICC, intraclass correlation. Source sections of the Supplement are given in brackets; the numbers in the Result column come from different data sets and are not directly comparable.				

Question	Analysis	Data	Result for the OR (reference 2.34 to 2.48)	Reading
Missing marker values	Delta-adjusted imputation (Section S21)	Development	1.25 at 1.6–1.8 development SD opposite shift (1.3–1.4 with doubled spread); 1.0 at 2.6–2.9	1.1–1.5 times the observed separation; not bounded by it
Echo selection	Selection factor B, feasible interval by linear programming (Section S22)	Development, index population	1.25 at B = 1.9–2.0 per marker SD; 1.0 at 2.4–2.5	Model-dependent feasibility limits: 1.2–1.5 (covariate-independent level; at most 1.16 if cases are never verified less often), 1.8 (linear level, all persons), 2.1–2.8 (central 90% of marker values); restrictive models incompatible with a complete explanation, selection not excluded in general
Left-bundle pattern, QRS spline (Section S26)	Firth stratum; QRS duration as spline	Development	1.45 and 1.40 within the stratum (intervals include 1); 2.45–2.51 with spline	Lower point estimates within block; exploratory pooled association; not explained by nonlinear QRS
Negative-control diagnoses (Section S27)	Eight diagnoses, B3 model	Development, 52% with index encounter	0.87–1.10 pooled (LVEF 2.35 in the same persons); retinopathy 1.27	No comparable positive associations with the selected coded diagnoses; general frailty confounding remains unresolved; retinopathy not a clean control
T-end shift in cases	Constant shift up to ± 32 ms, QTc recomputed arithmetically (Section S17)	Development	1.61 (+ 32 ms), 2.01 (-32 ms)	1.25 not reached; approximate coherence; not a measured error
Corrected bootstrap	Eligibility, replicate number, stopping rule (Section S18)	Assessment rows re-read	1,000 of 1,000 valid in all corrected rows	Point estimates and Wald intervals unchanged
LVEF, left ventricular ejection fraction; OR, odds ratio per development standard deviation; SD, standard deviation; B, selection factor; AUC, area under the receiver-operating-characteristic curve; ICC, intraclass correlation. Source sections of the Supplement are given in brackets; the numbers in the Result column come from different data sets and are not directly comparable.				

Question	Analysis	Data	Result for the OR (reference 2.34 to 2.48)	Reading
External replication	Second cohort for the cardiomyopathy code; EchoNext	Not available	Not available	Outstanding
LVEF, left ventricular ejection fraction; OR, odds ratio per development standard deviation; SD, standard deviation; B, selection factor; AUC, area under the receiver-operating-characteristic curve; ICC, intraclass correlation. Source sections of the Supplement are given in brackets; the numbers in the Result column come from different data sets and are not directly comparable.				

In the assessment partition, the OR was similar in women (2.37, 1.97–2.86) and men (2.34, 2.03–2.70), larger below 65 years (2.67) than from 65 years (2.20), and 2.50 in report-defined sinus rhythm (Fig. 2). With LVEF as a continuous reference in development data, the marker was associated with 3.7 percentage points lower LVEF per SD in training (3.5 in validation), and with an ordinal reference (< 40%, 40–49%, ≥ 50%) with an OR of 0.42 for a higher category in training (0.44 in validation; Section S13i).

Negative-control diagnoses

As post hoc negative controls in development data, we modeled eight diagnoses judged unrelated to left ventricular systolic function (list fixed before the data were examined; codes in the index hospital encounter) with the B3 covariates and the marker (Section S27). Only 52% of the main analysis set had an index encounter; in them the OR for LVEF below 40% was 2.40 (2.22–2.60) in training and 2.23 (1.93–2.57) in validation. With at least 100 cases in a fold, the ORs were 0.89 (0.77–1.03) and 0.90 (0.69–1.18) for glaucoma and 1.07 (0.97–1.19) and 0.90 (0.76–1.08) for prostatic hyperplasia in men; in training alone, 0.88 to 1.06 for cataract, hearing loss, diverticulosis, hemorrhoids and osteoarthritis (upper limits 1.06 to 1.40). Pooled over both folds (a row added after the case counts were seen), no negative control had an interval above 1 (point estimates 0.87 to 1.10). Coded retinopathy, added as a comparator, gave 1.29 (1.09–1.52) and 1.23 (0.96–1.58).

Discussion

The Fréchet distance between QRS and T loops tracks low ejection fraction beyond the measured QRS-T angle (OR 2.34 per development SD), across orientation comparators, echo timing and exclusion of pacing indicators (development data; half of the recordings could not be classified). The data do not show that the distance is a signal independent of true orientation: the angle coefficient fell to about 1 when the marker entered the model. The tipping values for attenuation to 1.25 and removal are given in the Results (Section S19); the data cannot test the marker over-measurement or the imputation shift, certain restrictive selection models are incompatible with a selection factor of about 2, but selection could not be excluded in general, and the scenarios do not bound combinations.

The early classical-error null (apparent OR 1.25) had been read as the main argument against an angle-proxy explanation, but a true OR of 1.0 at angle reliability 0.57 reproduces the observed uncorrected pattern. The case against a pure proxy then rests on repeat-based reliabilities and an error correlation that assume classical, nondifferential error changing between recordings, which outcome data cannot check. Persistent person- or device-level error, which repeat differences cannot detect, need not be large to matter: a persistent share of a few hundredths of the error variance already makes the correction inadmissible. We therefore describe the association as remaining under specific assumptions (repeat-based reliabilities and error correlation, nondifferential error, a linear angle effect), not as robust to measurement error in general; under other classical-error scenarios the observed pattern appears without an independent marker effect.

Error correction was method dependent: SIMEX estimates stayed positive, RC intervals were wide or unavailable, and joint RC was unstable near a singular latent covariance. A retained SIMEX estimate is therefore not evidence of an unattenuated association, and no simulation validates the real-world error model. Stronger angle, axis, gradient, TCRT and planarity comparators did not change the OR [4, 7, 9–14], but they shared most of the incremental AUC, which we do not offer as evidence of added information.

What the distance reflects is not established. It grows when the normalized QRS and T loops differ in shape and orientation, which can follow from abnormal activation (conduction disease), from scar or dilatation changing the loop geometry and from heterogeneous repolarization; we did not test these mechanisms and make no causal claim, and the distance correlates with the angle (Spearman 0.75). The reference is a functional measurement and the codes are diagnoses, not causes of dysfunction. AI-ECG studies show that internal assessment, external evaluation and implementation trials answer different questions [1, 51, 52]. QRS area is associated with outcomes in treated CRT cohorts [33, 53, 54] and preserves amplitude information that our normalization removes; none of this provides a selection threshold for this marker [2].

Conduction disease and rhythm are plausible alternative explanations, because wide QRS, bundle-branch block and ventricular pacing are common in reduced LVEF. The OR was similar with QRS < 120 ms and no block pattern (2.39, bootstrap 2.05–2.81; 329 cases), in report-defined sinus rhythm (2.50, 2.19–2.86) and in right-bundle pattern (2.44), which argues against a pure bundle-branch effect; a narrow QRS largely excludes ventricular but not atrial pacing. In development data the point estimates within left-bundle pattern were lower, with intervals including 1 in each fold and an exploratory pooled association; without an interaction test we claim no effect modification (32 events in validation; post hoc, rule-based surrogate). Neither a nonlinear QRS-duration term nor bundle-branch indicators removed the association elsewhere.

The negative-control diagnoses showed no comparable positive associations with the selected coded diagnoses: no interval excluded 1 from below, and the upper limits stayed at or below 1.40 against an LVEF OR of 2.4. General frailty confounding remains unresolved, because the intervals limit the individual diagnosis associations and not a shared bias component; the controls also depend on incomplete

inpatient coding of chronic conditions and covered only the 52% of persons with an index encounter. Retinopathy, mostly diabetic, stayed associated (OR about 1.3) and is not a clean negative control. Whether part of the association reflects general disease burden, conduction abnormalities or healthcare selection therefore remains open.

Limitations

This is a single-center, person-disjoint internal assessment in a partition previously used for related hypotheses; it is not an untouched or external replication. Selection after a large development screen allows for a winner's curse; similar development and assessment effect sizes do not show that selection bias is absent. All scenario analyses used development data, except the echo-interval reproduction and the corrected bootstrap intervals, which re-read assessment rows (Section S1). They are post hoc and do not validate any error, missingness or selection model.

Missing marker values, complete-case selection and echo verification are informative, and the weighting of the verification analysis assumes missingness at random. The scenarios give the size of single departures that would attenuate or remove the association; they do not bound the real departure or combinations of departures. The selection-factor check depends on the assumed level and shape of the verification ratio, on the calibration of the verification model in its upper range (deviation up to 0.025) and on the non-collapsible nature of the ORs. Differential marker error in cases is the most vulnerable point; repeat differences cannot reveal persistent error, and a uniform gain change of 5% to 10% altered the measured angle by root-mean-square 12.6° to 16.6° but the normalized distance by only 0.056 to 0.059, so device and gain differences between hospitals may behave differently. The LUDB files [49] are scaled to a 1 mV range per lead (author's inspection), so the boundary errors derived from them refer to that amplitude scale.

LVEF values are routine structured measurements, 97% of them generic, without adjudication or validation against reports, and the broad echo window permits changes in patient state. A blinded review of references and ECGs was not possible; it is a priority for the next step, as are a reliable pacing classification and independent replication. Further internal scenarios do not replace them. QRS adjustment does not remove secondary repolarization changes, and the model was not adjusted for antiarrhythmic, digoxin or QT-prolonging drugs beyond a seven-day order-start sensitivity. Electrolytes, renal function, body mass index, care setting and comorbidity remain potential confounders. Codes may misclassify disease and postdate the ECG. Prevalence, predictive values and individual risk cannot be inferred.

The results hold for MIMIC-IV-ECG from one hospital and for the LoopLens release named above; LUDB entered only for boundary errors. The marker was computed by a fixed software release (Supplement Section S0a), not validated against annotated wave boundaries in the outcome population; the release used is public (Code availability), and its checksums are reported so that the analysed version can be identified. Corrected bootstrap intervals and several scenario analyses were added after the first version of the analysis and carry the label post hoc. External replication is outstanding. Access to the Harvard-

Emory ECG database [55] for the cardiomyopathy-code endpoint has been requested and is pending; a replication in EchoNext [56] is not yet possible because its signal scale could not be reconciled (replication would require the provider's normalization constants, which have been requested).

What this study does not show

These findings do not support individual diagnosis, replacement of echocardiography, medication monitoring, prognosis or treatment selection. External evaluation, calibration and a prespecified referral threshold are required before an imaging-triage claim can be assessed. The next question is whether the distance adds reproducible information beyond true orientation in an independent cohort with verified references and explicit handling of selection.

Conclusions

Spatial QRS-T loop coupling retained an internal association with reduced LVEF beyond the measured integral angle (OR near 2.3 to 2.5), stable across orientation comparators, echo timing and exclusion of pacing indicators (development data; half of the recordings could not be classified). The marker carries substantial orientation information, and the early classical-error null that suggested an artifact bound of 1.25 was not a bound. Attenuation to 1.25 would need a differential marker over-measurement in cases of 0.35 to 0.38 control SD without error correction and 0.15 to 0.20 after correction at the three main anchors (about 0.25 at an assumed technical-agreement anchor; removal 0.47 to 0.49 and 0.16 to 0.22), or a shift of missing marker values of 1.4 to 1.5 times the observed separation (removal 2.6 to 2.9 SD). Certain restrictive selection models are incompatible with a complete explanation by selection on the echo reference, but selection could not be excluded in general. These analyses neither confirm nor exclude the departures. External replication, reference validation, reliable pacing classification and clinical utility remain open.

Declarations

Ethical Approval and Consent to participate: This is a secondary analysis of de-identified, publicly available research databases. The Institutional Review Board of Beth Israel Deaconess Medical Center (Boston, MA, USA) approved the sharing of the MIMIC-IV research resource and granted a waiver of informed consent [21]. Access to the credentialed resources MIMIC-IV and MIMIC-IV-ECHO was obtained under the PhysioNet Credentialed Health Data Use Agreement after completion of the required training. MIMIC-IV-ECG and LUDb are open-access PhysioNet resources; for LUDb, the Institutional Review Board of Lobachevsky State University of Nizhny Novgorod approved the original study (decision no. 23, 19 October 2017). The databases were de-identified or anonymized by their providers before release. No patients were contacted, no new data were collected and no attempt was made to re-identify individuals. Consistent with published secondary analyses of these databases, no additional ethics approval or informed consent was required.

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 [20], doi:10.13026/4nqg-sb35, is openly available on PhysioNet [23] under ODbL 1.0. MIMIC-IV version 3.1 [21,22], doi:10.13026/kpb9-mt58, and MIMIC-IV-ECHO version 1.0 [24], doi:10.13026/nrjh-5r77, are available to credentialed users under the PhysioNet Credentialed Health Data License 1.5.0. We used MIMIC-IV-ECHO version 1.0; version 1.0.1 (25 August 2026) removed 15 DICOM studies because of an identifier collision and nine out-of-scope files without naming them or stating whether the tabular files were affected. Our labels derive from the structured measurements rather than from the DICOM study identifiers, and removal of any 15 persons could change the main odds ratio by no more than 0.2 (Section S5). The hosp module supplied the clinical variables and codes; the structured echocardiography measurements supplied LVEF. LUDB version 1.0.1 [49,50], doi:10.13026/eegm-h675, is openly available under ODC-By 1.0 and supplied boundary errors only. The contributing tables and release selection rules are described in Section S5. Derived patient-level data cannot be shared under the data use agreement; aggregate results are provided in the Supplementary Material. A table linking each reported estimate to its aggregate source row is available from the author on request.

Code availability: Analysis scripts and the frozen LoopLens release used for this study (PolyForm Noncommercial 1.0.0) are available at <https://github.com/leoprobst/ecg-auswertung-code> (release v1.0.0), in line with clause 8 of the PhysioNet Credentialed Health Data License 1.5.0; aggregate results are provided in the Supplementary Material. Further development of the software remains private. Centering and scaling constants, coefficients and thresholds of the models tabulated in Supplement Section S25 are given there; the coefficients of the stronger model B4 and of the assessment refits are not tabulated (the refits are reported as ORs only). The analysed release is identified by the SHA-256 checksum d64c3074...8838 of its file list (1,959 files) and by the source hash 9bdad0ce...899d of its 875 Python files (shortened; full values from the author). The marker extraction is identical to the release of 7 October 2026: the code hash recorded by the extraction run (c166bd00...89b9) equals the value recomputed for the release. The analysis chain ran from a runtime snapshot (source hash d80fc3c3...ba89) that includes operational corrections of 11 October 2026 in seven files of later analysis stages; the signal processing, delineation, reconstruction and marker code is identical (Supplement Section S0a).

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 the other open databases used.

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; dated retrospective plans and checksummed specifications are not prospective registration. An earlier version was submitted to medRxiv in October 2026 and was not posted (outside the server's scope); it has not been posted or published elsewhere. A related manuscript by the same author, "Repeat variation and definition dependence of ECG/VCG markers: an exploratory audit of one implementation (LoopLens 10.1)", is in preparation for preprint posting. It supplies measurement definitions and repeatability context from the same research program and does not report the present ventricular association tables.

References

1. Yao X, Rushlow DR, Inselman JW, McCoy RG, Thacher TD, Behnken EM et al (2021) Artificial intelligence-enabled electrocardiograms for identification of patients with low ejection fraction: a pragmatic, randomized clinical trial. *Nat Med* 27:815–819. 10.1038/s41591-021-01335-4
2. Glikson M, Nielsen JC, Kronborg MB, Michowitz Y, Auricchio A, Barbash IM et al (2022) 2021 ESC Guidelines on cardiac pacing and cardiac resynchronization therapy. *Europace* 24:71–164. 10.1093/europace/euab232
3. Voulgari Ch, Tentolouris N, Moyssakis I, Dilaveris P, Gialafos E, Papadogiannis D et al (2006) Spatial QRS-T angle: association with diabetes and left ventricular performance. *Eur J Clin Invest* 36:608–613. 10.1111/j.1365-2362.2006.01697.x
4. Biering-Sørensen T, Kabir M, Waks JW, Thomas J, Post WS, Soliman EZ et al (2018) Global ECG Measures and Cardiac Structure and Function: The ARIC Study (Atherosclerosis Risk in Communities). *Circ Arrhythm Electrophysiol* 11:e005961. 10.1161/CIRCEP.117.005961
5. Cortez D, Schlegel TT, Ackerman MJ, Bos JM (2017) ECG-derived spatial QRS-T angle is strongly associated with hypertrophic cardiomyopathy. *J Electrocardiol* 50:195–202. 10.1016/j.jelectrocard.2016.10.001
6. Gleeson S, Liao YW, Dugo C, Cave A, Zhou L, Ayar Z et al (2017) ECG-derived spatial QRS-T angle is associated with ICD implantation, mortality and heart failure admissions in patients with LV systolic dysfunction. *PLoS ONE* 12:e0171069. 10.1371/journal.pone.0171069
7. Kück K, Isaksen JL, Graff C, Skaaby T, Linneberg A, Hansen T et al (2018) Spatial QRS-T angle variants for prediction of all-cause mortality. *J Electrocardiol* 51:768–775. 10.1016/j.jelectrocard.2018.05.011

8. Voulgari C, Pagoni S, Tesfaye S, Tentolouris N (2013) The spatial QRS-T angle: implications in clinical practice. *Curr Cardiol Rev* 9:197–210. 10.2174/1573403x113099990031
9. Acar B, Yi G, Hnatkova K, Malik M (1999) Spatial, temporal and wavefront direction characteristics of 12-lead T-wave morphology. *Med Biol Eng Comput* 37:574–584. 10.1007/BF02513351
10. Tse G, Gong M, Wong CW, Chan C, Georgopoulos S, Chan YS et al (2018) Total cosine R-to-T for predicting ventricular arrhythmic and mortality outcomes: A systematic review and meta-analysis. *Ann Noninvasive Electrocardiol* 23:e12495. 10.1111/anec.12495
11. Waks JW, Tereshchenko LG (2016) Global electrical heterogeneity: A review of the spatial ventricular gradient. *J Electrocardiol* 49:824–830. 10.1016/j.jelectrocard.2016.07.025
12. Draisma HH, van der SchaliJ MJ, Swenne CA (2006) Elucidation of the spatial ventricular gradient and its link with dispersion of repolarization. *Heart Rhythm* 3:1092–1099. 10.1016/j.hrthm.2006.05.025
13. Hnatkova K, Andršová I, Novotný T, Vanderberk B, Sprenkeler D, Junttila J et al (2024) QRS complex and T wave planarity for the efficacy prediction of automatic implantable defibrillators. *Heart* 110:178–187. 10.1136/heartjnl-2023-322878
14. Kors JA, van Herpen G, van Bommel JH (1999) QT dispersion as an attribute of T-loop morphology. *Circulation* 99:1458–1463. 10.1161/01.cir.99.11.1458
15. Kors JA, de Bruyne MC, Hoes AW, van Herpen G, Hofman A, van Bommel JH et al (1998) T-loop morphology as a marker of cardiac events in the elderly. *J Electrocardiol.* ;31 Suppl:54 – 9. 10.1016/s0022-0736(98)90289-8
16. Hasan MA, Abbott D, Baumert M (2012) Beat-to-beat spatial and temporal analysis for QRS-T morphology. *Annu Int Conf IEEE Eng Med Biol Soc* 2012:4193–4195. 10.1109/EMBC.2012.6346891
17. Pencina MJ, D’Agostino RB, Sr, D’Agostino RB Jr, Vasan RS (2008) Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 27:157–172. 10.1002/sim.2929
18. Cook NR (2007) Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 115:928–935. 10.1161/CIRCULATIONAHA.106.672402
19. Van Calster B, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ et al (2018) Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators. *Eur Urol* 74:796–804. 10.1016/j.eururo.2018.08.038
20. Gow B, Pollard T, Nathanson LA, Johnson A, Moody B, Fernandes C et al (2023) MIMIC-IV-ECG: Diagnostic Electrocardiogram Matched Subset. 10.13026/4nqg-sb35. RRID:SCR_007345
21. 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
22. Johnson A, Bulgarelli L, Pollard T, Gow B, Moody B, Horng S et al MIMIC-IV. Version 3.1. PhysioNet; 2024. RRID:SCR_007345. 10.13026/kpb9-mt58

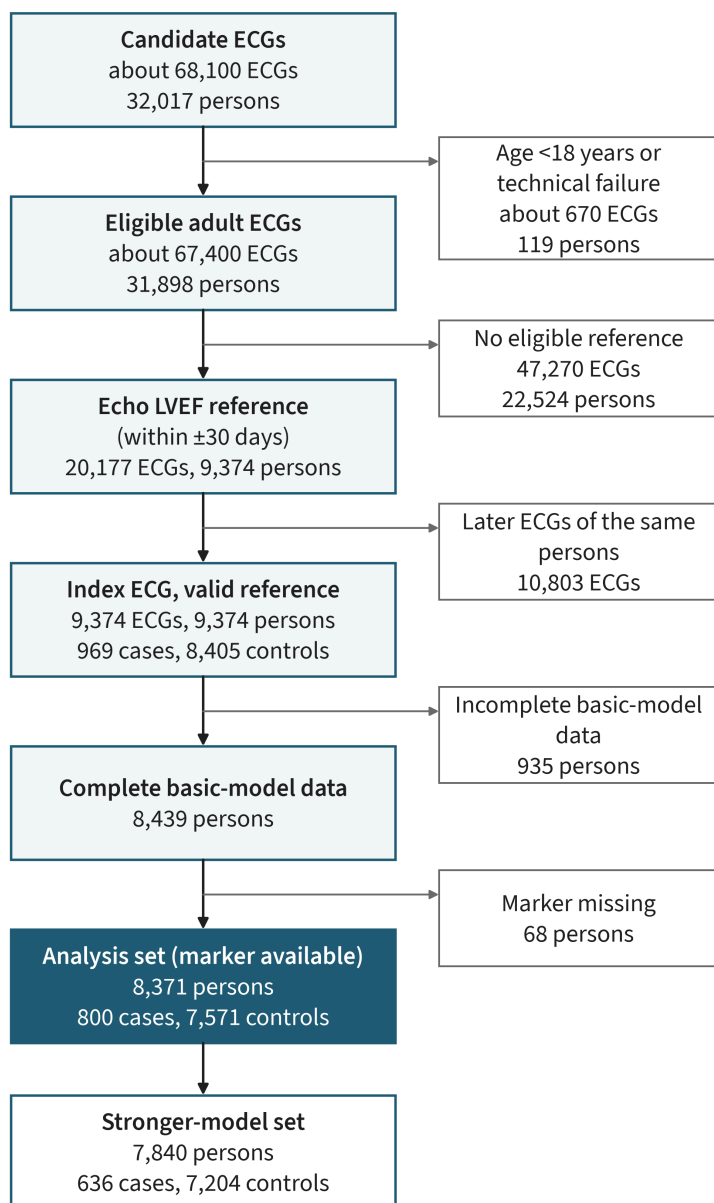
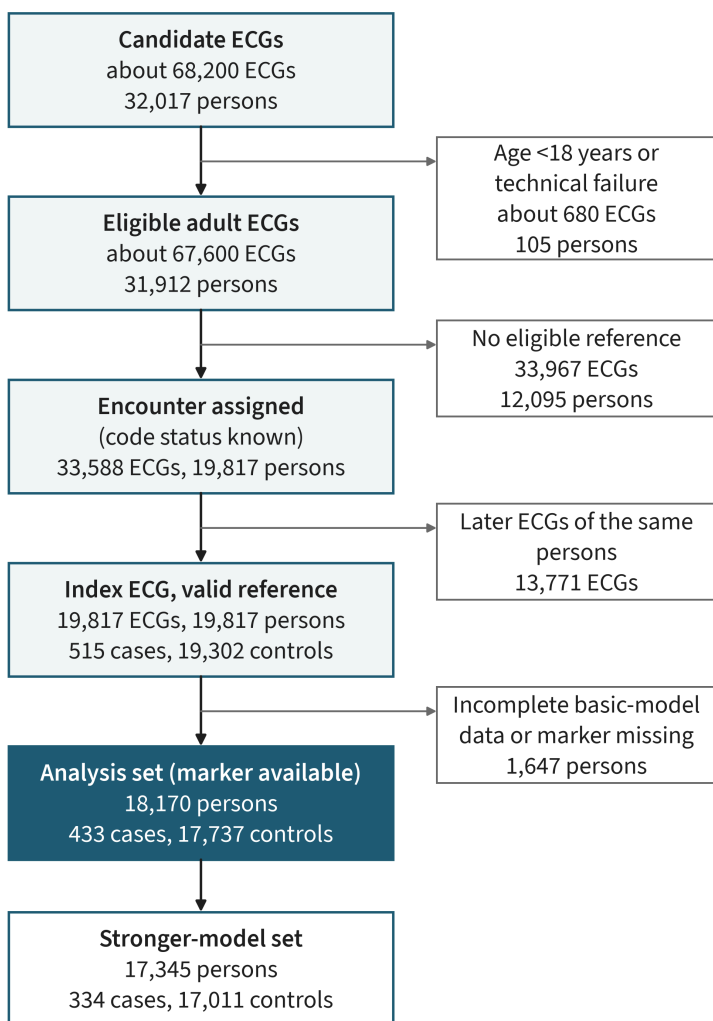
23. Pollard T, Moody B, Lehman LW, Gow B, Fernandes C, Xie C et al (2026) PhysioNet As A Global Platform for Biomedical Research. *Nat Health* 1:792–795. 10.1038/s44360-026-00096-z
24. Gow B, Pollard T, Greenbaum N, Moody B, Han A, Waks JW et al (2026) MIMIC-IV-ECHO: Echocardiogram Matched Subset. 10.13026/nrjh-5r77. RRID:SCR_007345
25. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M et al (2021) 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J* 42:3599–3726. 10.1093/eurheartj/ehab368
26. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M et al (2023) 2023 Focused Update of the 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J* 44:3627–3639. 10.1093/eurheartj/ehad195
27. Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L et al (2015) Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *J Am Soc Echocardiogr* 28:1–39e14. 10.1016/j.echo.2014.10.003
28. Kors JA, van Herpen G, Sittig AC, van Bommel JH (1990) Reconstruction of the Frank vectorcardiogram from standard electrocardiographic leads: diagnostic comparison of different methods. *Eur Heart J* 11:1083–1092. 10.1093/oxfordjournals.eurheartj.a059647
29. Eiter T, Mannila H (1994) Computing discrete Fréchet distance. Technical report CD-TR 94/64. Vienna: Christian Doppler Laboratory for Expert Systems, Vienna University of Technology
30. Okin PM, Roman MJ, Devereux RB, Kligfield P (1995) Electrocardiographic identification of increased left ventricular mass by simple voltage-duration products. *J Am Coll Cardiol* 25:417–423. 10.1016/0735-1097(94)00371-v
31. 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
32. Edenbrandt L, Pahlm O (1988) Vectorcardiogram synthesized from a 12-lead ECG: superiority of the inverse Dower matrix. *J Electrocardiol* 21:361–367. 10.1016/0022-0736(88)90113-6
33. Emerek K, Friedman DJ, Sørensen PL, Hansen SM, Larsen JM, Risum N et al (2019) Vectorcardiographic QRS area is associated with long-term outcome after cardiac resynchronization therapy. *Heart Rhythm* 16:213–219. 10.1016/j.hrthm.2018.08.028
34. Stabenau HF, Waks JW (2023) BRAVEHEART: Open-source software for automated electrocardiographic and vectorcardiographic analysis. *Comput Methods Programs Biomed* 242:107798. 10.1016/j.cmpb.2023.107798
35. Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S (2014) Good practices for quantitative bias analysis. *Int J Epidemiol* 43:1969–1985. 10.1093/ije/dyu149
36. Nab L, Groenwold RHH (2021) Sensitivity analysis for random measurement error using regression calibration and simulation-extrapolation. *Glob Epidemiol* 3:100067. 10.1016/j.gloepi.2021.100067
37. Cook JR, Stefanski LA (1994) Simulation-extrapolation estimation in parametric measurement error models. *J Am Stat Assoc* 89:1314–1328. 10.1080/01621459.1994.10476871

38. Extramiana F, Haggui A, Maison-Blanche P, Dubois R, Takatsuki S, Beaufile P et al (2007) T-wave morphology parameters based on principal component analysis reproducibility and dependence on T-offset position. *Ann Noninvasive Electrocardiol* 12:354–363. 10.1111/j.1542-474X.2007.00185.x
39. Perez-Alday EA, Hamilton C, Li-Pershing A, Monroy-Trujillo JM, Estrella M, Sozio SM et al (2010) The Reproducibility of Global Electrical Heterogeneity ECG Measurements. *Comput Cardiol* 2018;45. 10.22489/cinc.2018.162
40. Young WJ, van Duijvenboden S, Ramírez J, Jones A, Tinker A, Munroe PB et al (2021) A Method to Minimise the Impact of ECG Marker Inaccuracies on the Spatial QRS-T angle: Evaluation on 1,512 Manually Annotated ECGs. *Biomed Signal Process Control* 64:102305. 10.1016/j.bspc.2020.102305
41. Cro S, Morris TP, Kenward MG, Carpenter JR (2020) Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: A practical guide. *Stat Med* 39:2815–2842. 10.1002/sim.8569
42. Hernán MA, Hernández-Díaz S, Robins JM (2004) A structural approach to selection bias. *Epidemiology* 15:615–625. 10.1097/01.ede.0000135174.63482.43
43. Westreich D, Cole SR (2010) Invited commentary: positivity in practice. *Am J Epidemiol* 171:674–677. 10.1093/aje/kwp436
44. Greenland S, Pearl J, Robins JM (1999) Confounding and collapsibility in causal inference. *Stat Sci* 14:29–46. 10.1214/ss/1009211805
45. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP et al (2007) The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 370:1453–1457. 10.1016/S0140-6736(07)61602-X
46. Benchimol EI, Smeeth L, Guttman 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
47. Heidari S, Babor TF, De Castro P, Tort S, Curno M (2016) Sex and Gender Equity in Research: rationale for the SAGER guidelines and recommended use. *Res Integr Peer Rev* 1:2. 10.1186/s41073-016-0007-6
48. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B et al (2024) TRIPOD + AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 385:e078378. 10.1136/bmj-2023-078378
49. Kalyakulina A, Yusipov I, Moskalenko V, Nikolskiy A, Kosonogov K, Zolotykh N et al (2021) Lobachevsky University Electrocardiography Database. 10.13026/eegm-h675. RRID:SCR_007345
50. Kalyakulina AI, Yusipov II, Moskalenko VA, Nikolskiy AV, Kosonogov KA, Osipov GV et al (2020) LUDB: A New Open-Access Validation Tool for Electrocardiogram Delineation Algorithms. *IEEE Access* 8:186181–186190. 10.1109/ACCESS.2020.3029211
51. Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G et al (2019) Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat Med*

25:70–74. 10.1038/s41591-018-0240-2

52. Attia IZ, Tseng AS, Benavente ED, Medina-Inojosa JR, Clark TG, Malyutina S et al (2021) External validation of a deep learning electrocardiogram algorithm to detect ventricular dysfunction. *Int J Cardiol* 329:130–135. 10.1016/j.ijcard.2020.12.065
53. van Stipdonk AMW, Ter Horst I, Kloosterman M, Engels EB, Rienstra M, Crijns HJGM et al (2018) QRS Area Is a Strong Determinant of Outcome in Cardiac Resynchronization Therapy. *Circ Arrhythm Electrophysiol* 11:e006497. 10.1161/CIRCEP.118.006497
54. Plesinger F, van Stipdonk AMW, Smisek R, Halamek J, Jurak P, Maass AH et al (2020) Fully automated QRS area measurement for predicting response to cardiac resynchronization therapy. *J Electrocardiol* 63:159–163. 10.1016/j.jelectrocard.2019.07.003
55. Koscova Z, Li Q, Robichaux C, Junior VM, Ghanta M, Gupta A et al (2026) The Harvard-Emory ECG Database. *Sci Data* 13:516. 10.1038/s41597-026-06861-9
56. Poterucha TJ, Jing L, Ricart RP, Adjei-Mosi M, Finer J, Hartzel D et al (2025) Detecting structural heart disease from electrocardiograms using AI. *Nature* 644:221–230. 10.1038/s41586-025-09227-0

Figures

A Reduced LVEF (<40%)**B Coded cardiomyopathy**

The two endpoints share persons (8,306 at the reference stage, 7,052 in the analysis sets); counts cannot be added.

Figure 1

Participant flow in the assessment partition for (A) reduced left ventricular ejection fraction (LVEF <40%) and (B) coded cardiomyopathy. Main boxes give the numbers of ECGs and persons remaining after each step; side boxes give exclusions. One index ECG per person was analyzed, the first technically successful ECG with an eligible reference. Cases are LVEF <40% (A) or a cardiomyopathy code in the index hospital encounter (B). The stronger-model set contains persons with complete data for the additional voltage and axis terms. Counts below 10 are suppressed; in (B), the losses from incomplete basic-model data and from a missing marker are combined because the count for the missing marker alone is below 10. The ECG counts of the first two stages and the technical exclusions are rounded to the nearest 100 or 10. The endpoints share persons (8,306 at the reference stage, 7,052 in the analysis

sets), so counts must not be added. Source: Supplement Section S15a. Additional assessment sensitivities and descriptive summaries were specified after the primary results were known (post hoc); original specified association rows retain their separate status.

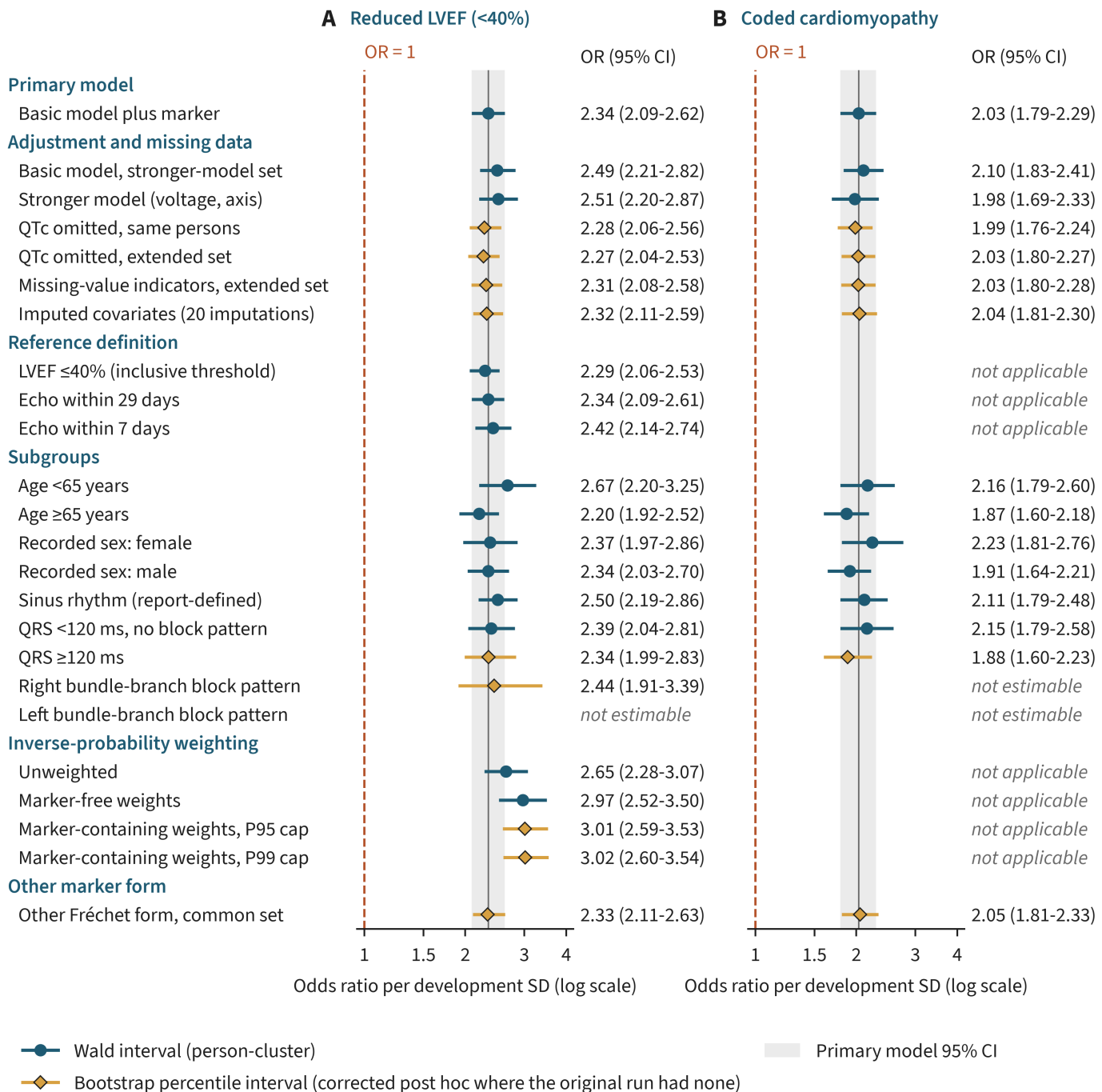


Figure 2

Odds ratios of the normalized Frechet distance for (A) reduced LVEF and (B) coded cardiomyopathy in the assessment partition, with sensitivities and strata of the primary model.

Association of the normalized Fréchet distance with (A) reduced LVEF (<40%) and (B) coded cardiomyopathy in the assessment partition. Points are odds ratios (OR) per endpoint-specific

development standard deviation of the marker, adjusted for age, sex, heart rate, QRS duration, QTc and the integral QRS-T angle; rows are sensitivities and strata of the primary model. Filled circles show person-cluster robust Wald 95% confidence intervals; diamonds show person-bootstrap percentile 95% intervals, including post hoc corrected intervals for rows whose bootstrap interval was missing in the original run (Supplement Section S18). The gray band is the Wald interval of the primary model; the dashed line marks OR = 1. “Echo within 29 days” and “within 7 days” restrict the interval between ECG and echocardiogram (primary analysis: within 30 days; the 29-day variant avoids the boundary region affected by hour-scale timestamp offsets). “Imputed covariates” is multiple imputation of missing covariates (20 imputations). Inverse-probability weighting rows use the first technically successful ECG, a different population from the main model. “Not estimable” marks strata with too few persons per class, “not applicable” analyses defined only for the LVEF reference. Further details, persons per row and post hoc status are in Supplement Section S16.

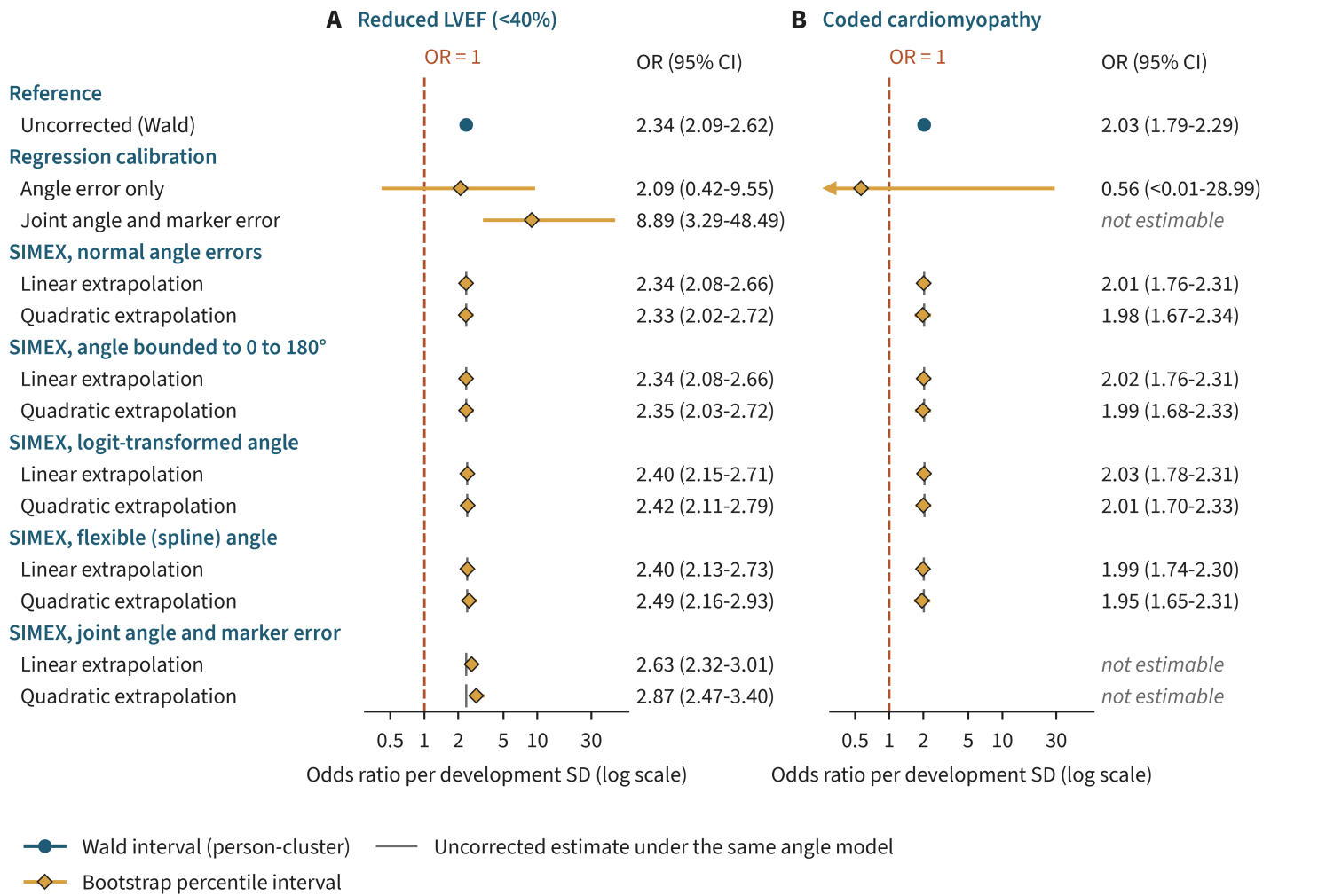


Figure 3

Marker odds ratio after correction for measurement error of the QRS-T angle (regression calibration and simulation-extrapolation) in the assessment partition, for (A) reduced LVEF and (B) coded cardiomyopathy. Regression calibration (RC) and simulation-extrapolation (SIMEX) use error variances

from clinical repeat ECGs and development-fixed settings; the error package itself is treated as fixed. SIMEX is shown with linear and quadratic extrapolation and four angle error models (normal, bounded to 0-180 degrees, logit-transformed, flexible spline angle); “joint” corrects angle and marker error together. Diamonds with whiskers are person-bootstrap percentile 95% intervals (1,000 attempted fits; angle-only RC in (B), 989 valid), the circle is the Wald interval of the uncorrected model, and the short gray bar marks the uncorrected estimate under the same angle model. The arrow marks a lower limit beyond the axis (4.2×10^{-5}). Joint correction for cardiomyopathy was not estimable because no error covariance was available for the beat-median marker. Repeat differences include true clinical change; neither a retained OR nor an unstable RC estimate proves the absence of measurement error. Additional assessment sensitivities and descriptive summaries were specified after the primary results were known (post hoc); original specified association rows retain their separate status. The upper limit of joint LVEF RC (48.49) is printed beside the estimate.

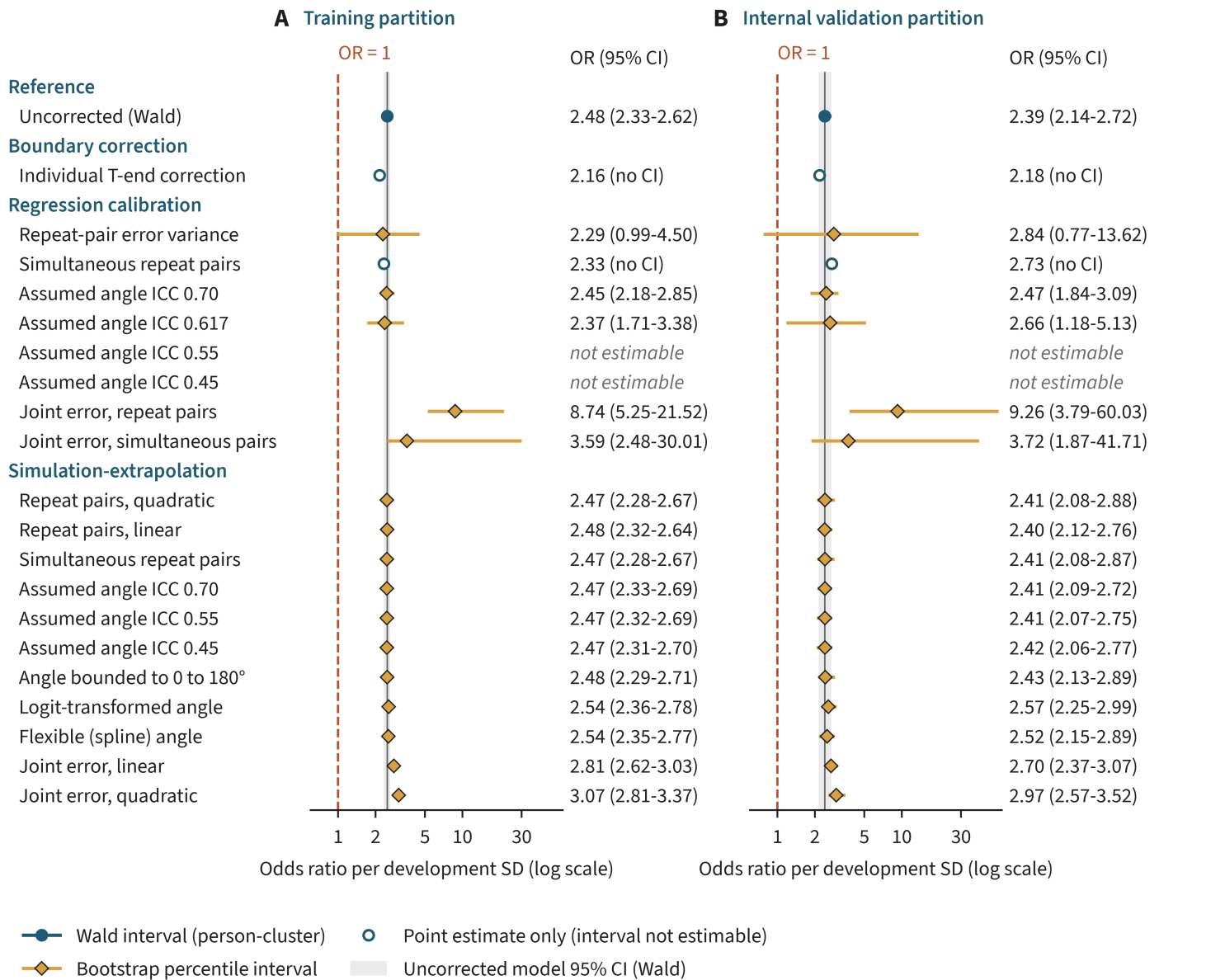


Figure 4

Developmental measurement-error sensitivities of the marker odds ratio for reduced LVEF in (A) the training and (B) the internal validation partition (development data, not independent replication). Rows show regression calibration (RC) and simulation-extrapolation (SIMEX) with error variances from all clinical repeat pairs or from simultaneous pairs, with assumed overall angle intraclass correlations (ICC), with bounded, logit-transformed and flexible (spline) angle models, and with joint angle and marker error. Filled circles are Wald 95% intervals, diamonds person-bootstrap percentile 95% intervals, open circles point estimates whose interval was not estimable (T-end correction: 17 of 500 and 100 of 500 valid draws on the finite-output sets of 25,388 and 8,495 persons; simultaneous-pair RC: fewer than 475 of 500 valid draws). RC was not estimable at assumed ICC 0.55 and 0.45 because conditional reliability was nonpositive. The gray band is the Wald interval of the uncorrected model; the logit rows use their own transformed-angle reference (2.524 and 2.484). Clinical repeats include physiological change.

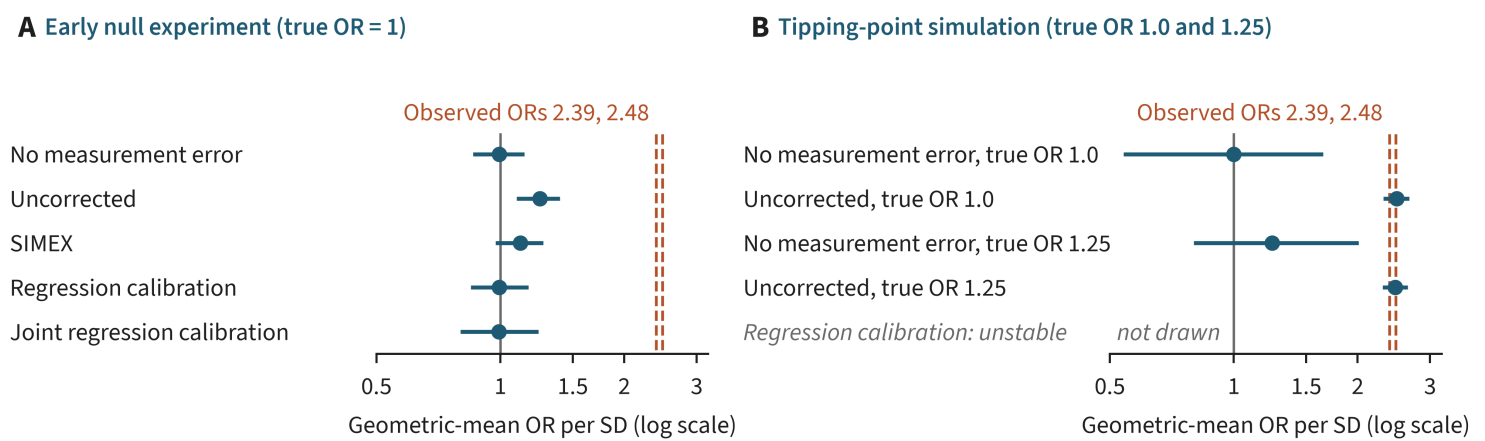


Figure 5

Two known-truth simulations of classical angle error: (A) early null experiment and (B) simulation at the angle-only tipping points.

Two known-truth simulations of classical angle error, shown with equal weight. (A) Early null experiment: 200 synthetic datasets of 4,000 observations, true angle log-effect 0.8, no independent marker effect (true OR = 1); the marker combines a latent-angle component with independent noise, and the angle error is independent of the marker. (B) Later simulation at the angle-only tipping points of the bivariate analysis (25,453 persons per dataset, 200 datasets per scenario, real values of the other covariates, error-free marker, strong true angle effect): true marker OR 1.0 at angle reliability 0.5688 and true marker OR 1.25 at 0.5698. Points are geometric-mean ORs of the marker and whiskers the 95% simulation quantiles across datasets (not confidence intervals for a biological effect). “No measurement error” uses the error-free angle; in (B) the regression-calibration estimates are inadmissible in 32 and 55 of 200 runs and scatter over several orders of magnitude, so they are not drawn. Dashed lines mark the observed uncorrected development ORs (2.48 training, 2.39 validation). Neither simulation bounds the OR that artifact can produce: (A) gives about 1.25, (B) about 2.5, at a true OR of 1 or 1.25. The genuine-effect scenario of the early experiment is in Supplement Table S47.

Supplementary Files

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

- [ProbstQRSTcouplingLVEFsupplement.pdf](#)