

Heart-rate-based clock alignment in intensive-care ECGs, serial-ECG matching and administrative follow-up in MIMIC-IV

Leo Probst

`leo.probst.research@proton.me`

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

Research Article

Keywords: electrocardiography, electronic health records, time alignment, MIMIC-IV, repeated measurements, data quality

Posted Date: October 20th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11410281/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

Objective:

To audit the timing, matching and observation requirements of serial electrocardiogram (ECG) studies that use routine clinical records in an intensive-care subset (9% of recordings).

Approach:

We used two person-disjoint, previously inspected development partitions of MIMIC-IV-ECG linked to MIMIC-IV. Intensive care unit (ICU) heart-rate alignment was assessed against identifiability criteria, simulated offsets and two null constructions. Pairs: development, internal assessment.

Main results:

Offsets were identifiable for 392 of 17,067 (2.3%) and 145 of 5,846 (2.5%) ICU-linked people; 91% were lost before estimation. Among people with at least three complete ECGs, the criteria were met in 27.1% as observed and in 19% to 22% under null constructions that weaken the alignment. Only 0.5% of ICU-linked people (117 of 1,979, 5.9%) had an offset with additional algorithmic support, not a calibrated uncertainty. Of the identifiable, 67.1% and 70.3% lay within ± 60 min (86% in simulation at zero offset), so the missed 90% criterion tested the procedure, not the clocks. Potassium matching limits rose from 352 and 69 to 562 and 136 sets when the baseline tolerance widened from 0.3 to 1.0 mmol/L. With exact timestamps, 34.6% and about 36% of index discharges had no later ECG or hospital discharge (36.7% and about 38% including deaths dated on or before discharge; emergency visits not counted as contact), and the follow-up span ended before 365 days in 26.4% and 25.6% of people with later contact and no documented atrial fibrillation or flutter (AF). A control in MC-MED (emergency department), locked before analysis and not a validation, missed its prespecified threshold (28.3% of 11,913 eligible visits), neither confirming nor refuting the MIMIC results.

Significance:

Individual clock offsets were rarely identifiable by heart-rate matching against intermittently charted rates. Matching counts depend on tolerances, and absent later contact is not absent events.

Key findings

- Individual ECG clock offsets could rarely be identified with a heart-rate match against charted rates: 2.3% and 2.5% of ICU-linked people met the criteria, and 0.5% (117 of the 1,979 people with at least three complete ECGs, 5.9%) had an offset with additional algorithmic support (a permutation test and stability under leave-one-ECG-out; 8.8 of these 117 people are expected in permuted data). This is not a calibrated uncertainty.
- Most of the loss was structural for the ± 12 h range chosen by design (narrower ranges were not tested). Of ICU-linked people, 91% lacked three ECGs with complete ± 12 h reference coverage; the

largest single cause was that every ECG lay within 12 h of a stay boundary (48% to 49%).

- “Identifiable” says little by itself: among the 1,979 people with at least three complete ECGs, the criteria were met by 27.1% as observed and by 19% to 22% under null constructions that weaken the alignment. The excess lies at small offsets (365 estimates within ± 60 min against 36 expected), not at large ones.
- Matching counts depend on the chosen calipers (maximum allowed differences between matched people): for potassium within ± 6 h, 352 and 69 matched sets rose to 562 and 136 when the baseline-potassium caliper was widened from 0.3 to 1.0 mmol/L, which equals the change threshold itself.
- For clinical readers: “no later AF documented” is not “no AF”. The follow-up span ended before 365 days in about a quarter of people with recorded contact after discharge and no documented AF. About a third of index discharges had no later ECG or hospital discharge, and emergency-department stays were not counted as contact.

Evidence status

All MIMIC results use person-disjoint, previously inspected development partitions; no independent held-out confirmation is reported. The atrial-fibrillation (AF) cohort selection was redesigned after inspection of earlier results, including historically opened test data elsewhere in a larger analysis project (not reported here); no test data were accessed for this note. Analyses added after the first analyses (simulation, null constructions, individual support, reference restrictions, loss decomposition, exact-timestamp count, MC-MED diagnosis) are post hoc and marked as such. The MC-MED control was locked (specification and script) before its waveform analysis, with its test splits closed. Heart-rate compatibility estimates are not calibrated clock errors, and matching limits are not completed treatment-effect samples. Contact cessation does not certify event freedom.

1. Introduction

MIMIC-IV-ECG contains diagnostic electrocardiograms (ECGs) that can be linked to records in the electronic health record (EHR). Its documentation warns that the ECG device clock was usually not synchronized to an external source (Gow *et al* 2023). The clinical database keeps relative times intact through a person-specific date shift (Johnson *et al* 2023a). This preserves the differences between recorded times but does not establish the physical accuracy of the source clocks.

Shared physiological signals are an established basis for synchronization (Xiao *et al* 2022, Goodwin *et al* 2023). Where devices can be co-recorded prospectively, device timestamps can instead be regressed on a master clock (Fruet *et al* 2026). Both approaches use richer temporal information than a short diagnostic ECG compared with intermittently charted heart rate. EHR timestamps also contain documentation error (Ward *et al* 2013, Tung *et al* 2022). Their error distributions cannot be transferred to MIMIC-IV-ECG.

Our question is what timing, matching and observation information should be reported before serial ECG-EHR comparisons are interpreted. Electrolyte-linked recordings are examples, because nominal laboratory proximity does not establish the order of physiological changes. We quantify these requirements without estimating treatment effects or diagnostic performance, and we provide a reusable reporting table. The numbers describe one hospital. The method (a population minimum with identifiability criteria, counting comparable opportunities before estimating, and explicit observation rules for event endpoints) can be applied elsewhere, and we made one prespecified technical check of it in a second data set.

2. Methods

2.1. Data and scope

MIMIC-IV-ECG version 1.0 identifies MIMIC-IV version 2.2 as its parent resource (Gow *et al* 2023, Johnson *et al* 2023b). We linked ECG records to the MIMIC-IV version 3.1 hospital and intensive-care unit (ICU) modules by the de-identified person key; the same person-specific date shift applies across modules (Johnson *et al* 2024), which is module consistency, not a correction of device-clock error. The documentation does not state explicitly that person identifiers and date shifts are unchanged between versions 2.2 and 3.1; we assume this, and the population heart-rate error minimum at 0 min (Fig. 1A) would be unlikely if the shifts differed, which is an indication and not proof. The ECG resource covers routine care at Beth Israel Deaconess Medical Center, Boston, 2008–2019; the documentation names Burdick/Spacelabs, Philips and General Electric machines, but the recorder type of an individual record is not coded. The two partitions comprise 640,204 recordings of 129,259 people (96,680 and 32,579); a third partition of the resource was reserved and not accessed.

LoopLens (version 10.1; release of 7 October 2026) is an in-house Python pipeline for ECG delineation and marker calculation; the author is its developer. It supplied the stored context data and, for the complete-case subset shown in Supplement S3.2 only, one stored marker. The LoopLens outputs come from the A155 run, whose feature-extraction code is identical to the release of 7 October 2026 (feature-code checksum recorded by the extraction run). Development and internal assessment are two previously inspected, person-disjoint development partitions. The temporal criteria were fixed before calculation; two implementation corrections followed diagnostic runs. The matching-window choices and the atrial-fibrillation (AF) cohort selection were developed iteratively. Everything added after the first analyses was specified before calculation but after the earlier results had been inspected, and is marked post hoc, including the correction of an arithmetic error (Section 2.4). The work was not prospectively registered.

2.2. Heart-rate alignment audit

ECG heart rate was 60,000 divided by the machine-reported mean RR interval in milliseconds (30 to 220 beats per minute; device and record timestamps within 60 s). ECGs were assigned to a unique ICU stay by their nominal timestamps. The reference was the charted heart rate (item 220045 of the ICU chart

events), reduced to its median at duplicate times. Of its timestamps, 87.2% fell on a full hour, and storage lagged charting by a median of 23 min (interquartile range 8–55). Whether a value came from a monitor or was typed in is not coded, so machine RR and charted rate are not independent measurements.

Candidate offsets ranged from -12 h to $+12$ h in 5 min steps. The ± 12 h range is a design choice; the structural losses refer to it, and narrower ranges were not tested. For each candidate ECG time, the nearest reference within 30 min in the same stay was selected (earlier first in ties); positive offsets compared the ECG with a later reference. Only ECGs with a reference at every offset contributed to complete error curves. The person-specific error curve was the median absolute heart-rate difference across complete ECGs, and the population curve was the median of these curves (Fig. 1A). An individual offset was called identifiable only with at least three complete ECGs, a connected near-optimal region (within 1 beat per minute of the minimum) no wider than 120 min, a minimum error of at most 5 beats per minute, and no near-optimal point at the search boundary. The specified criteria required a median offset within ± 15 min and at least 90% of identifiable offsets within ± 60 min. “Identifiable” means that these algorithmic criteria were met; it is not a statement about clock truth.

Uncertainty was summarized with 1,000 person bootstrap samples per partition; the intervals reflect between-person sampling, not individual clock-error uncertainty. A second audit pooled ECGs within ICU stays under stricter joint criteria (Supplement S1.5).

The following procedures are post hoc. Shifting ECG times with unchanged rates moves the objective function mechanically and only checks the implementation. To test recovery of unknown offsets, we simulated known offsets (0, ± 30 , ± 60 , ± 120 and ± 240 min) on the observed chart sampling, ECG times and stay boundaries, with an autocorrelated latent rate, calibrated noise and documentation delays of 0, 5 and 15 min. Two null constructions weaken the alignment: within-person permutation of the ECG rates (200 replications) and circular shifts of each stay’s chart series by at least $\min(24 \text{ h}, \text{one quarter of the stay span})$ (500 replications), which keep its autocorrelation; decoupling is weaker for short stays. Individual support was assessed by leave-one-ECG-out stability and a within-person permutation p value (1,000 permutations), calibrated against 20 permuted data sets. We restricted the analysis to ECGs with machine-reported sinus rhythm and no pacing text, tried stricter reference conditions (Supplement S1.12), and decomposed losses in a fixed order (Supplement S1.3).

2.3. Pair and matching feasibility

The potassium screen used consecutive, distinct ECG recordings from the same person and hospital stay, 2 h to 72 h apart, each linked to the nearest usable potassium measurement within ± 2 h or ± 6 h (assessed separately); different specimens, a consistent nominal order and values of 2.5 to 6.0 mmol/L were required. Change pairs differed by at least 1.0 mmol/L; control pairs had a small measured difference of at most 0.2 mmol/L, and intermediate pairs were not used.

The generous matching limits used all available opportunities, excluded self-matching and known device-label changes, and required adequate stored signal quality and heart rate. Matching was exact on partition, potassium item, sex category and laboratory timing side, with calipers (maximum allowed differences between the matched people) for age, initial heart rate, initial potassium and ECG interval (Table 1; Supplement S2). Each matched set links one change person and one control person. The limits omit unresolved requirements for treatment observation, clinical context, rhythm and full measurement eligibility; they are upper limits, not completed treatment-effect samples.

The calcium and heart-rate study used a ± 6 h window, screens for QRS duration, quality and device label, stored heart-rate and QRS proxies and total calcium; albumin, pH and ionized calcium were unavailable, so a stable total calcium does not indicate a stable electrophysiologic state. The common-beat population needed to compare QT correction formulas was not established, so these counts give no upper limit for the intended QT analysis (Supplement S2.3). The troponin design was excluded by rule (temporal-support indicator false throughout), not by an empirical zero-pair finding. Counts of 100 potassium sets and 50 calcium and QT sets per layer and partition were internal planning rules without power or precision derivation.

2.4. Administrative follow-up span

The AF example is a descriptive count over stored index discharges, not an incidence analysis. The first eligible discharge was selected, with the last reported-sinus ECG with regular waveform metadata 0–7 days before it. Known prior, index or through-discharge AF was excluded; risk time began at discharge. AF (atrial fibrillation or flutter) meant first later documentation, from diagnosis-linked admission-time proxies or ECG-report times, not adjudicated onset. Administrative codes for atrial fibrillation identify cases with a pooled sensitivity of 80% (95% confidence interval 72–86%) and a positive predictive value of 88% (82–94%) (Yao et al 2019), so a code-based label is a documentation target. The choice of index discharge and ECG used no marker, outcome or follow-up information; the risk set then required consistent event times and a positive span.

The administrative span runs from the index discharge to the last recorded contact, the later of the last ECG and the last hospital discharge; a positive span means at least one recorded ECG or hospital discharge after the index discharge. Emergency-department stays and laboratory events were not contact sources, and a later contact does not make the time in between observed. No calendar data-end date was applied, so truncation at the end of the data appears as contact end. The span was cut at the earliest of documented AF, its own end, recorded death or 365 days; AF took precedence when ends coincided, and other ties were counted separately. Death dates are calendar dates taken as 00:00, so a death on the discharge date counts as before discharge. Death dates in MIMIC-IV combine hospital records with linked Massachusetts state death records; deaths outside the hospital are recorded for up to one year after a person's last hospital discharge, and later deaths are censored by the provider (Johnson et al 2023a, 2024). A missing death date therefore means survival only within that window. Death was censored although it competes with AF (Austin et al 2016). The stored follow-up variable had

been built with a database function that counts crossed hour boundaries rather than elapsed time; all differences are now computed in seconds (Supplement S3.4).

2.5. External technical check in MC-MED

As an external technical check, we applied a specification locked before the waveform analysis to the training split of MC-MED (Kansal *et al* 2025a; adult emergency-department visits, one center, 2020–2022). A control compared waveform-derived heart rate with monitor heart rate, whose target offset is zero by construction, within ± 60 min. The primary relation, charted rate against waveform within ± 15 min, stayed locked unless the control identified at least 80% of eligible visits, at least 95% of these within ± 2 min, and a population minimum at zero (Supplement S7).

2.6. 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.

3. Results

3.1. Population and individual alignment

The population heart-rate error minimum was at 0 min in both partitions (Fig. 1A). ICU ECGs with a usable machine heart rate were 42,235 of 477,363 development recordings (8.8%) and 14,233 of 162,841 internal-assessment recordings (8.7%). Emergency-department ECGs could not be audited because the reference, the ICU chart, does not cover them.

Most people never reached the curve criteria (Fig. 2). Of 17,067 and 5,846 ICU-linked people, 91.3% and 91.5% did not have three ECGs with complete ± 12 h reference coverage. The largest single loss was proximity to a stay boundary: for 48.4% and 49.4% every ECG lay within 12 h of the start or end of an ICU stay, and 10.0% and 10.1% had only stays shorter than 24 h. Reference gaps removed a further 14.8% and 14.0%, fewer than three complete ECGs another 27.0% and 26.8%, and failure of the curve criteria 6.4% and 6.0% (Supplement S1.3). People with at least three complete ECGs had longer ICU stays (median 10.9 against 2.3 days), more ICU ECGs (7 against 1) and higher in-hospital mortality (27.4% against 12.2%), but similar age and sex.

Individual offsets were identifiable for 392 and 145 people: 2.3% and 2.5% of ICU-linked people, 26.5% (95% confidence interval 24.3–28.8%) and 29.2% (25.4–33.3%) of the 1,482 and 497 with at least three complete ECGs, and 0.41% and 0.45% of all people in the two ECG partitions. Among them, 263 of 392 estimates (67.1%; 95% confidence interval 62.2–71.6%) and 102 of 145 (70.3%; 62.4–77.3%) were within ± 60 min (Fig. 1B and 1C), 75.5% and 80.0% within ± 120 min and 88.0% and 91.0% within ± 360 min. Both partitions failed the specified criterion of 90% within ± 60 min (Section 3.2 gives the simulated

benchmark); the median-offset criterion was met in each (0 min). The near-optimal width had a median of 35 min (interquartile range 15–55) and 45 min (20–55); entire near-optimal zones lay inside ± 60 min for 54.8% and 57.2%. People without an identifiable estimate cannot be classified as having inaccurate clocks.

Pooling within ICU stays did not help: 350 of 4,467 and 129 of 1,477 eligible stays were identifiable, agreement between chronological halves was shown in 32 and fewer than 10 of them, and all specified stay-level criteria were unmet (± 60 min share 65.1% in both partitions; Supplement S1.5).

3.2. What the identifiability criteria carry (post hoc)

Shifting the ECG times with unchanged rates passed the fixed recovery check in both partitions; this checks the implementation, not whether an unknown offset can be found (Supplement S1.6). In the simulation, 85% to 87% of identifiable estimates lay within 60 min of the true offset and 43% to 48% within 15 min (74% at zero offset). With a simulated offset of zero, 32% of people met the criteria and 86% of the identifiable estimates lay within ± 60 min (89% at half the noise), so the specified 90% criterion could not have been met even with synchronized clocks. The observed yield of 27.1% and the 67.1% and 70.3% within ± 60 min should be read against these benchmarks and the null constructions; they describe the procedure, not the clocks. Estimates were pulled toward zero by a median of 10 to 15 min, a constant documentation delay of 15 min moved the median estimate to only about 5 min, and the identifiable share fell from 32% at zero offset to 24% to 26% at ± 240 min (Supplement S1.9).

Neither null construction made identifiability rare. Among the 1,979 people with at least three complete ECGs, the criteria were met in 22.0% (95% range 20.4–23.9%) after within-person permutation and in 19.0% (17.5–20.6%) after circular shifts, against 27.1% (537 people) as observed; none of 500 shifted replications reached the observed count. Identifiable estimates within ± 60 min numbered 365, against 36 (25–49) expected under circular shifts; beyond ± 60 min they numbered 172, against 339 (310–370). The criteria alone therefore separate little. The signal lies at small offsets, and large estimated offsets were not more frequent than under chance. A flag of “identifiable with an offset beyond ± 60 min” had a sensitivity of 25% for true offsets of at least 120 min and a positive predictive value of 7% to 21% if 5% of people had such offsets (Supplement S1.10).

Individual algorithmic support was narrower. Of the 537 people meeting the criteria, 232 (43.2%) had a permutation p value below 0.05 (31 expected in permuted data, range 18–41), and 117 (21.8%) also kept their estimate within 15 min when any single ECG was left out (9 expected, range 5–13). These 117 are 0.51% of the 22,913 ICU-linked people; in permuted data 8.8 of them (about 7.5%) are expected, and 2.7 of the more than 100 that also lay within ± 60 min (a small minority, below 5%). This support is not calibrated individual uncertainty. For 39% of identifiable people the estimate moved by more than two hours when one ECG was left out, against 70% in permuted data. For 129 people with only three usable ECGs a p value below 0.05 cannot be reached, so their individual uncertainty is undetermined (Supplement S1.11).

Restricting to sinus rhythm without pacing text left yield and location essentially unchanged (222 of 883 and 75 of 292 identifiable; 67.1% and 65.3% within ± 60 min). Stricter reference conditions (storage lag up to 15 min, off-hour times, narrower reference-matching windows) left no person with three complete ECGs, because complete ± 12 h coverage cannot be met with sparse, hour-rounded charting; this limits the design and says nothing about clock accuracy (Supplement S1.12).

3.3. Feasibility of controlled serial comparisons

Table 1 separates initial pairs from matched sets. The wider potassium window produced a larger initial pool as well as more possible matches; it was not a reanalysis of the same 615 pairs. Adequate treatment observation remained unresolved.

Table 1
Feasibility counts in the two development partitions.

Design and laboratory window	Initial pairs, development	Initial pairs, internal assessment	Matched sets, development	Matched sets, internal assessment	Interpretation
Potassium change, ± 2 h	615	234	≤ 41	< 10	Generous upper limits
Potassium change, ± 6 h	2,371	826	≤ 352	≤ 69	Separate wider-window pool; exploratory window choice
Calcium and heart-rate proxy, ± 6 h	355	122	79	11	Quality-screened proxy; common-beat population unknown

Table note: Each set links change and control opportunities from different people. $\leq$ marks generous upper limits; <10 marks a count below 10 that is not shown. Potassium initial pairs precede quality screening; calcium initial pairs are already quality/QRS/device screened (before those screens, 2,062 and 697 pairs). Rows are not additive. The gap class is the category of the time between the two ECGs (2 to < 6 h, 6 to < 24 h, 24 to 72 h).

At the original calipers, the ± 6 h development limit exceeded the internal planning level of 100 sets, whereas internal assessment did not. Relaxing baseline potassium from 0.3 to 0.5 and 1.0 mmol/L raised the ± 6 h limits to 425 and 562 matched sets in development and 99 and 136 in internal assessment; 1.0 mmol/L equals the change threshold itself. At ± 2 h, even combined relaxed calipers yielded only 134 and 37 sets (Supplement S2).

3.4. Administrative follow-up span

Compatible event and span references remained for 28,284 of 28,297 selected index discharges in development and for more than 9,745 in internal assessment. Of these, 10,371 and more than 3,660 had no positive administrative span after discharge: 597 and 202 had a recorded death date on or before the discharge date (day precision, so some also had a later contact), and the remainder (9,774 and more than 3,450) had no ECG or hospital contact after the index discharge. Emergency-department stays and laboratory events were not counted as contact. Thus 34.6% and about 36% of compatible index discharges had no later ECG or hospital discharge, and 36.7% and about 38% had no positive span when deaths dated on or before the discharge date are included. A positive span (the risk set) existed for 17,913 and 6,079 people (Table 2). A program-specific complete-case subset (finite loop-width marker, age, sex and heart-rate values) removes a further 35% and 34% of them and is shown in Supplement S3.2 (Table S20). Upstream selection was not reconstructed (Supplement S3).

Table 2

Administrative span at 365 days and the preceding loss of positive span; not incidence estimates.

Partition and stage	People	Documented AF	Span ended at last contact	Span ended by death	Span reached 365 days	Tied ends
Development, positive span	17,913	771	4,517	311	12,296	18
Internal assessment, positive span	6,079	275	1,485	113	> 4,190	< 10
Development, no positive span	10,371	n/a	n/a	n/a	n/a	n/a
Internal assessment, no positive span	> 3,660	n/a	n/a	n/a	n/a	n/a

Table note: AF = atrial fibrillation or flutter; n/a = not applicable; a positive span means at least one recorded ECG or hospital discharge after the index discharge. The last four columns apply to people without earlier documented AF, and each of the first two rows partitions its denominator exactly. A tied-end count below 10 is shown as < 10, and the horizon-end cell of that row is coarsened so that the tie count cannot be recovered. The last two rows are pre-risk selection losses. Times are exact to the second.

Among people with a positive span and no documented AF (17,142 and 5,804), the span ended at the last recorded contact in 26.4% and 25.6%. When contacts were required to follow each other within 30 days, fewer than 20 people kept a span to 365 days (Supplement S3.5). With integer-hour differences, 5,252 and more than 1,860 index discharges without any contact after discharge would have received a positive span, and the contact-end share would have been 43.6% and about 44%. Contact cessation marks loss of visible documentation, not all care or verified event freedom.

3.5. External technical check in MC-MED

The prespecified identifiability threshold was missed: only 28.3% of 11,913 eligible visits met the identifiability criteria. Among these, 98.5% lay within ± 2 min of the target offset zero, and null controls identified 5% to 6%. Identifiability did not fall monotonically with reference density, and two further prespecified controls missed their limits. The primary relation stayed locked. A post hoc diagnosis, outside the prespecified analysis, showed that the connectedness rule for the near-optimal region was the first rule violated for most non-identifiable visits and also separated data from null controls (without it, nulls were identifiable in 64% to 88% of visits). The chosen minimum lay within ± 2 min of zero in 82.0% of all eligible visits, against 5.9% and 4.7% under the nulls (Supplement S7).

4. Discussion

The documented clock warning for MIMIC-IV-ECG justifies questioning short-term ordering; our contribution concerns the procedure tested. A population minimum near nominal time coexisted with low individual identifiability, most of it structural before any estimation. The criteria alone were weakly informative, because about one in five people met them under null constructions that weaken the alignment, and an offset with additional algorithmic support remained in 0.5% of ICU-linked people. Heart-rate matching can assess compatibility with recorded clinical physiology, but it cannot separate clock error from physiological change, rhythm changes or charting delay; in the simulation estimates shrank toward zero and a 15 min delay stayed close to the nominal minimum. Without an external clock reference, neither a large estimated offset nor a flat curve identifies a true device-clock error. A concentration of estimates near zero is also what documentation coupling would produce, for example a rate read from the ECG device or charted at the time of the recording; because the reference origin is not coded, the excess at small offsets cannot be separated from aligned clocks. The population minimum stayed at 0 to 5 min when only values stored within 5 or 15 min, or only off-hour chart times, were used (Table S16), which argues against routine hourly charting as its only source but does not exclude coupling. Methods that use shared continuous signals can estimate alignment more directly (Xiao *et al* 2022, Goodwin *et al* 2023), and timestamp discrepancies reported elsewhere do not establish their magnitude here (Ward *et al* 2013, Tung *et al* 2022).

Laboratory windows describe nominal proximity, not physiological simultaneity, and enlarging a window after inspecting sparse counts is an exploratory design choice. At ± 6 h, relaxing the baseline-potassium caliper mattered more than relaxing heart rate, age or gap class (Supplement S2). For the design assessed here, comparing change pairs with pairs of small measured change requires separate recordings, comparable intervals and person-cluster uncertainty. That is not a universal requirement of serial analyses, and a small change selected by threshold can reflect measurement error and regression to the mean (Barnett *et al* 2005).

The loss of administrative follow-up is itself a finding of this count, in line with established concerns about informative presence and observation in routine health data (Sisk *et al* 2021). The span ended before 365 days in about a quarter of people with a positive span and no documented AF, and about a third of index discharges had no later ECG or hospital discharge; a recorded diagnosis is a

documentation target, and the span to the last contact is not continuous observation. Care outside a single health record system is common, and low record continuity goes with substantial misclassification of recorded patient characteristics (Lin *et al* 2018). An hour-boundary difference function silently gave index discharges without later contact a positive span, a reminder that time arithmetic belongs in an audit.

The MC-MED check neither confirms nor refutes the MIMIC results. The rule contributes at least in part to the low identifiability there, alongside the data or clocks; a rule calibrated on null controls is future work. The density explanation we had prepared was not supported. We make no statement on MIMIC device clocks and do not compare percentages across the two datasets.

The design has four main limits: only ECGs nominally inside ICU stays were used, complete offset coverage was required, heart rate was documented intermittently, and the reference origin is uncoded; no calibrated individual uncertainty interval for the offset was produced. A genuinely large offset could prevent nominal stay inclusion, and a constant offset per person ignores possible drift. The 5 min grid is finer than the effective resolution. Identifiability rose with ECG heart-rate variability (Supplement S1.7), consistent with a physiology-dependent measure. The internal assessment partition is an internal check, not independent validation, and the laboratory examples establish restricted feasibility, not biomarker effects. The clinical modules of MIMIC-IV version 3.1 cover admissions from 2008 to 2022 (on the shifted time axis), whereas the ECG resource ends in 2019; after 2019 only hospital discharges, not ECGs, can serve as contacts, and the effect on the contact-end counts was not examined. Table 3 lists the minimum items we propose to report.

Table 3
Minimum items to report in serial ECG-EHR studies.

Item	Minimum report
Population and units	Recordings, pairs, people and matched sets separately, with exclusive losses by stage.
Time semantics	Recording, specimen, storage and administration times; version linkage; reference origin; nominal versus supported ordering; durations from timestamps, not boundary counts.
Alignment	Candidate range, coverage, tie rule, a null with the same criteria, individual support (leave-one-out, permutation p value) and the population without estimates.
Controlled opportunities	Control recordings, calipers and sensitivity counts; limits and proxies separated from full eligibility.
Observation	Risk start, documented endpoint, last contact, death and horizon; no-positive-time losses and further selection steps.
Evidence and decisions	Inspected partitions, post-inspection choices, operational thresholds, limitations and validation status.

This audit informs study design and reporting. It does not support patient-level clock correction or a change in clinical management.

What this study does not show

No calibrated device-clock error, treatment effect, diagnostic performance or AF incidence is established. A favorable population curve or relaxed matching count does not resolve unobserved clinical eligibility.

5. Conclusion

For heart-rate matching against intermittently charted rates in the audited ICU subset, nominal linkage and a population minimum near zero did not establish individual clock accuracy or general ≤ 1 h ordering. Offsets met the identifiability criteria in 2.3% to 2.5% of ICU-linked people, mostly lost to structure before estimation; the criteria alone are weakly informative, and an offset with additional algorithmic support was found in 0.5%. Matching feasibility depends on the chosen calipers and should be counted before biological change is estimated from serial routine ECGs, and an event-free label requires explicit observation rules. The prespecified check in a second data set neither confirmed nor refuted these findings.

Declarations

Code availability

Analysis scripts and the frozen LoopLens release used for this study (PolyForm Noncommercial 1.0.0) are available at <https://github.com/leoprost/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.

Ethical Approval and Consent to participate

This study is a secondary analysis of de-identified research databases. 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 (Johnson *et al* 2023a). The creation and release of MC-MED was approved by the Stanford University Institutional Review Board and Secondary Data Use Committee with a waiver of consent (Kansal *et al* 2025a). Access to the credentialed MIMIC-IV and MC-MED data was obtained under the PhysioNet Credentialed Health Data Use Agreement after completion of the required Collaborative Institutional Training Initiative (CITI) training. MIMIC-IV-ECG is an open-access PhysioNet resource. The databases were de-identified by their providers before release; they were analyzed as anonymous data, and no attempt was made to re-identify individuals. Additional ethics approval or

informed consent was not sought, because the data are de-identified and the analysis follows the terms of the PhysioNet data use agreement, consistent with published secondary analyses of these resources.

Consent for publication

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

Availability of data and materials

All three resources are hosted on PhysioNet (Pollard *et al* 2026). MIMIC-IV-ECG (Gow *et al* 2023; version 1.0; doi:10.13026/4nqg-sb35; Open Data Commons Open Database License (ODbL) 1.0) is openly available. MIMIC-IV version 3.1 (Johnson *et al* 2023a, 2024; doi:10.13026/kpb9-mt58) and MC-MED (Multimodal Clinical Monitoring in the Emergency Department, version 1.0.1, published 25 September 2025; Kansal *et al* 2025a, 2025b; doi:10.13026/wvyw-g663) are available to credentialed users who sign the PhysioNet Credentialed Health Data Use Agreement 1.5.0, under the PhysioNet Credentialed Health Data License 1.5.0.

That agreement has no clause on aggregate or derived data; only aggregates are published here, with care to avoid identification of individuals or institutions (license clause 2). Derived patient-level data are treated as sensitive by PhysioNet and may only be shared on PhysioNet under the same agreement as the source data (PhysioNet guidance of 24 April 2024); they are therefore not provided here.

Aggregate table values, Figure 1B-C values and stay-level counts are provided in this note and the accompanying Supplement; the complete Figure 1A curve is supplied as Supplementary Data S1 (file supplementary_data_S5_S1_population_curve.csv, with a copy of identical content as .xlsx). An internal claim table linking each reported number to its source is held by the author; it is not a data repository. These documents do not release person-level data or private matching lists.

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

Leo Probst: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing (original draft), Writing (review and editing), Visualization, Project administration.

Acknowledgements

The author thanks the PhysioNet team and the investigators who created and maintain MIMIC-IV, MIMIC-IV-ECG and MC-MED.

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.

Reporting guidelines

The report follows the observational-data principles of Strengthening the Reporting of Observational Studies in Epidemiology (STROBE; von Elm *et al* 2007) and of the REporting of studies Conducted using Observational Routinely-collected health Data statement (RECORD; Benchimol *et al* 2015), together with the relevant agreement-reporting principles of the Guidelines for Reporting Reliability and Agreement Studies (GRRAS; Kottner *et al* 2011). Completed, applicability-aware checklists with explicit reporting gaps are in Supplement S5. Reporting also follows the Sex and Gender Equity in Research guidelines (SAGER; Heidari *et al* 2016). Sex as recorded in MIMIC-IV (an administrative field named gender) was an exact matching category in the laboratory feasibility screens; gender identity was not documented, and sex-specific clock accuracy was not evaluated.

Registration and related work

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. Two related manuscripts, one on repeat variation and definition dependence of an ECG/VCG marker catalog and one on QRS-T geometry and left-ventricular function, are in preparation for preprint posting. The first treats temporal linkage as a limitation; the present note addresses temporal support, controlled-pair feasibility and the AF administrative-span count, and its audit tables belong to this note. These projects share source data and measurement infrastructure, but their questions differ.

References

1. Austin PC, Lee DS, Fine JP (2016) Introduction to the analysis of survival data in the presence of competing risks *Circulation* 133 601-9 (10.1161/CIRCULATIONAHA.115.017719)
2. Barnett AG, van der Pols JC, Dobson AJ (2005) Regression to the mean: what it is and how to deal with it. *Int J Epidemiol* 34:215–220. 10.1093/ije/dyh299)
3. Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan S, M (RECORD Working Committee) 2015 The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med* 12 e1001885 (10.1371/journal.pmed.1001885)
4. Fruet D, Cimignolo S, Nollo G (2026) Multimodal data synchronization: a high-level software methodology for heterogeneous devices. *BMC Biomed Eng* 8 12. 10.1186/s42490-026-00114-x)
5. Goodwin AJ et al (2023) The truth Hertz-synchronization of electroencephalogram signals with physiological waveforms recorded in an intensive care unit. *Physiol Meas* 44:085002. 10.1088/1361-6579/ace49e)
6. Gow B et al 2023 MIMIC-IV-ECG: diagnostic electrocardiogram matched subset (version 1.0) *PhysioNet* (10.13026/4nqg-sb35)
7. 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)
8. Johnson AE, W et al 2023a MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data* 10 1 (10.1038/s41597-022-01899-x)
9. Johnson A, Bulgarelli L, Pollard T, Horng S, Celi LA, Mark R 2023b MIMIC-IV (version 2.2) *PhysioNet* (10.13026/6mm1-ek67)
10. Johnson A, Bulgarelli L, Pollard T, Gow B, Moody B, Horng S, Celi LA, Mark R 2024 MIMIC-IV (version 3.1) *PhysioNet* (10.13026/kpb9-mt58)
11. Kansal A, Chen E, Jin BT, Rajpurkar P, Kim DA 2025a MC-MED, multimodal clinical monitoring in the emergency department. *Sci Data* 12 1094 (10.1038/s41597-025-05419-5)
12. Kansal A, Chen E, Jin T, Rajpurkar P, Kim D (2025b) Multimodal Clinical Monitoring in the Emergency Department (MC-MED) (version 1.0.1) *PhysioNet*. 10.13026/wvyw-g663
13. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, Roberts C, Shoukri M, Streiner DL (2011) Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *J Clin Epidemiol* 64:96–106. 10.1016/j.jclinepi.2010.03.002)
14. Lin KJ, Glynn RJ, Singer DE, Murphy SN, Lii J, Schneeweiss S 2018 Out-of-system care and recording of patient characteristics critical for comparative effectiveness research *Epidemiology* 29 356 – 63 (10.1097/EDE.0000000000000794)
15. Pollard T, Moody B, Lehman L-W, Gow B, Fernandes C, Xie C, Johnson A, Mark R, Heldt T 2026 *PhysioNet as a global platform for biomedical research. Nat Health* 1 792–795 (10.1038/s44360-026-00096-z)

16. Sisk R, Lin L, Sperrin M, Barrett JK, Tom B, Diaz-Ordaz K, Peek N, Martin GP 2021 Informative presence and observation in routine health data: a review of methodology for clinical risk prediction. *J Am Med Inf Assoc* 28 155–166 (10.1093/jamia/ocaa242)
17. Tung T-H, DeLaurentis P, Yih Y 2022 Uncovering discrepancies in IV vancomycin infusion records between pump logs and EHR documentation. *Appl Clin Inf* 13 891–900 (10.1055/s-0042-1756428)
18. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative) (2007 The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies *Lancet* 370 1453-7 (10.1016/S0140-6736(07)61602-X)
19. Ward MJ, Froehle CM, Hart KW, Lindsell CJ (2013) Operational data integrity during electronic health record implementation in the ED *Am. J Emerg Med* 31:1029–1033. 10.1016/j.ajem.2013.03.027)
20. Xiao R, Ding C, Hu X 2022 Time synchronization of multimodal physiological signals through alignment of common signal types and its technical considerations in digital health. *J Imaging* 8 120 (10.3390/jimaging8050120)
21. Yao RJR, Andrade JG, Deyell MW, Jackson H, McAlister F, A, Hawkins NM (2019) Sensitivity, specificity, positive and negative predictive values of identifying atrial fibrillation using administrative data: a systematic review and meta-analysis. *Clin Epidemiol* 11:753–767. 10.2147/CLEP.S206267)

Figures

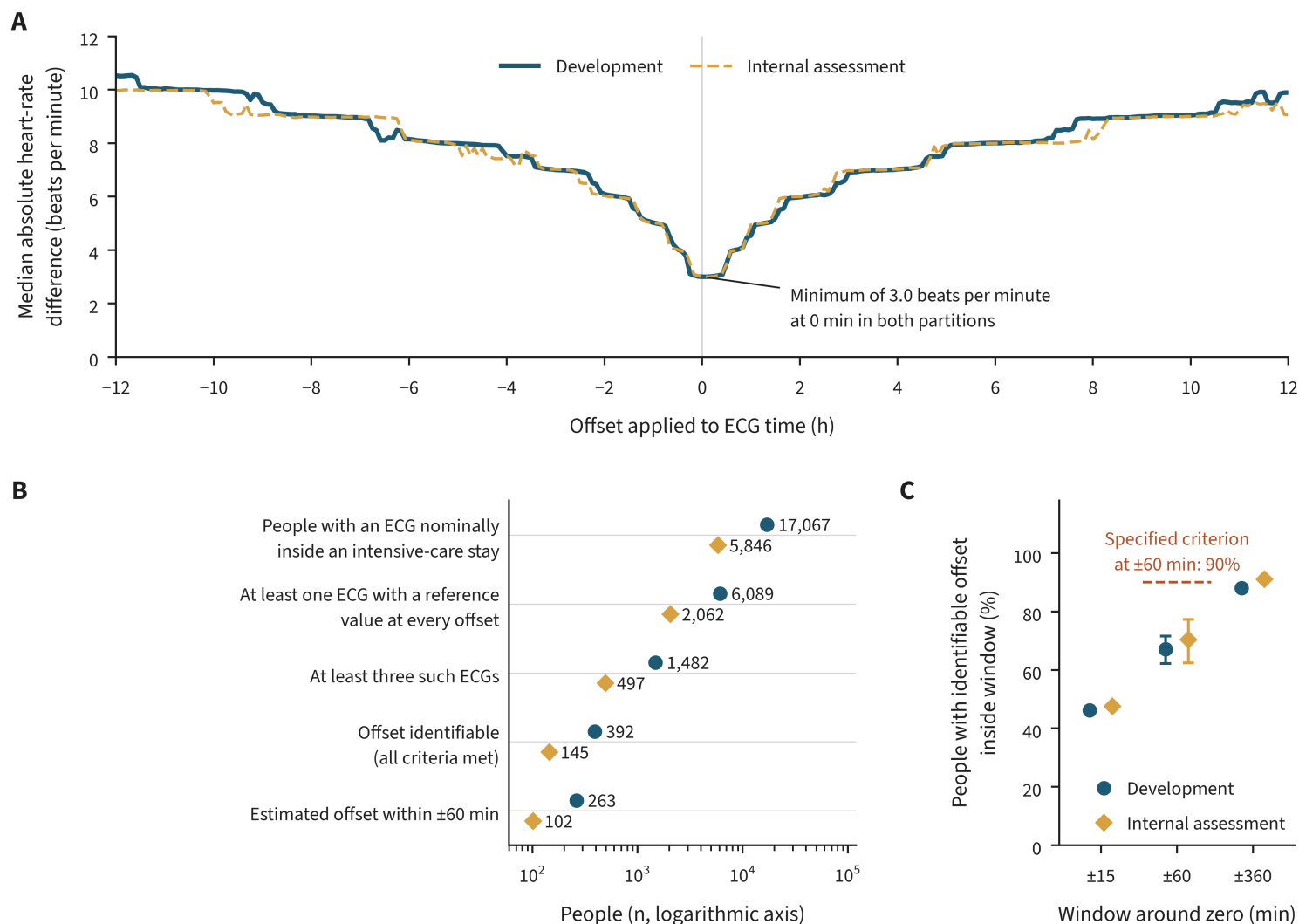


Figure 1

Heart-rate agreement over candidate clock offsets and its individual-level identifiability in two person-disjoint development partitions (A, population error curve; B, people at each step; C, share of identifiable people within the offset windows).

Heart-rate agreement over candidate clock offsets and its individual-level identifiability in two person-disjoint development partitions (development, 17,067 people; internal assessment, 5,846 people). (A) Population error curve: median over people of each person's median absolute difference between ECG-derived and charted heart rate, by the offset applied to ECG time (positive values compare the ECG with a later chart value; no interval computed). (B) Number of people at each step, from an ECG nominally inside an ICU stay to an estimated offset within ± 60 min of zero (logarithmic axis). (C) Share of the 392 and 145 people with an identifiable offset whose estimate lay within ± 15 , ± 60 and ± 360 min of zero; error bars at ± 60 min are 95% bootstrap intervals, and the dashed line marks the specified 90% criterion. Circles, development; diamonds, internal assessment. The offsets are heart-rate alignment estimates, not calibrated clock errors.



Figure 2

Where people were lost before an individual clock offset could be estimated, by step, in the development and internal assessment partitions. Bars show the share of ICU-linked people who remained at each step (blue) and who were lost at that step (ochre) in the development (A, 17,067 people) and internal assessment (B, 5,846 people) partitions; numbers give people and, in parentheses, the share of all ICU-linked people. The hatched part of the loss at the third step is people whose ICU stays were all shorter than 24 h, so that the 12 h margin could never be met; the rest of that loss is people with longer stays whose ECGs all lay within 12 h of a stay boundary. The steps are cumulative and in fixed order (Supplement S1.3). All values are complete counts; no interval is shown.

Supplementary Files

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

- [ProbstECGEHRtimealignmentsupplement.pdf](#)
- [supplementarydataS5S1populationcurve.csv](#)