

Supplementary material: temporal support, controlled-pair feasibility and administrative follow-up in MIMIC-IV ECG-EHR studies

Leo Probst · Johannes Kepler University Linz, Medical Faculty, Altenberger Strasse 69, 4040 Linz, Austria · Supplement to preprint version 1 · 16 October 2026

This supplement accompanies the technical feasibility note. Development and internal assessment denote previously inspected, person-disjoint development partitions, conventionally named train and validation. Counts from parallel designs and nested horizons must not be added. The source resource was routine care at Beth Israel Deaconess Medical Center, Boston; the ECG resource period was 2008-2019. Local audit-specific calendar distributions were not calculated. Private person-level datasets and matching lists are not supplied. Counts of persons or ECGs below 10 are shown as <10, and complementary cells are coarsened so that they cannot be recovered. Sections S1.3, S1.4, S1.9 to S1.12, S3 and S7 contain analyses added after the first analyses; they are post hoc, were specified before calculation and were run after the earlier results had been inspected. Where a number appears in both the manuscript and this supplement, the supplement gives the source table.

S1. Heart-rate alignment: definitions, recipe and flow

S1.1. Measurement and time semantics

ECG heart rate = $60,000 / \text{machine-reported mean RR interval in milliseconds}$. RR is the machine summary, not an independently remeasured beat series; the stored RR is taken from the machine measurements of the MIMIC-IV-ECG records and not measured by LoopLens. 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. Eligibility required 30-220 beats per minute and device/record timestamp discrepancy ≤ 60 s. Charted intensive-care unit (ICU) heart rate was the clinical reference (Heart Rate item 220045); duplicate observations at the same person/stay/time were reduced to their median. The origin, storage lag and rounding of this reference are described in S1.12. The source-specific date shift preserves recorded within-person differences but does not correct source clock error.

ECGs were nominally assigned to one ICU stay. At each offset from -12 to +12 h in 5-min steps (289 candidates), choose the nearest chart observation within 30 min; prefer the earlier chart time on a tie. Both target and reference must remain inside that stay. A complete ECG error curve requires a reference at every candidate. Incomplete curves do not enter the optimized error curve. Person curves are medians of absolute heart-rate differences over their complete ECG curves; the population curve is the median across person curves, giving people equal weight.

The near-optimal region contains candidates with error $\leq$ minimum +1 beat per minute. An identifiable estimate requires at least three complete ECG curves, a connected near-optimal region ≤ 120 min wide, minimum error ≤ 5 beats per minute, and no near-optimal search-boundary candidate. Positive offsets mean comparison with a later clinical reference. Exact argmin ties prefer

the smallest absolute offset, then the smaller signed offset. The specified criteria were median offset within ± 15 min (met: 0 min in each partition) and $\geq 90\%$ of identifiable offsets within ± 60 min (unmet). It is an internal alignment criterion, not a clinical norm or physical clock-error tolerance established externally.

S1.2. Person flow

The following text diagram is the complete reported person flow from nominal ICU linkage. Entries are development/internal assessment.

Flow S1. Person flow from nominal ICU linkage to identifiable offsets, development / internal assessment.

- People with ECG nominally inside an ICU stay: 17,067 / 5,846
 - No complete error curve: 10,978 / 3,784
 - At least one complete curve: 6,089 / 2,062
 - Fewer than three complete ECG curves: 4,607 / 1,565
 - At least three complete ECG curves: 1,482 / 497
 - Identifiability criteria unmet: 1,090 / 352
 - Identifiable: 392 / 145
 - Estimated offset outside ± 60 min: 129 / 43
 - Estimated offset within ± 60 min: 263 / 102

Exclusive losses in this diagram are arithmetic differences between published aggregate stages, not newly adjudicated reasons. The last stage means an identifiable alignment estimate within the window; the others cannot be classified as having inaccurate clocks. Complete curves comprised 13,201/4,419 ECGs, a different unit from people. The within-window shares among identifiable people were:

Table S1. Original alignment flows and window counts.

Window	Development	Internal assessment	Uncertainty
± 15 min	181/392 (46.2%)	69/145 (47.6%)	No interval calculated
± 60 min	263/392 (67.1%)	102/145 (70.3%)	95% percentile CI 62.2-71.6% / 62.4-77.3%
± 360 min	345/392 (88.0%)	132/145 (91.0%)	No interval calculated

The uncertainty recipe used 1,000 person bootstrap resamples per partition, seed 110. Repeated ECGs remain within person clusters. These intervals condition on the stored point estimates and do not measure individual clock uncertainty or re-estimate identifiability curves. The population curve has no computed confidence band. Its minimum at 0 min in both partitions does not certify physical simultaneity. Figure 1 of the manuscript plots these aggregate results; it does not add a new estimate.

S1.3. Structural losses before offset estimation (post hoc)

The decomposition below follows people through fixed stages. Stage L1 is the person flow's starting point; each later stage requires the previous one. Counts are people; percentages are shares of L1.

Table S2. Exclusive loss decomposition in fixed order.

Stage	Development	% of L1	Internal assessment	% of L1
L1 ECG nominally inside a unique ICU stay	17,067	100	5,846	100
L2 at least one RR-eligible ICU ECG	16,884	98.9	5,770	98.7
L3 at least one ECG at least 12 h from both stay boundaries	8,619	50.5	2,882	49.3
L4 at least one ECG with a reference at every one of the 289 offsets	6,089	35.7	2,062	35.3
L5 at least three complete ECGs	1,482	8.7	497	8.5
L6 identifiable	392	2.3	145	2.5

Losses by stage, with reasons (development; internal assessment):

- L1 to L2: 183 (1.1%) and 76 (1.3%) without an RR-eligible ECG.
- L2 to L3, stay length and boundary position: 8,265 (48.4%) and 2,888 (49.4%). Of these, 1,709 (10.0%) and 589 (10.1%) had only stays shorter than 24 h, in which the window never fits, and 6,556 (38.4%) and 2,299 (39.3%) had longer stays but every ECG within 12 h of a boundary.
- L3 to L4, reference gaps: 2,530 (14.8%) and 820 (14.0%), practically all with charted values in the stay but a gap over 30 min at one or more offsets; 1,898 and 620 of them covered at least 90% of the offsets.
- L4 to L5, number of complete ECGs: 4,607 (27.0%) and 1,565 (26.8%), of whom 3,372 and 1,152 had exactly one complete ECG and 1,235 and 413 exactly two; 889 and 309 had at least three boundary-free ECGs but failed on reference gaps.
- L5 to L6, curve criteria: 1,090 (6.4%) and 352 (6.0%). In fixed order, minimum above 5 beats per minute removed 80 and 32, zone wider than 120 min 893 and 286, and a non-contiguous zone or one at the search boundary 117 and 34. Taken singly, the criteria removed 80 and 32 (minimum), 922 and 297 (width), 1,019 and 326 (non-contiguous) and 105 and 29 (boundary). The median zone width of people who were not identifiable was 435 and 460 min.

Structurally unavailable people (L1 to L5) numbered 15,585 (91.3%) and 5,349 (91.5%). Failure of the curve criteria itself affected 6.4% and 6.0% of ICU-linked people. The conditional yield of identifiability (Wilson 95% intervals) was:

Table S3. Conditional identifiability yield by denominator.

Quotient	Development	Internal assessment
identifiable / at least three complete ECGs	392/1,482 = 26.5% (24.3-28.8)	145/497 = 29.2% (25.4-33.3)
identifiable / at least one complete ECG	6.4% (5.9-7.1)	7.0% (6.0-8.2)
identifiable / at least one boundary-free ECG	4.5% (4.1-5.0)	5.0% (4.3-5.9)
identifiable / ICU-linked people	2.3% (2.1-2.5)	2.5% (2.1-2.9)

S1.4. People with and without three complete ECGs

Included means at least three complete ECGs; excluded means the other ICU-linked people. Stay length, ECG count and heart-rate variability are linked to the selection by construction, so their differences are descriptive and not confounder findings. SMD is the standardized mean difference (development; internal assessment).

Table S4. Included against excluded people.

Characteristic	Included (median [Q1; Q3] or share)	Excluded	SMD
Age at first ICU ECG, years	68 [56; 77]	67 [56; 79]	-0.03; -0.09
Male sex (administrative field)	59.3%	58.6%	0.01; 0.00
Longest ICU stay, days	10.9 [6.3; 18.1]	2.3 [1.3; 4.4]	1.14; 1.07
ICU-linked ECGs, number	7 [5; 10]	1 [1; 2]	1.58; 1.73
SD of ECG heart rate, beats per minute	17.5 [11.4; 24.1]	10.7 [5.2; 18.8]	0.50; 0.46
Mean ECG heart rate, beats per minute	91.3 [79.7; 102.7]	82.1 [69.5; 96.4]	0.43; 0.43
In-hospital mortality of the linked admission	27.4%	12.2%	0.39; 0.39
First ICU unit: cardiac vascular (CVICU)	12.1%	27.6%	-0.40
First ICU unit: medical (MICU)	34.3%	19.9%	0.33

Among people with at least three complete ECGs, identifiable and non-identifiable people had similar age, sex, stay length and mortality (absolute SMD up to 0.14; mortality 0.09 and -0.03). Identifiable people had higher heart-rate variability (SD 20.1 against 16.1 beats per minute; SMD 0.39 in development and 0.23 in internal assessment) and a slightly higher mean rate (SMD 0.34 and 0.09), in line with a physiology-dependent measure.

S1.5. Stay pooling and half-sample flow

Stay curves instead aggregate complete ECG error curves within a stay. Identifiability criteria are retained. The specified joint requirements were: identifiability in $\geq 50\%$ of stays with at least three RR-eligible ECGs; estimated offsets agreeing within 30 min between chronological halves in $\geq 80\%$

of identified stays; and $\geq 90\%$ of identified offsets within ± 60 min. Halves may have at least one complete ECG each, but must pass the curve-shape/error criteria. Unestimable halves are absence of agreement evidence, not measured discordance.

Flow S2. Stay-level flow from stays with at least three RR-eligible ECGs to agreement of chronological halves, development / internal assessment.

- Stays with at least three RR-eligible ECGs: 4,467 / 1,477
 - Identifiability unmet or unavailable: 4,117 / 1,348
 - Identifiable stays: 350 / 129
 - Halves not both estimable: 274 / 105
 - Both halves estimable: 76 / 24
 - Estimated offsets differ by more than 30 min: 44 / >14
 - Agreement within 30 min: 32 / <10

Within ± 60 min among all identifiable stays: 228/350 / 84/129

Identifiability was 7.8%/8.7%; half-agreement evidence on the criterion denominator was 9.1%/<7%, compared with 42.1%/<40% among estimable halves only. Within-window proportions were 65.1%/65.1%. All specified criteria failed. Stay uncertainty used 1,000 person-cluster resamples per partition, seed 115: all eligible stays of a resampled person move together. Person-cluster 95% intervals were 7.0-8.6% and 7.3-10.4% for stay identifiability; 6.1-12.1% for half-agreement evidence in development (the internal-assessment interval is not shown because it rests on fewer than 10 stays); and 60.3-70.0% and 56.8-73.0% for conditional ± 60 min shares. The stricter split requiring at least three ECGs per half yielded 18 and fewer than 10 estimable stays, with fewer than 10 agreeing in each partition, among 95 and 37 identified stays with at least six ECGs. Person and stay denominators are not interchangeable.

S1.6. Consistency check and permutation diagnostic

Recipes were fixed after inspection of earlier results and before these computations. Shifting ECG times while the heart rates stay unchanged moves the objective function mechanically; the check therefore tests the implementation, boundary behavior and coverage, and says nothing about the ability to find an unknown offset (S1.9). ECG timestamps were shifted by 0, ± 30 , ± 60 , ± 120 and ± 240 min; chart times, nominal stay boundaries and the original complete-coverage/identifiability rules were retained. The input population had at least three complete baseline ECG curves (1,482 and 497 people). Expected shifted estimates equal baseline estimate minus the injected shift. Bias is the mean signed difference from that expected estimate; recovery shares and zone inclusion are conditional on identifiability at both settings. A near-optimal zone is not a confidence interval. Zero shift is a reproduction check, not known correct clock truth. Consistency success was de-

defined as median absolute differential error ≤ 15 min and $\geq 90\%$ within 60 min in each shift/partition; every variant met that operational rule, which does not calibrate absolute clock error.

Table S5. Consistency check: recovery of injected shifts relative to unshifted fitted offsets.

Partition	Injected min	Baseline ≥ 3	Shifted ≥ 3	Shifted identifiable	Both identifiable	Bias min	Median absolute error min	Within 15 min	Within 60 min	Zone contains expected
Development	0	1,482	1,482	392	392	0.0	0.0	100.0%	100.0%	100.0%
Development	-30	1,482	1,448	≥ 370	369	-3.4	0.0	89.2%	$\geq 97\%$	$\geq 97\%$
Development	30	1,482	1,470	393	376	7.4	0.0	84.8%	$\geq 97\%$	$\geq 97\%$
Development	-60	1,482	1,413	381	350	-5.6	0.0	85.7%	$\geq 97\%$	$\geq 97\%$
Development	60	1,482	1,452	384	356	13.3	0.0	79.2%	$\geq 97\%$	$\geq 97\%$
Development	-120	1,482	1,351	360	313	-11.2	0.0	80.2%	96.8%	95.8%
Development	120	1,482	1,432	364	320	19.8	0.0	74.7%	$\geq 97\%$	$\geq 97\%$
Development	-240	1,482	1,282	370	266	-20.8	0.0	78.2%	96.2%	94.4%
Development	240	1,482	1,382	366	280	28.7	5.0	67.1%	96.1%	95.4%
Internal assessment	0	497	497	145	145	0.0	0.0	100.0%	100.0%	100.0%
Internal assessment	-30	497	483	141	≥ 136	-6.3	0.0	83%	$\geq 93\%$	$\geq 93\%$
Internal assessment	30	497	≥ 488	146	≥ 136	6.6	0.0	84%	$\geq 93\%$	$\geq 93\%$
Internal assessment	-60	497	475	≥ 135	134	-8.3	0.0	79.1%	$\geq 93\%$	$\geq 93\%$
Internal assessment	60	497	≥ 488	145	132	7.6	0.0	83.3%	$\geq 93\%$	$\geq 93\%$
Internal assessment	-120	497	459	133	122	1.7	0.0	73.8%	$\geq 92\%$	$\geq 92\%$
Internal assessment	120	497	476	134	118	18.6	5.0	72.9%	$\geq 92\%$	$\geq 92\%$
Internal assessment	-240	497	424	122	94	-10.5	0.0	73.4%	$\geq 90\%$	$\geq 90\%$
Internal assessment	240	497	459	125	98	19.6	5.0	72.4%	$\geq 90\%$	$\geq 90\%$

Cells marked $\geq$ in Table S5 are lower bounds that keep fewer than 10 people inside a loss or outside a recovery window from being recovered; two within-15-min shares are rounded to whole percentages for the same reason.

Twenty permutations (seed 181001) randomly reassigned ECG-derived heart rates to ECG times within each person. Chart times and physiological series remained unchanged. This breaks the within-person ECG-time/heart-rate association while retaining person-level heart-rate distributions. Physiological stability and autocorrelation limit the null: these are diagnostics, not calibrat-

ed false-positive probabilities or hypothesis tests. This initial 20-permutation diagnostic is superseded by the 200-permutation and circular-shift nulls in S1.10.

Table S6. Within-person permutation diagnostic (initial, 20 permutations).

Partition	Baseline ≥ 3 people	Permutations	Identifiable count range	Within ± 60 among identifiable range
Development	1,482	20	287-351	16.4-23.7%
Internal assessment	497	20	94-128	12.5-25.4%

S1.7. Ties, uncertainty zones and effective resolution

Exact argmin ties occurred in 324 of 392 and 121 of 145 identifiable people. The tie-midpoint rule uses the midpoint of the first and last exact argmin candidate, even if that midpoint is not itself a minimizer. Near-optimal-zone endpoints/midpoint are separate sensitivities. Excluding exact ties changes the denominator. Whole-zone and overlap rules classify the original near-optimal zone; neither is a probabilistic confidence statement.

Table S7. Argmin and near-optimal-zone sensitivities.

Partition	Rule	People	Median min	Within ± 120 count	Within ± 360 count	Within ± 60 count	Within ± 60 share
Development	offset	392	0.0	296	345	263	67.1%
Development	tie mid	392	7.5	287	343	250	63.8%
Development	tie exclude	68	5.0	n/a	≥ 59	45	66.2%
Development	near low	392	-15.0	291	344	248	63.3%
Development	near high	392	25.0	289	342	n/a	n/a
Development	near mid	392	7.5	288	343	251	64.0%
Development	whole near zone inside60	392	n/a	n/a	n/a	215	54.8%
Development	near zone overlaps60	392	n/a	n/a	n/a	n/a	n/a
Internal assessment	offset	145	0.0	116	132	102	70.3%
Internal assessment	tie mid	145	2.5	115	132	99	68.3%
Internal assessment	tie exclude	24	5.0	≥ 15	≥ 15	≥ 15	$\geq 62\%$
Internal assessment	near low	145	-15.0	115	131	95	65.5%
Internal assessment	near high	145	30.0	113	131	n/a	n/a
Internal assessment	near mid	145	7.5	115	132	99	68.3%
Internal assessment	whole near zone inside60	145	n/a	n/a	n/a	83	57.2%
Internal assessment	near zone overlaps60	145	n/a	n/a	n/a	n/a	n/a

Near-optimal widths were median 35 min (interquartile range 15-55) and 45 min (20-55). The ± 15 min share is finer than this typical region. In Table S7, n/a in the tie-exclusion row of development and in the ± 60 min cells of the overlap rule and of the near-high rule marks counts that are withheld because their difference from a neighbouring count (the same people within ± 60 against ± 120 min, or a nested rule) would be a number of people below 10. The overlap count follows from the near-low, near-high and whole-zone counts (a zone no wider than 120 min that overlaps the ± 60 min window has an end inside it), so the near-high count is withheld together with it. Me-

dian-reduced chart observations numbered 2,941,325 and 1,001,797; 2,564,374 (87.2%) and 873,796 (87.2%) had minute zero, and all had second zero. These counts concern chart observations of the nominal ICU cohort, not raw documentation acts; minute zero alone does not establish timestamp rounding. The ECG heart-rate standard deviation was calculated across baseline complete recordings, and tertile cutoffs were partition-specific. Strata are descriptive and conditioned on at least three complete curves; no rhythm/pacing classification was inferred from RR values.

Table S8. Identifiability by physiology and repeated-recording availability.

Partition	Stratifier	Group	People	Identifiable	Share
Development	ECG heart-rate SD	lower	494	71	14.4%
Development	ECG heart-rate SD	middle	494	154	31.2%
Development	ECG heart-rate SD	upper tertile	494	167	33.8%
Development	Complete ECG count	3	607	135	22.2%
Development	Complete ECG count	4-5	483	133	27.5%
Development	Complete ECG count	≥ 6	392	124	31.6%
Internal assessment	ECG heart-rate SD	lower	166	36	21.7%
Internal assessment	ECG heart-rate SD	middle	165	50	30.3%
Internal assessment	ECG heart-rate SD	upper tertile	166	59	35.5%
Internal assessment	Complete ECG count	3	208	51	24.5%
Internal assessment	Complete ECG count	4-5	160	48	30.0%
Internal assessment	Complete ECG count	≥ 6	129	46	35.7%

Tertile boundaries for ECG heart-rate SD were 9.42 and 17.99 beats per minute in development and 9.26 and 17.58 in internal assessment.

S1.8. Aggregate source data and descriptive heterogeneity

Supplementary Data S1 (CSV, with an .xlsx copy of identical content) contains all 289 offset/error pairs per partition behind Figure 1A, with the complete-curve person denominators. Selected points show the width of the population minimum; no curve confidence band was computed.

Table S9. Selected points of the person-level population curve.

Partition	Offset min	Population absolute heart-rate error, beats per minute
Development	-720	10.5
Development	-360	8.1
Development	-240	7.5
Development	-120	6.1
Development	-60	5.0
Development	-30	4.0
Development	-15	3.1
Development	0	3.0
Development	15	3.0
Development	30	3.5
Development	60	4.5
Development	120	6.0
Development	240	7.0
Development	360	8.0
Development	720	9.9
Internal assessment	-720	10.0
Internal assessment	-360	8.1
Internal assessment	-240	7.5
Internal assessment	-120	6.0
Internal assessment	-60	5.0
Internal assessment	-30	4.0
Internal assessment	-15	3.5
Internal assessment	0	3.0
Internal assessment	15	3.0
Internal assessment	30	3.5
Internal assessment	60	4.9
Internal assessment	120	6.0
Internal assessment	240	7.0
Internal assessment	360	8.0
Internal assessment	720	9.1

Table S10. Descriptive anchor-year, first-unit and device-label strata; small groups pooled.

Partition	Dimension	Group	Complete-curve people	Identifiable	Conditional ± 60 share
Development	Anchor-year group	2011 - 2013	1,552	84	70.2%
Development	Anchor-year group	2008 - 2010	2,151	146	67.8%
Development	Anchor-year group	2014 - 2016	1,511	105	65.7%
Development	Anchor-year group	2017 - 2019	875	57	63.2%
Internal assessment	Anchor-year group	2008 - 2010	718	36	$\geq 75\%$
Internal assessment	Anchor-year group	2014 - 2016	486	37	70.3%
Internal assessment	Anchor-year group	2011 - 2013	537	37	67.6%
Internal assessment	Anchor-year group	2017 - 2019	321	35	n/a
Development	First ICU unit	Medical Intensive Care Unit (MICU)	1,621	158	73.4%
Development	First ICU unit	Medical/Surgical Intensive Care Unit (MICU/SICU)	854	49	67.3%
Development	First ICU unit	Cardiac Vascular Intensive Care Unit (CVICU)	957	37	59.5%
Development	First ICU unit	Coronary Care Unit (CCU)	1,190	74	67.6%
Development	First ICU unit	Surgical Intensive Care Unit (SICU)	773	44	50.0%
Development	First ICU unit	Trauma SICU (TSICU)	500	20	$\geq 55\%$
Development	First ICU unit	Neuro units and groups with <30 complete-curve people	194	10	n/a
Internal assessment	First ICU unit	Medical Intensive Care Unit (MICU)	550	55	74.5%
Internal assessment	First ICU unit	Cardiac Vascular Intensive Care Unit (CVICU)	308	20	$\geq 55\%$
Internal assessment	First ICU unit	Coronary Care Unit (CCU)	416	26	$\geq 65\%$
Internal assessment	First ICU unit	Surgical Intensive Care Unit (SICU)	291	17	n/a
Internal assessment	First ICU unit	Medical/Surgical Intensive Care Unit (MICU/SICU)	289	14	n/a
Internal assessment	First ICU unit	Trauma SICU (TSICU) and groups with <30 complete-curve people	208	13	n/a
Development	Device label	Device label A	210	18	n/a
Development	Device label	Device label B	388	31	$\geq 70\%$
Development	Device label	Device label C	573	73	71.2%
Development	Device label	Device label D	325	34	$\geq 73\%$
Development	Device label	Device label E	230	<10	n/a
Development	Device label	Device label F	194	<10	n/a

Partition	Dimension	Group	Complete-curve people	Identifiable	Conditional ± 60 share
Development	Device label	Device label G	258	24	$\geq 62\%$
Development	Device label	Device label H	361	14	n/a
Development	Device label	Device label I	414	15	n/a
Development	Device label	Device label J	49	<10	n/a
Development	Device label	Device label K	470	24	$\geq 62\%$
Development	Device label	Device label L	147	<10	n/a
Development	Device label	Device label M	224	<10	n/a
Development	Device label	Device label N	388	18	n/a
Development	Device label	Device label O	62	<10	n/a
Development	Device label	Device label P	148	16	n/a
Development	Device label	Device label Q	281	<10	n/a
Development	Device label	Device label R	69	<10	n/a
Development	Device label	Device label S	214	17	n/a
Development	Device label	Device label T	347	19	n/a
Development	Device label	Device label U	135	14	n/a
Development	Device label	Device label V	37	<10	n/a
Development	Device label	Device label W	203	15	n/a
Development	Device label	Device label X	115	<10	n/a
Development	Device label	Device label Y	79	<10	n/a
Development	Device label	Device label Z	37	<10	n/a
Development	Device label	Labels with <30 complete-curve people	131	<10	n/a
Internal assessment	Device label	Device label X	37	<10	n/a
Internal assessment	Device label	Device label H	127	<10	n/a
Internal assessment	Device label	Device label D	128	11	n/a
Internal assessment	Device label	Device label W	77	<10	n/a
Internal assessment	Device label	Device label B	142	10	n/a
Internal assessment	Device label	Device label K	149	11	n/a
Internal assessment	Device label	Device label C	181	22	$\geq 59\%$
Internal assessment	Device label	Device label S	64	10	n/a
Internal assessment	Device label	Device label N	132	<10	n/a
Internal assessment	Device label	Device label T	130	<10	n/a
Internal assessment	Device label	Device label G	105	13	n/a
Internal assessment	Device label	Device label U	56	<10	n/a
Internal assessment	Device label	Device label A	63	<10	n/a
Internal assessment	Device label	Device label L	51	<10	n/a
Internal assessment	Device label	Device label E	72	<10	n/a
Internal assessment	Device label	Device label M	71	<10	n/a

Partition	Dimension	Group	Complete-curve people	Identifiable	Conditional ± 60 share
Internal assessment	Device label	Device label F	61	<10	n/a
Internal assessment	Device label	Device label I	118	<10	n/a
Internal assessment	Device label	Device label Q	95	<10	n/a
Internal assessment	Device label	Device label Y	32	<10	n/a
Internal assessment	Device label	Device label P	37	<10	n/a
Internal assessment	Device label	Labels with <30 complete-curve people	134	10	n/a

Strata use the most frequent label across each person's complete ECG curves; each person appears once per dimension. Labels with fewer than 30 complete-curve people are pooled, but identifiable subsets may still be small. Identifiable counts below 10 are shown as <10 without a share (n/a); first-unit groups with fewer than 10 identifiable people were pooled. Shares that would reveal fewer than 10 people inside or outside ± 60 min are shown as lower bounds or as n/a. These counts describe complete-curve availability, not all ICU people. Device labels do not establish manufacturer or physical recorder identity; first ICU unit is not necessarily the location at ECG acquisition. Anchor-year groups and shifted dates do not recover a real device calendar. Cart-by-shifted-pseudo-month pooling yielded conditional ± 60 shares of 64.6% and 62.0%; it cannot validate true device-month stability.

The coarse admission-boundary check compared nominal ECG time with the stored assigned hospital admission. Among 243,432 and 83,347 linked ECGs, 65,799 and 22,638 were outside that interval; none lacked admission boundaries. This is a context-link check, not a clock-error rate: contextual admission assignment and outpatient recordings can contribute. The ICU-eligible audit itself had zero outside-stay timestamps by construction (42,235 and 14,233 ECGs), so it is not an independent validation of stay timing.

S1.9. Simulation of known offsets (post hoc)

The consistency check in S1.6 shifts ECG times while the heart rates stay unchanged, which moves the objective function mechanically. To ask whether the procedure can find an unknown offset, the reference side was kept real (charted values, ECG times, stay boundaries, full-coverage rule) and only the ECG heart rate was generated: a latent rate at time $t + \tau$ plus noise, estimated with the unchanged procedure. The population was the 1,979 people (1,482 and 497) with at least three complete ECGs and 17,620 complete ECGs. Offsets τ were 0, ± 30 , ± 60 , ± 120 and ± 240 min. The estimator should return $\tau + d$, where d is a documentation delay (a charted value with chart time c is taken to originate at $c - d$).

- Model A (optimistic bound): linear interpolation of the charted values shifted by d .

- **Model B (main model):** a latent Ornstein-Uhlenbeck process whose autocorrelation was estimated from the real chart series of the 1,979 people (robust variogram from the median of absolute differences, because the curve criteria are median-based): nugget 18.0, variance 125.9, time constant 25.6 h; the variogram rises from 4.4 (3.5 min lag) through 17.6 (60 min) and 39.6 (3.5 to 5.5 h) to 110 (33 h). Given the two neighbouring charted values, the latent rate at ECG times was drawn from the Markov bridge; draws of one person's ECGs are conditionally independent given the chart points.
- **Noise:** white noise on the ECG rate, calibrated so that the median per-person curve minimum at $\tau = 0$ and $d = 0$ matches the real data (1.08 beats per minute; 94.3% of real people have a minimum of at most 5): 1.20 (model B), 1.59 (model A) and 0.74 (model B with person-scaled noise). Sensitivity at half and double noise.
- **Delay:** $d = 0, 5$ and 15 min, and the empirical storage lag (storage time minus chart time; median 23 min) as an upper limit, because chart time is not certainly the measurement time minus the storage lag.
- **Replications:** 20 per scenario (10 for the empirical delay and the sensitivities); seeds from `SeedSequence([242, scenario, replication])`; replicate spread of the identifiable share 0.3 to 0.6 percentage points.

Table S11. Recovery of known offsets, model B, both partitions pooled, delay 0 (shares among 1,979 people or among identifiable estimates).

tau, min	Identifiable	Within 15 min of tau	Within 60 min of tau	Median estimate minus tau, min	Estimates within ± 60 min of zero
0	32.3%	73.7%	86.3%	0	86.3%
+30 / -30	30.1% / 30.2%	47.8% / 46.3%	85.5% / 84.5%	-10 / +10	83.9% / 82.3%
+60 / -60	29.4% / 27.5%	46.7% / 43.5%	86.9% / 85.2%	-10 / +15	68.6% / 67.6%
+120 / -120	27.6% / 26.2%	45.5% / 43.3%	84.9% / 85.4%	-15 / +15	5.8% / 4.6%
+240 / -240	26.1% / 24.3%	45.3% / 43.3%	85.2% / 84.8%	-15 / +15	2.1% / 1.5%

Reading. Under idealized generation, the true offset is found within 60 min for 85% to 87% of identifiable people and within 15 min for 43% to 48% (74% at $\tau = 0$); zone width and hour-rounded charting limit the resolution. The estimate shrinks toward zero by 10 to 15 min because of the tie rule and wide zones. With a constant delay of 15 min and $\tau = 0$, the median estimate is only about 5 min instead of 15 min, so a systematic delay of up to about 15 min stays close to the nominal population minimum. The identifiable share falls from 32% to 24% to 26% with growing $|\tau|$ because zones reach the search boundary. Model A, person-scaled noise and half noise give the same picture (identifiable at $\tau = 0$: 31.7%, 32.8%, 34.3%; within 60 min: 86.8%, 86.5%, 89.4%); double noise lowers these to 29.7% and 78.4%. The empirical delay (upper limit) lowers the share within 15 min to 28%, within 60 min to 75%, and the identifiable share to 25.4%. The simulated yield at $\tau = 0$ (32%) exceeds the observed one (27%): real data are less favorable than the idealized model. Limits: a constant offset (no drift or device change), the chart treated as an interpo-

lated truth, noise calibrated to the curve minimum and not to the whole error distribution, and no rhythm disturbance, pulse deficit, pacing or treatment effect.

S1.10. Null constructions, mixture and predictive value (post hoc)

Two null constructions weaken the alignment between ECG rates and chart series; the decoupling is not complete. Within-person permutation reassigns the ECG-derived rates within the person across all usable ECGs (200 replications, seed family 242). Circular shift treats each stay's chart series as a circle of length span plus mean spacing and replaces each value by that of the nearest original point at the circularly shifted time, independently per stay; ECG times, chart times and coverage are unchanged and only the chart values move. N1 draws the shift uniformly between $\min(24 \text{ h}, \text{span}/4)$ and span minus that value (500 replications), N2 requires at least $\min(72 \text{ h}, \text{span}/4)$ (500), and N3 (diagnostic, 300) shifts by whole multiples of 24 h of at least 2 days, which keeps the daily phase (seed family 243).

Table S12. Observed and null identifiability among 1,979 people with at least three complete ECGs (mean and 2.5 to 97.5 percentiles).

Condition	Replications	Identifiable	Share	Identifiable within ± 60 min	Identifiable beyond ± 60 min
Observed	n/a	537	27.1%	365	172
Within-person permutation	200	435.8 (403.9-473.1)	22.0%	87.0 (70-103)	348.8 (321-379)
N1, circular shift of at least $\min(24 \text{ h}, \text{span}/4)$	500	375.5 (347-408)	19.0%	36.1 (25-49)	339.4 (310-370)
N2, circular shift of at least $\min(72 \text{ h}, \text{span}/4)$	500	364.6 (332.5-396)	18.4%	35.4 (25-47)	329.3 (297.5-359.5)
N3, multiples of 24 h (diagnostic)	300	375.7 (347.5-404)	19.0%	29.1 (20-38)	346.6 (318-374.6)

No replication of any null reached the observed number of identifiable people (highest values N1 420, N2 412, N3 435), the observed number within ± 60 min (highest 55, 53, 41) or fell below the observed number beyond ± 60 min (lowest 294, 278, 301). One-sided empirical p values were at most 0.002 (N1, N2), 0.0033 (N3) and 0.005 (permutation). By partition, identifiable people numbered 392 (development) against 278.8 (254-305) under N1 and 145 against 96.6 (81.5-112) in internal assessment; within ± 60 min 263 against 26.5 (18-37) and 102 against 9.6 (4-16). The permutation null gave 324.3 (298-360) and 111.4 (93-126) identifiable people, and 65.2 (50-80) and 21.8 (15-30) within ± 60 min. Preserving the daily phase (N3) did not raise the number within ± 60 min, so a daily rhythm alone does not create the small-offset excess.

Mixture reading (post hoc; not used in the conclusions of the manuscript). With $a = (\text{observed share} - \text{null share}) / (1 - \text{null share})$, the excess over the null can be written as a share: 10.1% (8.2-11.6%) under N1, 10.7% (8.9-12.4%) under N2, 10.1% (8.4-11.6%) under N3 and 6.6% (4.2-8.5%) under permutation. Reading a as the share of people who carry a recoverable time signal rests on

three assumptions: (1) the population is a mixture of people with and without such a signal; (2) people without a signal meet the criteria at the null rate; (3) every person with a signal meets the criteria. Assumption (3) is contradicted by the simulation in S1.9: even under idealized generation only 24% to 32% of people met the criteria (32.3% at offset 0), so the criteria detect only part of the people with a signal. Assumption (2) is uncertain too, because the null constructions decouple incompletely. The quantity a is therefore not an estimate of the share of people with a time signal; it is the excess share of people meeting the criteria over the null rate. A three-class reading (within 60 min, beyond 60 min, not identifiable) against the idealized simulation and the permutation null, which placed 60.2% (53.3-67.1%) of people as idealized-like (61.1% with model A), rests on the same kind of assumptions and is likewise not a prevalence.

Predictive value. The flag “identifiable and estimated offset beyond ± 60 min” had a sensitivity of 25.1% for true offsets of at least 120 min (model B), and false-positive rates of 4.9% in simulated aligned clocks (τ 0 to ± 30) and 17.6% under the permutation null.

Table S13. Positive predictive value of the flag by prevalence of large true offsets.

Prevalence	PPV, false-positive rate from simulation (4.9%)	PPV, false-positive rate from permutation (17.6%)
1%	5.0%	1.4%
5%	21.4%	7.0%
20%	56.3%	26.3%

Limits. The permutation null keeps each person’s variability but destroys the time structure of the ECG rates. The circular shift keeps autocorrelation, but for stays shorter than 4 days the shift is limited to $\text{span}/4$ and decoupling is weaker; N2 differs little from N1. The null affects only the criteria evaluated here and says nothing about individual clocks.

S1.11. Individual algorithmic support of the offset (post hoc)

For each of the 1,979 people, two quantities were added to the identifiability criteria. Leave-one-ECG-out stability is the largest deviation of the offset from the all-ECG estimate when each complete ECG is left out in turn (same tie rule; with three ECGs the estimate then rests on two). The permutation p value uses the person's minimum curve error as the statistic and permutes the ECG rates within the person across all usable ECGs (for at most 6 usable ECGs by exact enumeration of all $m!$ permutations, with p equal to the share of these permutations, the identity included, whose statistic is at most the observed one, so that the smallest attainable p is $1/m!$; otherwise by 1,000 random permutations, with $p = (1 + \text{number of permutations with a statistic at most the observed}) / (R + 1)$). Expected numbers without a time signal come from 20 pseudo data sets in which each person's ECG rates were permuted beforehand and the same criteria, p values and stability were computed; this includes the selection by the criteria, under which p values among the identifiable are not uniform.

Table S14. Individual support among the 537 identifiable people.

Measure	Observed	Expected without signal (mean; 2.5-97.5 percentile)
Identifiable (criteria met), of 1,979	537 (27.1%)	430.9
of these, p below 0.05	232 (43.2%)	30.7 (18.4-41.1)
of these, p below 0.01	123 (22.9%)	4.7 (2.0-9.6)
of these, stability within 15 min	187 (34.8%)	46.4 (37.5-60.1)
of these, stability within 60 min	298 (55.5%)	100.2 (79.0-121.1)
p below 0.05 and stability within 15 min	117 (21.8%)	8.8 (5.0-13.0)
and offset within ± 60 min	more than 100 (about 20%)	2.7 (0.5-6.1)

The strictest group (117) is 0.51% of the 22,913 ICU-linked people (90 of 17,067, 0.53%, and 27 of 5,846, 0.46%); with the further restriction to offsets within ± 60 min it is more than 100 people (about 0.5%). In permuted data 8.8 of the 117 people (p below 0.05 and stability within 15 min) are expected, about 7.5% of the observed group; 2.7 of the more than 100 people that also lie within ± 60 min are expected, a small minority (below 5%); for p below 0.05 alone, 30.7 of 232, about 13%. These measures show algorithmic support relative to a permutation null of unknown fidelity. They are not calibrated individual uncertainty intervals for the offset, no such interval was produced, and they do not establish clock accuracy. With Benjamini-Hochberg control at $q = 0.05$ over all 1,979 tests, 178 hypotheses were rejected, 80 of them among the identifiable.

By number of complete ECGs, 186 identifiable people had three, 181 had four or five and 170 had six or more. Among those with three, 14 had p below 0.05 and the strictest group had fewer than 10 people; among those with three to five, 38 were in the strictest group (null 4.5); among those with six or more, 135 had p below 0.05 and 79 were in the strictest group (null 4.3). The evidence is concentrated in people with six or more complete ECGs (79 of 117). For 129 people with only three usable ECGs, a p value below 0.05 cannot be reached ($3! = 6$ permutations, smallest $p = 1/6$; 31 of them identifiable), so their individual uncertainty is undetermined, not small.

Stability. Among the identifiable, the largest leave-one-out deviation was 0 min for 87 (16%), at most 5 min for 123 (22.9%) and above 120 min for 211 (39.3%); without signal 303 of 431 (70%) exceeded 120 min. Permuting only among the complete ECGs gave 185 instead of 232 people with p below 0.05 (correlation of log p values 0.91); the conclusion is unchanged. Limits: the permutation null destroys the temporal structure of the rates, p values are per person without multiplicity correction (the pseudo data sets calibrate the group counts), and stability and p concern the estimator, not the correctness of a true clock error.

S1.12. Reference origin and reference sensitivity (post hoc)

The reference was item 220045 (heart rate) of the ICU module's charted events: 3,943,122 median-reduced observations for the 22,913 ICU-linked people with a rate of 30 to 220 beats per minute (at most one row per person, stay and chart time), of which 87.2% fell on minute zero (all on second zero). Storage lagged charting by a median of 23 min (quartiles 8 and 55; 90th percentile 99; 99th percentile 255; 0.9% negative); 18.0% of values were stored within 0 to 5 min and 38.5% within 0 to 15 min. A warning flag was set for 0.38%. The tables do not code whether a value was transferred from a monitor or entered manually. Of 56,468 RR-eligible ICU ECGs, 39,929 (70.7%) carried a machine-reported sinus rhythm without pacing text and 2,403 (4.3%) carried pacing text (pace, paced, pacing, pacemaker).

Table S15. Reference variants under the unchanged full-coverage rule (development; internal assessment).

Variant	People with at least one complete ECG	At least three	Identifiable	Within ± 60 min	Null identifiable (20 permutations)
V0 baseline	6,089; 2,062	1,482; 497	392; 145	263 (67.1%); 102 (70.3%)	321.9; 108.7
Va sinus rhythm without pacing text	4,573; 1,557	883; 292	222 (25.1%); 75 (25.7%)	149 (67.1%); 61.3-73.4); 49 (65.3%; 54.7-74.7)	213.1; 72.1
Vb storage lag up to 5 min; Vb15 up to 15 min; Vc minute not zero; Vd10 and Vd15 reference-matching window 10 and 15 min; Ve1 and Ve2 combinations	0 to <10	0	0	n/a	n/a

Every restriction of the chart side ends the estimation, because full coverage requires a reference within the window at all 289 offsets (every 5 min over ± 12 h), which an hour grid satisfies only with the 30 min window. Comparable measurement conditions cannot be produced with this cohort and design. Restriction to sinus rhythm leaves the yield (25.1% and 25.7% against 26.5% and 29.2%; null 24.1% and 24.7%) and the share within ± 60 min essentially unchanged. At pair level (people with at least one complete ECG; each ECG-offset pair counted only if the chosen reference met the condition; population curve as the median over people for offsets with at least 100 people), the population minimum stays near zero:

Table S16. Pair-level restrictions, development.

Variant	Valid pairs	Minimum at, min	Value at 0	At +60	At -60
V0 baseline	100%	0	3.00	4.54	5.02
Va sinus without pacing	66%	5	2.91	4.00	4.49
Vb5 lag 0 to 5 min	17%	0	2.97	4.94	5.02
Vb15 lag 0 to 15 min	38%	0	2.99	4.85	5.03
Vc minute not zero	4.1%	5	3.89	6.16	6.99
Vd10 distance up to 10 min	37%	5	2.85	4.98	5.08
Vd15 distance up to 15 min	54%	5	2.95	4.95	5.09
Ve1 sinus, lag up to 15 min, distance up to 15 min	13%	5	2.03	4.05	4.52
Ve2 minute not zero, lag up to 15 min, distance up to 15 min	1.5%	5	3.21	7.11	7.09

Values are median absolute differences in beats per minute. Internal assessment shows the same pattern (minimum at 0 to 10 min in all variants; Ve2 with only 13 evaluable offsets). The minimum does not disappear when delayed or rounded times are excluded, and it still says nothing about individual clocks. In one partition, Vb15, Vc and Vd each had fewer than 10 people with a complete curve; these are very few cases.

S2. Serial-pair feasibility: implemented rules and design-specific calipers

S2.1. Control of switching ECG recordings

Each opportunity consists of two distinct, consecutive ECG recordings from the same person, nominally within the same hospital stay and 2-72 h apart. A control opportunity likewise consists of two distinct recordings from a control person with a small measured change and comparable interval. Threshold-based change groups can select measurement error and regression to the mean; effect estimates would require repeated measurements and treatment context. Segments

of one recording do not supply this control. Different hospital admissions are a separate sensitivity, not a requirement for the primary control. Change and control opportunities are connected between people for matching; persons in a matched set are different. A completed biological analysis would compare their pair differences and account for person clustering. No such effect or difference of pair differences is estimated here, and no error band from another population is transferred.

S2.2. Potassium design

Nearest eligible chemistry potassium measurement within $\pm w$ h of each ECG, $w=2$ or 6 , with same item within a pair, distinct nonmissing specimens, nominal temporal ordering consistent with the ECG order, usable unrounded numerical values, known units and no reported hemolysis. Laboratory ties prefer the pre-ECG side, then the documented time and deterministic source-record order. The documented laboratory event time is used; storage time does not substitute for collection or event time. Unknown hemolysis status is not proven absence. Both potassium values must lie in 2.5-6.0 mmol/L. Change = $|\Delta K| \geq 1.0$ mmol/L; control = $|\Delta K| \leq 0.2$ mmol/L. Intermediate changes are neither arm. The archived recipe normalizes mmol/L and mEq/L equivalently for potassium; chemistry and whole-blood sources are not interchanged. Chemistry items were 50971/52610; matching is item-specific.

The generous bound admits all available supported episodes, excludes self-matching and known within-pair device-label changes, and requires stored global quality ≥ 0.6 and heart rate 30-180 beats per minute on both ECGs. Unknown device information does not establish a stable physical recorder. Persons with both kinds of opportunities remain available for this upper-limit calculation; a fully eligible disjoint first-opportunity selection can only reduce that bound.

Table S17. Implemented matching rules.

Implemented matching feature	Rule for both laboratory windows
Partition	Exact; no cross-partition matching
Potassium item	Exact
Administrative sex category	Exact, including an explicit unknown category
Laboratory timing side	Exact before/after class separately at each state
ECG separation class	Exact [2,6], [6,24], [24,72] h
Ratio of ECG separations	0.8-1.25
Baseline potassium difference	≤ 0.3 mmol/L
Baseline heart-rate difference	≤ 10 beats per minute
Age difference	≤ 10 years

Construct a bipartite person graph with an edge if supported change/control episodes satisfy the implemented rules. Maximum cardinality, one-to-one without replacement, was checked independently with a network-flow calculation. Stored private graph matches are not a clinically finalized matching sample. Minimum-distance optimization, lexicographic final selection, balance/standardized mean difference (SMD) and model-rank checks were not completed for a fully eligible effect sample. No marker outcome was a matching feature.

The full intended eligibility additionally requires adequate treatment-interval observation, baseline medication/dialysis context, rhythm, complete measurement eligibility and a first qualifying opportunity with mutually exclusive change/control persons. These remain unresolved. Absence of an administration entry and a stay-level electronic medication administration record (eMAR) presence flag do not certify untreated status or interval coverage. Earlier complete accompanying-laboratory calipers were removed from primary matching by a documented pre-effect revision; accompanying laboratories became covariates with train-bound imputation/missingness, while treatment/rhythm and the central potassium reference cannot be imputed. The old accompanying-laboratory variant remains a sensitivity, not the reported bound.

Flow S3. Potassium design: change-pair pool and maximum person-matched sets at each laboratory window, development / internal assessment.

- Potassium ± 2 h: change-pair pool 615 / 234
 - maximum person-matched sets ≤ 41 / < 10
 - fully eligible clinical sample unknown; no effect fit
- Potassium ± 6 h: separate change-pair pool 2,371 / 826
 - maximum person-matched sets ≤ 352 / ≤ 69
 - fully eligible clinical sample unknown; no effect fit

These are successive objects with different units, not a single exclusive person-exclusion cascade. Availability between the initial pair pool and the matched sets is not fully decomposed into exclusive reasons. The wider window changes the candidate pool; it is not rematching the same ± 2 h pairs. The internal planning level (not a statistical feasibility criterion) was 100 matched sets per partition. The internal-assessment upper limit was below that level at both windows; additional full eligibility cannot increase the limit. Direction-specific planning levels were also unmet in the wider-window internal assessment (at most 29 potassium-decrease sets versus 30). Direction-specific maxima share controls and are not additive.

S2.3. Calcium/heart-rate proxy design

This is a screen for the heart-rate-change layer of a QT-correction feasibility design; it does not constitute a QT-formula comparison. Nearest eligible total-calcium samples within ± 6 h, same item within pair, distinct specimens, consistent nominal ordering, calcium 1.75-3.00 mmol/L. Known mg/dL calcium values are divided by 4.008. Calcium change ≥ 0.20 mmol/L; small calcium change ≤ 0.05 mmol/L. Stored heart rate must be 40-120 beats per minute; the heart-rate-change layer requires $|\Delta HR| \geq 15$ beats per minute and controls $|\Delta HR| \leq 5$ beats per minute with a small calcium change. Quality ≥ 0.6 , stored QRS < 120 ms and $|\Delta QRS| \leq 10$ ms on both recordings; known device-label change excluded.

Table S18. Implemented matching rules.

Implemented matching feature	Proxy screen rule
Partition, calcium item, administrative sex	Exact, unknown sex category explicit
Laboratory timing side at each state	Exact before/after class
ECG separation class	Exact [2,6), [6,24), [24,72] h
ECG separation ratio	0.8-1.25
Age difference	≤ 10 years
Baseline proxy heart-rate difference	≤ 10 beats per minute
Baseline calcium difference	≤ 0.10 mmol/L
Baseline proxy QRS difference	≤ 10 ms

First provisional opportunities were matched on person nodes, one-to-one without replacement, with control persons disjoint from change persons. Maximum cardinality and independent network flow agreed. Exact care-setting/exposure/dialysis attributes and complete treatment coverage were not available; no successful balance or distance-minimum claim is made. The previously planned albumin caliper had been removed by revision, not silently omitted. A small change in total calcium after removal of the albumin caliper does not indicate a stable electrophysiologically relevant calcium state; albumin, pH and ionized calcium remain open requirements, and the proxy counts are technical selection counts only.

Flow S4. Calcium and heart-rate proxy design from the nominal screen to matched proxy sets, development / internal assessment.

- Heart-rate-change layer after nominal calcium/heart-rate screen: 2,062 / 697 pairs
 - after quality/QRS/device-label screen: 355 / 122 pairs
 - first provisional change opportunities: 339 / about 110 persons
 - matched proxy sets: 79 / 11
 - identical eligible common QT/RR/early-T beat population unknown
 - no QT-formula fit or selection

The proxy heart rate and QRS are stored context values. General accepted-beat counts are not evidence for a common eligible QT/RR/T-onset/T-peak beat population. The proxy can alter case assignment, so 79/11 is not an upper limit for the intended common-beat design. The generous

all-episode proxy screen yielded 93/11 sets and is a different selection, not the primary Table 1 count of the manuscript. The internal planning level for the fully eligible design was at least 50 sets per layer and partition; falling short of it in the proxy count does not numerically prove failure of an unknown true common-beat pool.

S2.4. Troponin reporting-transition design

Flow S5. Troponin reporting-transition design and its structural exclusion.

- Potential laboratory reporting-transition opportunities
 - prerequisite: independently supported ECG-to-clinical temporal alignment
 - prerequisite unmet throughout this development implementation
 - zero eligible pairs / zero matched sets in both partitions

No numeric caliper was applied after this structural exclusion. Initial transition-pair counts for a timing-supported population are not established. This is not the prevalence of troponin changes, absence of infarction or a negative biomarker effect. Laboratory reporting transition and physiological onset must not be equated.

S2.5. Exploratory caliper sensitivities

The same marker-free supported opportunities were rematched using the original person-edge definition. Each sensitivity changes one caliper unless labeled combined. All retain exact partition, potassium item, administrative sex, laboratory timing side, distinct persons and the 0.8-1.25 ECG-interval ratio. The combined variant uses potassium ≤ 1.0 mmol/L, heart rate ≤ 15 beats per minute, age ≤ 15 years and no exact gap class. Stricter variants are also shown. Episode combinations and person-matched sets are different units; the number of distinct person edges is not shown. Independent maximum flow agreed with one-to-one bipartite matching. Counts remain upper limits with unresolved treatment/rhythm and mutually exclusive first-opportunity eligibility.

Table S19. Matching sensitivity counts at fixed nominal laboratory windows.

Window h	Partition	Variant	K caliper mmol/L	HR caliper beats per minute	Age years	Exact gap class	Episode combina- tions	One-control sets / links
2	Develop- ment	base	0.3	10	10	yes	67	41
2	Develop- ment	K02	0.2	10	10	yes	44	31
2	Develop- ment	K05	0.5	10	10	yes	101	58
2	Develop- ment	K10	1.0	10	10	yes	184	95
2	Develop- ment	HR05	0.3	5	10	yes	≥ 35	27
2	Develop- ment	HR15	0.3	15	10	yes	89	56
2	Develop- ment	age05	0.3	10	5	yes	37	27
2	Develop- ment	age15	0.3	10	15	yes	92	≥ 45
2	Develop- ment	gap relaxed	0.3	10	10	no	82	≥ 45
2	Develop- ment	combined	1.0	15	15	no	430	134
2	Develop- ment	two controls	0.3	10	10	yes	67	54
2	Internal assessment	base	0.3	10	10	yes	<10	<10
2	Internal assessment	K02	0.2	10	10	yes	<10	<10
2	Internal assessment	K05	0.5	10	10	yes	<20	<20
2	Internal assessment	K10	1.0	10	10	yes	23	<20
2	Internal assessment	HR05	0.3	5	10	yes	<10	<10
2	Internal assessment	HR15	0.3	15	10	yes	11	<20
2	Internal assessment	age05	0.3	10	5	yes	<10	<10
2	Internal assessment	age15	0.3	10	15	yes	13	<20
2	Internal assessment	gap relaxed	0.3	10	10	no	<10	<10
2	Internal assessment	combined	1.0	15	15	no	58	37

Window h	Partition	Variant	K caliper mmol/L	HR caliper beats per minute	Age years	Exact gap class	Episode combinations	One-control sets / links
2	Internal assessment	two controls	0.3	10	10	yes	<10	<10
6	Development	base	0.3	10	10	yes	1,210	352
6	Development	K02	0.2	10	10	yes	857	310
6	Development	K05	0.5	10	10	yes	1,841	425
6	Development	K10	1.0	10	10	yes	3,454	562
6	Development	HR05	0.3	5	10	yes	623	270
6	Development	HR15	0.3	15	10	yes	1,739	410
6	Development	age05	0.3	10	5	yes	672	272
6	Development	age15	0.3	10	15	yes	1,692	398
6	Development	gap relaxed	0.3	10	10	no	1,577	368
6	Development	combined	1.0	15	15	no	9,514	637
6	Development	two controls	0.3	10	10	yes	1,210	554
6	Internal assessment	base	0.3	10	10	yes	124	69
6	Internal assessment	K02	0.2	10	10	yes	81	48
6	Internal assessment	K05	0.5	10	10	yes	198	99
6	Internal assessment	K10	1.0	10	10	yes	358	136
6	Internal assessment	HR05	0.3	5	10	yes	62	47
6	Internal assessment	HR15	0.3	15	10	yes	183	85
6	Internal assessment	age05	0.3	10	5	yes	60	46
6	Internal assessment	age15	0.3	10	15	yes	169	84
6	Internal assessment	gap relaxed	0.3	10	10	no	153	≥70
6	Internal assessment	combined	1.0	15	15	no	952	180
6	Internal assessment	two controls	0.3	10	10	yes	124	93

For the two-controls variant, each change-person node has capacity two and each control-person node capacity one: the final column is links, not sets of two fully matched controls or independent change people. Baseline one-to-one sets remain 41/<10 and 352/69; two-control link counts are 54/<10 and 554/93. Sets and links below 10 appear as <10. A few cells are shown as bounds (<20 or ≥) because the exact value would make the difference from a neighbouring cell, in the same row or in a nested variant, a number below 10. Persons can occur as candidates in both graph arms in these generous upper limits; no fully disjoint clinical cohort is asserted. Baseline potassium is the strongest of the individually varied bottlenecks at ±6 h: loosening it to 1.0 mmol/L permits 562/136 one-control sets, compared with 410/85 for heart rate, 398/84 for age and 368/≥70 without exact gap class. Caliper relaxation trades comparability for count and is not a new validated effect de-

sign. The 100 and 50 set levels are internal planning rules, not derived from power or precision and not statistical feasibility criteria; development alone exceeded the 100 level for potassium at ± 6 h.

S3. AF administrative-span count

S3.1. Selection, time zero, event definition and contact

The archived selector uses the first eligible discharge, with the last eligible reported-sinus ECG and regular waveform metadata in 0-7 days before discharge. Selection proceeds in this order: usable machine metadata; machine-reported sinus rhythm (rule below) with regular waveform rhythm context; no coded AF in a prior or the index admission and no AF mentioned in the current report; discharge 0 to 7 days after the ECG; no known AF through discharge; consistent documented event and span references; positive administrative span after discharge. The first eligible discharge per person is kept (earliest discharge time; within it the latest eligible ECG, then study order). The choice of the index discharge and ECG (all selection steps up to and including no known AF through discharge) uses no marker value and no information about events or contact after discharge, and no marker value enters at any step. The last two selection steps of the risk set do use post-discharge information: consistency of a documented AF time with the span and the death date (13 index discharges in development and fewer than 10 in internal assessment were excluded) and a positive administrative span after discharge. The selection is therefore not blind to follow-up in the risk set, only in the choice of the index. There is no separately established fixed disease-free washout or complete pre-index healthcare coverage.

Reported sinus rhythm is a text rule on the machine report (case-insensitive): at least one match of “sinus rhythm”, “sinus tachycardia”, “sinus bradycardia” or “sinus arrhythmia” as whole words, and no match of “no sinus” or “no longer sinus”, of “sinus” followed within 25 characters by “has been replaced” or “not present”, or of “possible”, “probable” or “uncertain” followed within 20 characters by “sinus”.

Documented AF (atrial fibrillation or flutter) is the earliest of two sources: the admission time of the first admission with a diagnosis in the code group atrial fibrillation or flutter (ICD-9-CM 42731 and 42732; ICD-10-CM codes beginning I48), and the time of the first ECG whose machine report was classified as AF positive (the stored classification was used as is; its handling of flutter is not documented separately). Because the diagnosis source carries the admission time, not the onset, AF that began after the ECG in the same admission is classified as timing unknown at the index and is not used to exclude. These are diagnosis-linked proxies and report times, not adjudicated onset; administrative codes for atrial fibrillation have a pooled sensitivity of 80% and a positive predictive value of 88% (Yao *et al* 2019, listed in the manuscript), so a code-based label is a documentation target.

Last contact is the later of the last ECG time of the person (all ECGs of the master index) and the last hospital discharge time. Emergency-department stays, laboratory events and the date of

death are not contact sources. The administrative span runs from the index discharge to the last contact; it is not a sum of observed time, and a later contact does not make the time in between observed. No calendar data-end date was applied: the span ends at the later of the last ECG and the last hospital discharge present in the data, so truncation by the end of the data window, or by the one-year limit of death records after the last hospital discharge (Johnson *et al* 2023a, 2024), cannot be distinguished from contact end and appears as contact end. Death dates in MIMIC-IV combine hospital records with linked Massachusetts state death records; a missing death date means survival only within that one-year window. Death is a calendar date (taken as 00:00), so a death on the discharge date counts as occurring before discharge. AF takes precedence over contact end, death and horizon when its time is above zero, within the span and within the horizon. For people without AF, a coincidence of two or three ends is counted as a tie, with the minimum as the time.

The historical risk-set algorithm chooses the earliest documented AF, last contact, recorded death or horizon. A death competes with AF but was censored in that recipe. These counts do not estimate cumulative incidence or competing-risk probabilities. Time differences are computed in seconds from timestamps without fractional seconds.

S3.2. Stored selector flow (exact timestamps)

Flow S6. AF selector from fixed eligible index discharges to complete cases, development / internal assessment, exact timestamps.

- Earlier candidate ECGs/discharges: upstream count not reconstructed
 - fixed eligible index discharges (first discharge per person): 28,297 / >9,745
 - incompatible event/span reference: 13 / <10
 - compatible reference: 28,284 / >9,745
 - no positive administrative span after discharge: 10,371 / >3,660
 - recorded death date on or before the discharge date: 597 / 202
 - no contact after the index discharge: 9,774 / >3,450
 - positive administrative span (at risk): 17,913 / 6,079
 - incomplete marker/age/sex/heart-rate data: 6,304 / 2,076
 - complete cases: 11,609 / 4,003

Excluding the landmarks with a death date on or before the discharge date, 9,774 of 28,284 (34.6%) in development and more than 3,450 (about 36%) in internal assessment had no ECG or hospital discharge after the index discharge; with those deaths, 36.7% and about 38% had no positive administrative span. Emergency-department stays and laboratory events are not contact sources. The finite-positive-duration check removed no further cases after the positive-time stage. The

reproduction of the index discharge selection with the stored values gave the sealed counts (28,297 and more than 9,745). Person uniqueness and partition disjointness were checked. Of the 597 index discharges with a death date on or before the discharge date, 18 had a later ECG or admission, which points to date-level conflicts (death at 00:00 against contact on the same day). Fourteen index discharges that were incompatible under integer-hour arithmetic are compatible with exact times. Complete cases required finite values of the marker (loop width), age, administrative sex and machine heart rate.

Table S20. 365-day count with exact timestamps, before and after completeness selection.

Partition	Stage	People	Documented AF	Contact end	Death end	Horizon end	Tied end	Contact end among people with a positive span and no documented AF
Development	at risk	17,913	771	4,517	311	12,296	18	26.4%
Development	complete cases	11,609	413	2,900	177	>8,090	<20	25.9%
Internal assessment	at risk	6,079	275	1,485	113	>4,190	<10	25.6%
Internal assessment	complete cases	4,003	141	993	70	>2,780	<10	25.7%

Each row sums exactly to its person denominator. End reasons are mutually exclusive; tied ends remain separate. Counts below 10 are shown as <10, and the horizon-end cell of the same row is coarsened so that they cannot be recovered. The tied-end count of the development complete-case row is shown as <20 (horizon-end cell coarsened accordingly), because its difference from the at-risk row would otherwise be a count below 10. The denominators of the last column are the people with a positive span and no documented AF (17,142; 11,196; 5,804; 3,862). Loss through the completeness requirement was 6,304 of 17,913 (35.2%) and 2,076 of 6,079 (34.2%). Contact cessation is not verified event freedom, and no recorded AF by the horizon is not proof that AF never occurred. Pure finite counts have no inferential interval, but ascertainment error and generalizability remain uncertain.

The earlier AF design selected the first master ECG before checking discharge eligibility. The revised design selected the first eligible discharge and its ECG. That post-inspection change alters the target population and is not a replication. Earlier AF work used a historically opened test partition; the present count used only train and validation and supplies no held-out performance claim. Only the 365-day horizon is reported. Censuses at 30 and 90 days, and counts of recorded deaths and of deaths after contact cessation, were based on integer-hour differences, were not recomputed with exact timestamps and are not reported.

S3.3. Administrative span

The span from the index discharge to the last visible contact (exact, not capped at the horizon) among at risk people had a median of 1,110.1 days in development (quartiles 259.0 and 2,338.5; 5th and 95th percentiles 14.7 and 3,997.5) and 1,148.7 days in internal assessment (quartiles 277.6 and 2,359.4; 5th and 95th percentiles 15.0 and 4,042.3); the 1st percentile was 3.7 and 4.0 days. The span is long because later admissions and ECGs extend it, and it says nothing about the continuity of contact (S3.5).

S3.4. Integer-hour arithmetic (corrected)

The stored follow-up variable had been built with a database date-difference function in hour units that counts crossed hour boundaries, not elapsed hours: it equals the difference of the floor values of the two timestamps in hours (100% of rows). It is never smaller than the elapsed time rounded down and at most one hour larger. Compared with a threshold in whole units of 30 days or more it can only over-count by at most one hour; compared with the index discharge in seconds (a threshold of zero hours after discharge) it matters, because a person with no contact after discharge can obtain a difference of one hour. Table S21 shows both variants.

Table S21. Integer-hour (archive) against exact timestamps.

Quantity	Development, integer-hour	Development, exact	Internal assessment, integer-hour	Internal assessment, exact
Compatible index discharges	28,270	28,284	9,745	>9,745
Without positive administrative span	5,119	10,371	1,800	>3,660
At risk	23,151	17,913	7,945	6,079
Contact end, at risk	9,769	4,517	>3,340	1,485
Contact end among people with a positive span, no documented AF	43.6%	26.4%	about 44%	25.6%
Complete cases	15,092	11,609	5,257	4,003
Contact end among people with a positive span, no documented AF, complete cases	43.5%	25.9%	about 44%	25.7%

Under integer-hour arithmetic, 5,252 (development) and more than 1,860 (internal assessment) at-risk people had exactly zero contact after the index discharge and were given a positive span. Exact contact spans of all people counted as at risk under integer-hour arithmetic were (internal assessment: the first two classes are merged so that the effect of the correction cannot be recovered cell by cell):

Table S22. Exact contact span after the index discharge among the people counted as at risk under integer-hour arithmetic.

Exact span	Development	Internal assessment
exactly 0 h	5,252	merged with next row
over 0 to 7 days	427	2,013 (exactly 0 h to 7 days)
over 7 to 30 days	1,097	374
over 30 to 365 days	3,695	1,220
over 365 days	12,680	4,338

In development, the last contact of these people was the index discharge itself for 5,252, the discharge of a later admission for 13,598, and an ECG after the last discharge (outpatient, emergency or other) for 4,301; internal assessment showed the same pattern. After their last contact, 2,738 (development) and 921 (internal assessment) people still had an emergency-department stay, which is not a contact source in this definition. A search of the other analysis code for the same construction found state changes for at most 10 ECGs per endpoint and partition at thresholds of 30 days or more.

S3.5. Contact chain (sensitivity)

One reading of uninterrupted contact requires each contact (admission or ECG) to begin within a maximum gap after the previous chain end; the chain end then replaces the last contact, and AF documented after the chain end is censored. This was computed once and not independently repeated; another reading (contact ends only after a gap) was not computed.

Table S23. Contact chain with maximum gap, development.

Maximum gap	At risk	Documented AF	Contact end	Contact end among people without AF	Horizon end	AF documented after chain end (up to 365 days)
90 days	8,791	511	7,943	95.9%	104	88
30 days	5,934	317	5,487	97.7%	<20	123

In internal assessment, 2,998 people were at risk with a 90-day gap (contact end 2,684, 95.7%; AF after the chain end 32) and 2,065 with a 30-day gap (contact end 1,907, 97.7%; AF after the chain end 46). A span of 365 days of uninterrupted contact is almost never documented (below 0.1% of index discharges at a 30-day gap and 0.4% at 90 days). The horizon is an administratively defined span from the index to the last visible contact.

S3.6. Scope of the count

The other pre-risk selection steps (upstream candidates) were not reconstructed. Demographic characteristics of the exact-timestamp stages were not recomputed (S4).

S4. Administrative sex, age and missingness

MIMIC-IV names its administrative field gender. The source paper calls it administrative gender (Johnson *et al* 2023a), whereas the table documentation describes it as genotypical sex; we treat it as a recorded category. We use its recorded F/M/unknown values as sex categories without verification of biological sex or gender identity. Sex was an exact category in laboratory matching. It was not a tested modifier of clock accuracy. The following count describes the selected AF index discharges only; excluded upstream records and the clock and laboratory samples have no new demographic table in this audit. Age and sex of the exact-timestamp stages (at risk, complete cases) were not recomputed after the correction in S3.4, and the earlier values for those stages, which rested on integer-hour arithmetic, are withdrawn.

Table S24. Age and administrative sex of the selected AF index discharges.

Parti- tion	Stored stage	People	F	M	Unknown	Age miss- ing	Age Q25	Age median	Age Q75
Devel- opment	Select- ed index dis- charges	28,297	14,642	13,655	0	0	47	59	70
Internal assess- ment	Select- ed index dis- charges	>9,745	4,980	>4,764	0	0	47	59	70

The internal-assessment selected-index discharge count and its male count are coarsened, because together with Flow S6 they would reveal the number of index discharges (fewer than 10) excluded for incompatible references. Age summaries are stored context values in years; upstream age derivation was not independently reconstructed. Unknown=0 applies only to these selected stages. No sex interaction, equality of clock accuracy or population-wide absence of missing sex is asserted. ED/ICU setting attributes are not stored in this fixed AF cache and were not joined from mixed sources. Missingness causes loss at complete-case selection (35.2% and 34.2% of at-risk people, S3.2), and selection can be informative. Variable-level audit missingness and sex-specific results outside these aggregate stages remain unreported, rather than silently imputed or assigned normal status. A technical feasibility count with no biological sex-effect objective does not justify an invariance claim.

S5. Reporting checklists

STROBE (von Elm *et al* 2007), RECORD (Benchimol *et al* 2015), GRRAS (Kottner *et al* 2011) and SAGER (Heidari *et al* 2016) were checked against published records. Locations refer to the accompanying manuscript (M) and this supplement (S). “Partial” states a reporting gap; it does not mean the item passed. Applicability is to a retrospective technical/observation audit, not a causal treatment or diagnostic-accuracy study.

S5.1. STROBE

Table S25. Reporting checklist.

Item	Status	Location and explanation
1a design in title/abstract	Complete	Title and Abstract identify technical feasibility audit.
1b balanced summary	Complete	Abstract reports identifiability, null comparison, individual support, limits, exact-timestamp contact ends and the MC-MED control.
2 background	Complete	M1, known clock warnings and synchronization literature.
3 objectives	Complete	M1; question concerns individual support and controlled opportunities.
4 design	Complete	M2.1, retrospective collation, previously inspected partitions.
5 setting/dates	Partial	M2.1/S opening: source site/resource period; sample-specific date distributions unreported.
6a selection/follow-up	Partial	M2.2-2.3/S1-S3; AF upstream candidate flow not reconstructed.
6b matching criteria	Complete for implemented screens	S2 tables, person units and unresolved full eligibility distinguished.
7 variables	Complete for stated targets	S1-S3 define RR/chart HR, offset, K/calcium/heart-rate, documented AF and end reasons.
8 sources/measurement	Partial	M2.1/S1-S3; no external clock reference, independent assay or AF-onset adjudication.
9 bias	Complete as description	M4, nominal ICU selection, complete curves, observation and treatment missingness. No bias removal asserted.
10 study size	Complete	Census/feasibility denominators, fixed internal thresholds in M2-3/S1-S3; no power claim.
11 quantitative variables	Complete	S1-S2 enumerate limits, grid, arm thresholds and calipers.
12a statistical methods	Complete for reported estimates	M2.2/S1 bootstrap; S2 maximum cardinality; S3 exact count, no model fit.
12b subgroups/interactions	Not applicable to an effect claim	No new sex interaction; M Reporting/S4 explicitly limits inference.
12c missing data	Partial	S1 unavailable curves, S2 unresolved eligibility, S3 complete-case loss; full variable-level missingness absent.
12d follow-up/matching	Complete for stored design; upstream partial	S2 matching; S3 risk start/end/censoring. No full upstream AF reconstruction.
12e sensitivity	Complete for reported scope	Stay pooling, simulation, two nulls, reference restrictions, two separate K windows, first/all-episode Ca proxy; S6 decision timing.
13a numbers at each stage	Partial	S1.3 exclusive decomposition for alignment; S3 flow with exact timestamps; pre-cache AF and exclusive laboratory flow incomplete.
13b reasons for nonparticipation	Partial	Explicit unavailable stages and eligibility limitations in S1-S3; not all reasons adjudicated.

Item	Status	Location and explanation
13c flow diagram	Complete at documented starting stages	S1-S3 text diagrams; missing upstream branches labeled unknown.
14a characteristics	Partial	S1.4 included against excluded ICU-linked people; S4 selected AF index discharges only; demographics of exact-time stages and of the laboratory audit not calculated.
14b missing values per variable	Partial	S4 age/sex and complete-case stage; comprehensive field-level table absent.
14c follow-up time	Partial	365-day end reasons and exact span quantiles in S3; 30 and 90 day horizons not reported. No follow-up estimand for clock/pair count.
15 outcome data	Complete for stored count	M3/Tables 1-2/S3 365-day events and end reasons; no clinical onset incidence.
16a estimates/precision/adjustment	Complete for reported audit	M3.1 person CI; S1 clustering. Limits/proxy/count are not adjusted effects and receive no invented effect CI.
16b category boundaries	Complete	S1-S2 full implemented categories and numeric cutoffs.
16c absolute risks	Not applicable	No risk-effect claim or cumulative-incidence estimate.
17 other analyses	Complete for reported scope	Failed pooling and negative feasibility decisions retained; no favorable-only window reporting.
18 key results	Complete	M4-5.
19 limitations	Complete as disclosure	M4/S1-S4; unknown magnitude of bias explicitly remains.
20 interpretation	Complete	M4-5 distinguishes nonidentifiability/nonestimability from biological null, and contact from event freedom.
21 generalizability	Complete as limit	M2.1/4, inspected internal development, no external clock or disease validation.
22 funding	Complete	M Funding, self-funded, no specific grant.

S5.2. RECORD extensions

Table S26. Reporting checklist.

Item	Status	Location and explanation
1.1 data type	Complete	Title/Abstract/M2.1, routine ECG and EHR audit.
1.2 database/region/time	Partial	M2.1 source database/site/period; local branch periods not quantified.
1.3 linkage	Complete	Abstract/M2.1-2.3 ECG-clinical linkage.
6.1 population codes/algorithms	Partial	S1/S2 rules and laboratory items; S3.1 gives selector order, report-text rule and AF code group; the complete code lists are in the repository (release v1.0.0).
6.2 validation of selection	Partial	Technical disjointness and count checks S3/S6; no independent clinical adjudication.
6.3 linkage flow	Partial	S1-S3; pre-cache and exclusive all-stage linkage losses unavailable.
7.1 variable code lists/algorithms	Partial	S1 HR item, S2 calipers and item-specific rules, S3 end recipe and AF code group; the complete variable dictionary is in the repository (release v1.0.0).
12.1 access to database	Complete	M Ethical statement/Data availability; credentialed clinical data, credentialed training/data use agreement, local analyses.
12.2 data cleaning	Complete for summarized recipes	S1 duplicate medians/timestamp selection criteria; S2 numerical/unit/specimen/order rules; S3 consistency/time rules. Historical raw inputs not regenerated.
12.3 linkage quality	Partial	S1 unique nominal stays, S2 within-person/stay pairing; physical timing and upstream selector binding incomplete.
13.1 selection from database	Partial	All reported starting stages S1-S3; not claimed to start from all released records.
19.1 routinely collected-data limitations	Complete	M4/S3-S4 missing care, diagnosis proxies, informative observation and changing index.
22.1 protocol/code/data access	Partial	M Code/Data availability/S6; aggregates and recipes supplied, code in the public repository (release v1.0.0), aggregate results in this Supplement and Supplementary Data S1, raw data access restricted.

S5.3. Relevant agreement-reporting principles (GRRAS)

This is a thematic application, not certification of the complete original item-level GRRAS checklist. The study compares physiological compatibility and has no independently synchronized clock gold standard or observer panel. Agreement target, population selection, measurement timing/repetition, reference limitations, fixed criteria, bootstrap cluster, unavailable estimates and generalizability are reported in M2.2/3.1/4 and S1. Observer training/blinding and inter-rater reliability are not applicable to automated RR/chart matching; source-clock blinding and true-

clock calibration are not established. Complete original-checklist verification remains a publication-stage task.

S5.4. SAGER

Table S27. Reporting checklist.

Domain	Status/location
Title/abstract	No sex-specific population or effect claim; no sex-restricted title warranted.
Design/definition	M Reporting/S4: administrative gender used as sex category, distinct from gender identity.
Methods	S2 exact sex-category matching; no newly authorized sex-effect analysis.
Results	S4 selected AF index discharges F/M/unknown and age; other audit demographics remain missing.
Discussion	M4/S4 limits generalization; no sex invariance or equal clock accuracy claim.

The absence of complete branch demographics remains a reporting limitation.

S6. Decisions, provenance and reproducibility

S6.1. Decision history and test exposure

The person-alignment specification was bound before its calculation (11 October 2026 local time; locked 10 October 2026 at 23:34:49 UTC); stay pooling was specified after person-audit results were known and locked before pooling. The two diagnostic-run corrections enforced stay boundaries and connected near-optimal regions without changing the criteria. Fixed identifiability and agreement criteria were not relaxed after failure. The specification SHA-256 and full calculation protocol are retained in the internal claim matrix; they are not printed as internal identifiers in this public supplement. The potassium accompanying-laboratory revision followed counts without new effect inspection. The ± 6 h potassium window followed counts and descriptive marker distributions, so is explicitly exploratory. The calcium/heart-rate screen preceded QT-value inspection and did not establish the required beat population. AF index revision followed inspection, changing selection rather than replicating a fixed index. The stored-selector count was subsequently reproduced without refitting prognostic models, and its time arithmetic was corrected after the first analyses (S3.4). The simulation, null constructions, individual-support analysis, loss decomposition, reference restrictions and exact-timestamp count were specified before calculation but after the earlier results had been inspected; no formal prespecification was locked for them. The MC-MED control (S7) was locked by specification and script hash before its waveform analysis; its post hoc diagnosis was not.

Both internal partitions had been inspected. The research program test partition was previously opened for other exploratory research, including earlier documented-AF design and follow-up revisions. It is not an unseen external or independent confirmatory sample. The retained clock/matching analyses and revised AF discharge selector count are development-only; this au-

dit accessed no test data, footer or result. Historical test contact is disclosed from the access register, without re-opening those materials. No person identifiers are included in this supplement.

S6.2. Reproducible recipes

Person-level heart-rate alignment uses the offset range and rules in S1, with original bootstrap seed 110; stay pooling retains person clusters and seed 115. Shift and 20-permutation sensitivities use the recipes in S1.6 and seed 181001; the simulation and 200-permutation null use seed family 242, the circular-shift null seed family 243 and the individual-support analysis seed family 245 (S1.9 to S1.11). Laboratory matching uses S2's exact categories/calipers and maximum cardinality/flow agreement. AF administrative-span count uses S3's first-eligible-discharge selection, documented-event references and 365-day end rules with exact timestamps, both before and after completeness filtering. These describe computable recipes; upstream raw clinical-event reconstruction remains incomplete.

The aggregate curve is supplied as Supplementary Data S1 (CSV, with a copy of identical content as .xlsx). Private selectors, matching lists, clinical records and full texts are not distributed. 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 this Supplement and in Supplementary Data S1, and no data or result tables are deposited in the repository. The repository holds the original scripts of the person-level alignment with its bootstrap intervals (seed 110; folder `population_alignment`) and of the stay-level pooling with person clusters (seed 115; folder `stay_analysis`), independent recomputations of the person-level estimates and the scripts of the analyses added after the first analyses. The time fields (ECG time, chart time, storage time, admission and discharge times, last ECG time), the AF code group and the report-text rule (S3.1), the completeness variables (loop-width marker, age, administrative sex, machine heart rate) and the selection order are defined in this supplement. LoopLens version 10.1 (release of 7 October 2026) supplied the stored context data and, for the complete-case subset (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). The analyses added after the first analyses used separate scripts that read the stored data. The internal claim matrix retains source hashes and row-level aggregate provenance; it is not a public repository deposit. A contemporary hash does not prove every historical generator/input identity.

S6.3. Interpretation safeguards

Table S28. Interpretation safeguards.

Safeguard	Position for the reported finding
mechanical coupling	Restricted: optimized HR agreement is not independent clock truth; AF is a documentation endpoint.
selection	Restricted: ICU repeated/complete curves and selected discharge index discharges; no positive/negative predictive values or population clock-error rate.
analysis-choice selection	Restricted: fixed clock criteria, inspected development data; wider K window and AF index revised after inspection.
multiplicity/confirmation	Restricted: many correlated offsets; no universal priority, independent test or favorable-only feasibility claim.
informative missingness	Open: full-curve availability, treatment coverage and post-discharge observation.
comparator	Open for clock calibration: chart HR is not synchronized hardware truth; no clinical effect inferred from proxy matching.
verification/within-person	Restricted/open: same-person recordings distinguished from between-person matching; AF onset/absence unadjudicated.
quality/device/text	Open: stored quality/device labels and machine rhythm summaries do not establish hardware identity, clinical rhythm or absence of measurement artifacts.

A synchronized independent source or a justified new eligibility/observation contract could change feasibility. It would not retrospectively validate existing offsets or make missing AF diagnoses observed. No biological equivalence margin or causal effect is estimated.

S7. External technical check in MC-MED

S7.1. Status and scope

This section reports a prespecified technical control, not a validation of the MIMIC results. It was specified and locked before any waveform-derived heart rate was calculated, and it was run once on the training split. The prespecified primary relation (charted heart rate against waveform heart rate within ± 15 min, relation B) was made conditional on a control (relation A) and stayed locked. The post hoc diagnosis in S7.6 was carried out after the prespecified result was known; it does not change the prespecified outcome and is not a rescue of the threshold. The test splits and the chronological splits of MC-MED were not opened. The validation split served for a pilot without verdict and for the choice of the beat detector.

S7.2. Data

MC-MED version 1.0.1 (Kansal *et al* 2025a, 2025b) holds 118,385 adult emergency-department visits at one academic medical center in 2020 to 2022, with continuously monitored vital signs, waveforms (electrocardiogram, photoplethysmogram, respiration), charted vital signs and clinical events. The release (project page checked on 16 October 2026: “Multimodal Clinical Monitoring in the Emergency Department (MC-MED)”, version 1.0.1, published 25 September 2025, doi:10.13026/wvyw-g663) is available only 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. The agreement has no clause on aggregate or derived data, and the project page has no sentence on publishing results. Two clauses of the license text distributed with the data bear on this report: clause 2, “The LICENSEE will exercise all reasonable and prudent care to avoid disclosure of the identity of any individual or institution referenced in PhysioNet restricted data in any publication or other communication” (observed here by reporting aggregates only, with counts below 10 masked), and clause 8, “The LICENSEE agrees to contribute code associated with publications arising from this data to a repository that is open to the research community” (the code is deposited; see the manuscript’s Code availability). The same clauses appear in the data use agreement (as clauses 2 and 9). Only the visit table, the numerics table (heart rate and heart-rate variability, source monitor or chart), the waveform summary, the lead II waveforms (500 Hz) and the random training and validation split lists were read. Training had 94,657 visits of 56,436 patients, of which 44,182 had a lead II waveform; validation had 12,062 visits of 7,054 patients, with no patient in both. Waveform times lie on the visit’s time axis, shifted by the data providers in arbitrary seconds per patient; a minute-zero check as in the MIMIC audit is therefore not possible, and 98% of charted values lie on the monitor’s minute grid. Waveform segments begin at or after the room time, and at least 1% of them begin within 0.01 s of it, which suggests that the providers cut the monitor stream to the room period; relation A therefore has a target offset of zero by construction. Charte heart rate is sparse (median 2 values per visit within the visit window, third quartile 4) and lies mostly outside monitored periods (about one in 25 charted values has a monitor stamp within 60 s), and it is not a copy of the monitor value (1.7% equal within 60 s, in the range of chance).

S7.3. Design

Units are 5-min raster windows of the lead II waveform within the longest single segment of a visit (chosen by header duration, the earlier segment on equal duration; applied identically to the original analysis, the null controls and all sensitivities), within the visit window shortened by 600 s at each end. Waveform heart rate came from an R-wave detector (relation A: trailing 60 s, to match the monitor value, which is the mean of the preceding 60 s; relation B: centered 10 s with at least 4 RR intervals). Reference series were the monitor heart rate (relation A) and the charted heart rate (relation B). The estimator followed the MIMIC audit: candidate offsets on a grid, nearest reference (earlier on ties), complete-unit rule, person-level median curve, near-optimal zone within 1 beat per minute of the minimum and identifiability criteria (at least three complete units, connected zone no wider than 120 min, minimum error at most 5 beats per minute, no zone point at the grid boundary).

Table S29. Grids.

Relation	Search range	Step	Reference distance	Anchor condition
A, monitor heart rate (technical control, target offset zero)	± 60 min	2 min	at most 90 s	none
B, charted heart rate (primary, gated)	± 15 min	5 min	at most 30 min	at least 2 distinct charted anchors at offset zero

Threshold H0 (relation A): at least 80% of eligible visits identifiable, at least 95% of the identifiable within ± 2 min of zero, and the population minimum at 0 ± 2 min (all three parts). Prespecified further checks were H3 (identifiable share falls monotonically with reference spacing of 1, 5, 15, 30, 60 and 120 min, at least 80% at 1 min and at most 40% at 60 min), H4 (a within-visit permutation null identifies some visits, and a null with another patient's reference series at most 5%), H5 (inserted shifts of 0, ± 10 and ± 30 min are recovered), H6 (median zone width at most 10 min) and NK4 (heart rate rescaled by 15% keeps at most 5% identifiable). The thresholds were set before the analysis; the expectations were not blind to the MIMIC results, and a zero population minimum was expected by construction of the data set (the target offset is zero), whereas the identifiable share was not constrained by it. For relation B a status followed from the number of visits usable for a structure-preserving null (cyclic rotation of the unit heart rates by at least 60 min, at least 5 admissible rotations): at least 5,000 visits and 2,500 patients confirmatory, at least 1,000 and 500 exploratory, otherwise description only. If H0 failed, relation B was locked.

S7.4. Locking of the specification and detector choice

The specification, the script, the configuration, the helper functions and the numerical software version were locked by hash on 16 October 2026 (08:05 local time) before the waveform extraction (hashes retained internally). Only the beat detector's quality limits could change, within pre-set ranges, and only on the validation split, by technical quality alone: agreement with the monitor heart rate within 5 beats per minute in at least 95% of windows and a yield of at least 0.80, with no

relation-B quantity involved. Five candidates were extracted on the full validation population (101,937 windows, 166 read errors counted as invalid windows) and one was chosen by the pre-set rule and frozen before any estimate of the offsets.

Table S30. Detector candidates on the validation population.

Candidate	Maximum RR variation	Minimum cover-age, s	Agreement within 5 beats per minute	Yield	Criterion met	Chosen
K1	0.50	48	0.983	0.894	yes	no
K2	0.40	48	0.984	0.893	yes	no
K3	0.30	48	0.988	0.886	yes	yes
K4	0.50	45	0.983	0.894	yes	no
K5	0.50	54	0.983	0.894	yes	no

S7.5. Results, training split (bound analysis)

Population: 42,914 visits had units in the longest segment; 6,505 had at least one complete unit and 1,823 at least three (relation B, raw). The anchor condition excluded 1,316 of these, leaving 507 for relation B, of which 37 (37 patients) were admissible for the structure-preserving null. Reasons for non-admissibility were too few admissible rotations (376), non-stationary heart rate (45), seam jump (36) and gaps (13). The status was therefore description only, and no verdict on the target statement was issued.

Table S31. Relation A (technical control), training split.

Quantity	Value
Visits with units	42,914
Eligible (at least three complete units)	11,913
Identifiable	3,376
Identifiable share (95% interval, patient bootstrap)	28.3% (27.5-29.2)
Median offset, min	0
Within ± 2 min among the identifiable	98.5% (98.1-98.9)
Population minimum at, min	0
Median zone width, min	0 (a single accepted grid point, not zero timing uncertainty)
Ties	0%
Gate H0	not met (share below 80%; the other two parts met)

A zone width of 0 min means that one grid point (on the 2 min grid) was accepted; it does not mean zero timing uncertainty, and the same holds for H6. Identifiability rose with heart-rate variability (terciles of the SD of the unit heart rate): 7.7% (307 of 3,971), 26.2% (1,040) and 51.1% (2,029), with at least 98% within ± 2 min in each tercile. The upper tercile also missed the gate.

Table S32. Prespecified checks, training split.

Check	Result	Label
H3 density: identifiable share by reference spacing 1, 5, 15, 30, 60, 120 min	28.3%, 32.5%, 37.8%, 43.3%, 35.2%, 9.2% (own denominators about 11,900; 11,443; 12,375; 12,654; 11,823; 9,770)	not met (not monotone; 1 min below 80%)
H4 within-visit permutation null	5.0% at most (599 of 11,913 at most)	permutation identifies some visits
H4 other-patient reference null	6.0% (range of replications 5.3% to 6.5%)	not met (limit 5%)
H5 recovery of inserted shifts (0, ± 10 , ± 30 min)	median error 0 min in every condition (jointly identifiable visits 2,235 to 3,376)	met
H6 median zone width	0 min	met
NK4 heart rate rescaled by 15%	8.4% (7.9-8.9) identifiable	not met (limit 5%)

Sensitivities of relation A: search range 30 min with 1 min step 21.4% of 25,374 visits; range 120 min with 2 min step 26.5% of 1,969; unit spacing 900 s 15.3% of 8,239; at most 6 units 8.2% of 3,625; unshortened visit window 28.8% of 12,457. The within-visit null identified 4.7% on average (replications 512 to 599 of 11,913), the other-patient null 6.0% (denominator 5,637 visits on average, with 12 visits without a partner).

S7.6. Post-hoc diagnosis (not bound)

The diagnosis used the same training split (relation A, 11,913 visits) after the prespecified result was known and reproduced all prespecified quantities exactly (denominator, identifiable count, share, terciles, nulls and all six density stages). Variants are diagnostic and are not alternative rules.

Exclusion reasons. Two views must be kept apart. The priority order was (i) minimum error above 5 beats per minute, (ii) zone touching the grid boundary, (iii) near-optimal set not connected, (iv) width above 120 min. In this order (the first violated rule), of 8,537 non-identifiable visits 7,118 (83.4%) failed first on the connectedness of the near-optimal zone, 1,407 (16.5%) on the boundary rule, 12 on the minimum error, and none on the width, which cannot occur on a grid of ± 60 min. Independently of order, more than 8,400 visits (more than 70% of all eligible visits) violated connectedness, about 12% of all eligible visits the boundary rule, and 1,306 both. Because no visit violated the width rule, the 7,118 visits that failed first on connectedness are the same visits that violated only connectedness; in general priority and exclusivity are different. The overlap counts confirm it: 7,118 visits violated (iii) alone, and none violated (iv) in any combination. The minimum error played no role (median minimum 0.38 beats per minute). Among the more than 8,400 visits with a connectedness violation, the nearest foreign component of the near-optimal set lay a median of 8 min from the chosen minimum (quartiles 6 and 14); in 1,795 visits exactly 4 min (a gap of one grid point), in 5,401 up to 10 min and in 742 over 30 min. Across the variability terciles, the share with a connectedness violation fell from 90% to 74% and 49%, and the share with a boundary violation from 25% to 8% and 3%.

Table S33. Relaxing the connectedness rule on the data and on the null controls.

Rule	Data: identifiable	Within ± 2 min among the identifiable	Within-visit permutation null identifiable	Other-patient null identifiable
Bound rule	28.3%	98.5%	4.7%	6.0%
C1, connectedness rule dropped	88.1%	85.0%	63.8%	65.9%
C2, rules on the component around the minimum only	97.9%	83.0%	87.9%	84.0%

Without the connectedness rule, the null controls are identifiable in 64% to 88% of visits, so the rule is the part that separates data from null. Location of the chosen minimum, without any zone rule: 82.0% (81.3-82.7) of all eligible visits had their minimum within ± 2 min of zero, against 5.9%

under the within-visit permutation null and 4.7% under the other-patient null (chance expectation 4.9%). Among the 7,118 visits that violated only connectedness, 5,593 (78.6%) had their minimum within ± 2 min. By variability tercile, 75.8%, 82.7% and 87.6% of visits had the minimum within ± 2 min. The prespecified rule therefore discards mostly visits whose chosen minimum lay within the prespecified location window (± 2 min of the target offset zero); it behaves as a conservative test of a unique minimum with a narrow tolerance, where conservative means low yield under strict criteria and not calibrated error control.

Density curve. On the common denominator of visits present at all six spacings (6,481 on average; 2,588 in all 20 replications), the identifiable share was 37.8%, 38.5%, 42.8%, 45.9%, 36.7% and 10.5% at 1, 5, 15, 30, 60 and 120 min, so the non-monotone course is not a denominator effect. It is the sum of two monotone but opposite contributions: violations of connectedness fell from 52.8% to 1.9% with growing spacing, while boundary violations rose from 9.3% to 76.3% and minimum-error violations from below 1% to 11.3%. Without the connectedness rule, the share fell monotonically (C1 90.6% to 12.4%). The location of the minimum fell monotonically from 90.3% to 21.8%; among visits identified by the prespecified rule, 99.2% (1 min), 60.0% (15 min) and 40.5% (30 min) lay within ± 2 min. The rising part of the curve up to 30 min therefore measures in part the behavior of the rule (smooth, wide zones become connected more easily), not additional timing information.

Correction of a label. The bound results table described the denominator as falling with reference spacing. That is wrong: the denominator rises from about 11,900 at 1 min to 12,654 at 30 min and falls afterwards (11,823 at 60 min and 9,770 at 120 min), because the admissible reference distance grows with spacing (the larger of 90 s and half the spacing) while thinning removes reference points. The stages compare different sets of visits (6,481 present at all stages; 16,113 in at least one). The sentence “reference density does not explain the MIMIC result” should read: the prepared density explanation was not supported, and the mechanism is described by the rule structure (connectedness against boundary).

S7.7. What the check cannot show

- It does not bear on relation B, on unknown offsets or on MIMIC. Relation A has target offset zero by construction; the location share of 82.0% applies to that target and is not a general accuracy. A transfer of the rule’s strictness or of the tercile gradient to MIMIC was not examined.
- The gradient by variability is an association; the terciles are built from the same unit heart rates that form the error curve, and unit count, observation length and signal quality were not separated.
- The null controls have a different reference geometry and, for the other-patient null, a different denominator; they are not calibrated error rates.

- The mechanism of “gaps in the zone around the minimum” is described but not explained: noise in 60 s monitor values, quantization of the heart-rate estimate to whole units and the 2 min grid were not separated. Alternatives to the fixed 1 beat per minute tolerance (relative tolerance, other bands, zone width as the measure) were not examined, and recovery of inserted shifts was not recomputed under the variants.
- The intervals of the density curve do not include the randomness of the thinning, and the density curve confounds reference points, admissible distance, complete units and the visit denominator.
- A rule calibrated on null controls, or a joint rule for location and zone, would have to be specified in advance with its own null control; none was set here. Charted heart rate in the emergency department mixes clock difference, documentation delay and measurement method, and no such component is separated.
- Percentages cannot be compared directly with the MIMIC audit, which differs in unit, search range, reference and denominator.

References

Bibliographic entries are given in the accompanying manuscript: von Elm *et al* (2007), Benchimol *et al* (2015), Kottner *et al* (2011), Heidari *et al* (2016) and, for MC-MED, Kansal *et al* (2025a, 2025b). No guideline supplies an external numerical accuracy threshold for this audit.