

Supplementary Material: How much does the monitor T wave change per mmol/L of potassium? A within-patient analysis in emergency department patients with potassium below 6 mmol/L

Supplement to the manuscript of 17 October 2026. Leo Probst, Johannes Kepler University Linz, Medical Faculty, Altenberger Strasse 69, 4040 Linz, Austria; leo.probst.research@proton.me. Reference numbers [n] refer to the reference list of the main text.

Contents: S1 Specification, timeline and post hoc analyses; S2 Methods in detail; S3 Flow of pairs; S4 Repeatability and individual changes; S5 Sensitivity analyses of the T-amplitude slope; S6 Controls; S7 T-wave steepness; S8 Timing of the ECG segments; S9 Reference reading; S10 Other questions of the specification; S11 Characteristics of primary and other pairs (post hoc); S12 Reporting.

Conventions. Counts of 1 to 9 are shown as <10, and counts from which such a number could be recalculated by subtraction are rounded ("about"). Patient counts are given rounded down to tens. All amplitudes are in mm at 10 mm per mV unless stated. Sections are numbered S1 to S12 and tables S1 to S14 independently of each other. Confidence intervals (CI) are 95% percentile intervals from a patient-cluster bootstrap unless stated. The primary analysis and the other analyses of the specification are not adjusted for multiple comparisons, and the sensitivity analyses share the same pairs.

S1. Specification, timeline and post hoc analyses

S1a. What was fixed and when

The analysis was written down in a specification (version 2.2) and implemented in a script with a self-test (40 of 40 checks passed). Both were locked together with the configuration on 16 October 2026 at 20:07:45 local time (18:07:45 UTC) by a lock file held outside the project folder; the lock has no independent time stamp. At every start the script checks the hashes of the specification, itself, three measurement modules, the configuration and all input files against the lock file and stops on any mismatch. The specification and script went through four rounds of code review by an AI tool (OpenAI Codex) before the lock. No analysis of the relation between potassium and any ECG measure was run on the data before the lock.

Object	SHA-256
Specification v2.2	8e09350dab33bc5ce81e2a11127899c2ba8afa84de5a5b4f6833a3577e42381e
Analysis script	04ac15dae0f8a6a0a9b2d154e43b4ff3d240acd410d10707498fa53d48c9dcea
Configuration of the script (all measurement parameters, seeds, thresholds)	12a93e65e801a48094bec1c454f4ea1323fd34983e249599afabf0c9b826f823
Lock file	bed63aff161d627afe62c256705927b48c14b134bd8f7b02f7a7e1d766ddd886

Software: Python 3.12.4, NeuroKit2 0.2.13, numpy 2.5.3, pandas 2.3.3, scipy 1.18.1, statsmodels 0.15.0, random numbers from numpy PCG64 with seeds fixed in the configuration. The reference reading of the author (S9) was exported on 17 October 2026 at 10:56 local time and the main run

started at 10:59, so the reading was completed after the analysis was locked and before the main analysis was run. The main run covered all three random splits of the database (training, validation, test) in one pass and was completed on 17 October 2026. The splits are random by visit and not independent by patient; results by split are a stability description, not a confirmation.

S1b. What had been seen before the lock

A feasibility check on the training split counted pairs, laboratory availability, ECG coverage, signal quality, the repeatability of the measurement algorithm and the marginal distributions of the outcomes (median T amplitude 0.19 to 0.22 mV; SD of the T-amplitude change over all pairs 0.127 mV; SD of the potassium change in clean pairs 1.34 mmol/L). It did not examine any relation between potassium and an ECG measure and did not read the validation or test splits. After the lock, 177 of the 216 primary pairs proved to lie in the training split, whose marginal distributions had been seen. Two external results were known: an intraoperative database (VitalDB [14,15]; unpublished, non-peer-reviewed analysis of the author, slope 0.0213 mV per mmol/L, 95% CI 0.0103 to 0.0331, that is 0.21 mm per mmol/L) provided the expected slope, and an earlier analysis in another database provided the expectation for the QRS question (S10). The expectations were therefore not blind.

S1c. Analyses added after the results were known

After the locked results were seen, the robust repeatability limit of the amplitude was calculated (S4), and a robustness analysis of the steepness outcome was added (S7): sensitivity analyses, seeds, repeatability of steepness, controls on steepness, a common scale for steepness and amplitude, and an analysis of steepness adjusted for amplitude. Its script was written and tested on artificial data before it was run. These runs also read the test split because all splits are part of the locked primary set. A descriptive comparison of the primary pairs with other pairs (S11) and a reader check of the influential pairs (S5) were also added after the results were known. These analyses are exploratory, they change no locked category, and they are marked post hoc wherever they appear. The repeatability limit of steepness and its ratio to the slope are also post hoc.

S2. Methods in detail

Data. MC-MED version 1.0.1 (credentialed access): labs, orders, visits, minute-level monitor values and continuous lead II waveforms. Times are UTC seconds; a time check required the median offset between the first monitor value and the first waveform segment to lie between -60 and 300 s (observed median 41 s). Lead II had to be in mV; potassium, bicarbonate and sodium in mmol/L (no rows with another unit were found).

Samples and pairs. A sample is one potassium value per visit, source and order time. Serum values stay serum values; the “all sources” analysis prefers serum and otherwise takes a unique point-of-care or whole-blood value. Conflicting values of the same source and order time are dropped. Censored values ($>$ or $<$) are never analysed as exact values and remove the pair. A pair

is a sample and the next sample of the same visit with 1 to 12 hours between the order times. Each end carries the result time of its own source; a pair without a serum result time is excluded.

Treatment. Sources: medication orders with administration time and dialysis orders (which have only an order time). Classes: potassium-lowering (regular insulin, glucose boluses, bicarbonate, albuterol, binders), ECG-active (intravenous calcium), potassium supplements, dialysis (haemo-dialysis, peritoneal dialysis, continuous renal replacement), weak or unclear (other insulins, loop diuretics). The window of a pair runs from the earlier of the order time and the ECG time of the first end to the latest of the second order time, second result time and second ECG time. A pair is “documented” if a class was given from 120 minutes before the window start until its end, or a dialysis order lies from 360 minutes before the start; “unclear” if a class was given 120 to 360 minutes before the start, an order lacks an administration time, a weak class was given, or dialysis was ordered elsewhere in the visit; and “no indication” otherwise. 34% of medication orders have no administration time (79% for 50% glucose), so “no indication” means no documented or assumed treatment, not untreated. The primary analysis used “no indication” and an initial potassium below 6.0 mmol/L.

ECG segments and windows. The search window for an end ran from 15 minutes before the earlier of order and result time to the earlier of 15 minutes after the later of the two and 3 hours after the earlier. Within it we took the clean 10-second segment (grid of 5 s) whose midpoint was nearest to the order time. A segment was clean if it had no missing value, no saturation (above 4.5 mV), no flat line of 60 ms or more, a range of at least 0.3 mV, 5 to 37 beats, a mean RR interval of 0.27 to 2.0 s, a median beat of at least 0.15 mV and high correlation of the beats with their template. The measurement window was the 60 seconds around the segment midpoint, shifted at the true segment ends. Pairs whose two windows overlap were excluded in every analysis. A second window for the repeatability analysis was the next clean segment at least 90 s from the first midpoint whose window did not overlap the first.

Measurement algorithm. The recording is sampled at 500 Hz. Segment selection uses a zero-phase band-pass filter (second-order Butterworth, 0.5 to 40 Hz) and a 5 to 20 Hz filter for the QRS envelope. For the measurement, NeuroKit2 version 0.2.13 [16] cleans each 60-second window and detects R peaks, and the algorithm forms a median beat aligned to the R peak (200 ms before to 0.75 RR intervals after). The baseline is the median of the 32 ms ending 8 ms before QRS onset (PR isoelectric level). The T peak is the extremum of largest magnitude in a search window that scales with the square root of the RR interval (a biphasic T wave is therefore measured at its larger deflection); steepness is the slope of the descending limb from the T peak over 0.12 times the square root of the RR interval. A window is invalid if more than 30% of the beats are irregular or if fewer than 20 s of good beats remain. The main QRS direction is the larger deflection from the baseline; polarity is “uncertain” if the smaller deflection is at least 70% of the larger. The signal is turned so that the main direction is positive, and the T amplitude (peak minus PR isoelectric level) and the steepness of the descending limb are measured. The primary outcome is the QRS-relative T am-

plitude (positive: T in the QRS direction). A raw signed lead II amplitude is a sensitivity analysis. The algorithm was built on earlier measurement modules of the author (used in an intraoperative analysis) with the minimum good-beat coverage reduced from 60 to 20 s; all parameters are part of the locked configuration. The T-wave measures do not use the wavelet delineation of NeuroKit2; it is used for other intervals only.

Heart rate. Monitor heart rate, median of the values within 5 minutes of the segment midpoint (at least 3 values). The covariates are the three components of the difference of a restricted cubic spline (knots 55, 70, 85, 110 per minute) between the two ends.

Model. $\Delta T = \alpha + \beta \Delta K + \sum \gamma_j \text{dB}_j + \varepsilon$ by least squares, free intercept. The unadjusted and the adjusted model run on the same pairs. Intervals: 2,000 patient-cluster bootstrap draws, percentile method; sensitivity analyses on subsets used 500 draws. Estimates need at least 30 pairs and 10 patients. For the primary slope the specification set a margin of 0.1 mm per mmol/L (0.010 mV) and categorised results as imprecise (lower limit at or below zero and upper limit at or above the margin), positive above the margin, and others. The T-amplitude slope was “imprecise”. For steepness the margin was 0.2 mV/s per mmol/L (the amplitude margin times the ratio of steepness to amplitude of 20.8 per second observed in the intraoperative database).

Precision. The specification gave a planning formula, standard error = residual SD / (SD of $\Delta K \times \sqrt{(\text{pairs} / \text{design effect})}$), the half-width 1.96 times that, and the smallest slope detectable with 80% power as 2.8 times the standard error. For about 200 pairs, an SD of ΔK of 1.34 mmol/L and a residual SD of 0.06 to 0.127 mV, the planned half-width was 0.06 to 0.14 mm per mmol/L and the smallest detectable slope 0.09 to 0.19 mm. With the observed values (216 pairs, SD of ΔK 0.70 mmol/L, residual SD 1.43 mm) the formula gives a standard error of 0.14 mm and a smallest detectable slope of 0.40 mm per mmol/L; the bootstrap standard error was 0.17 mm and the half-width of the interval 0.33 mm.

Repeatability. Within-sample SD, intraclass correlation (1,1), bias and the repeatability limit $1.96 \times \sqrt{2} \times \text{SD}$ within, from two non-overlapping windows of the same sample [17,18]; reporting follows GRRAS [19]. A robust version uses a robust SD; it does not correspond to the 95th percentile of the differences. The specification used the classical limit; the robust limit was calculated after the results were known (post hoc). Both limits describe single measurements of the same sample and are not a yardstick for a group slope.

S3. Flow of pairs

Table S1. Flow from consecutive pairs to primary pairs

Step	Pairs	Share of the previous step
Consecutive serum pairs, 1 to 12 h apart	5,661	
Result time present, values not censored	5,585	98.7%
ECG segment at both ends within the window	970	17.4%
Valid measurement at both ends	592	61.0%
Same and clear polarity	522	88.2%
No overlapping windows, heart rate at both ends, no documented or unclear treatment	324	62.1%
Initial potassium below 6.0 mmol/L (primary pairs)	216	66.7%

Primary pairs: 3.9% of 5,585 pairs, about 200 patients. Patient counts are rounded down to tens. The step to 324 pairs combines three exclusions (overlapping windows, missing heart rate, documented or unclear treatment), which are not shown separately.

S4. Repeatability and individual changes

Table S2. Repeatability of the two measures

Measure	Samples	SD within	Repeatability limit (classical)	Repeatability limit (robust)	ICC	Bias (window 1 minus 2)
T amplitude, all samples	21,055	0.63 mm	1.73 mm	0.53 mm	0.960	-0.03 mm
T amplitude, ends of primary pairs only	414	0.46 mm	1.27 mm	0.44 mm	0.973	0.03 mm
T steepness, all samples	21,055	1.47 mV/s	4.08 mV/s	2.06 mV/s	0.864	0.24 mV/s
T steepness, ends of primary pairs only	414	1.48 mV/s	4.11 mV/s	1.66 mV/s	0.865	0.04 mV/s

The repeat differences have heavy tails, so the robust limit (calculated post hoc) is two to three times smaller than the classical one (pre-specified). The main text reports both limits from all samples as context for single measurements. A comparison of the group slope with these limits is not a like-for-like comparison, because a limit describes the difference between two single measurements and the slope describes a group mean; extrapolating the slope linearly to the potassium change at which it would reach a limit goes far beyond the observed range (SD of ΔK 0.70 mmol/L) and is therefore not given. The expected SD of the difference between two windows of the same sample is $\sqrt{2}$ times the within-sample SD, 0.89 mm (0.65 mm using the ends of the primary pairs); the SD of the change between two samples not explained by potassium was 1.43 mm. The limits cover variation within minutes of one sample, not over hours.

Table S3. Change in T amplitude by category of potassium change

Category of potassium change (mmol/L)	Pairs	Median (mm)	Interquartile range (mm)
≤ -1.0	33	-0.08	-0.70 to 0.27
-1.0 to -0.5	28	-0.01	-0.96 to 0.41
Absolute change below 0.5	132	-0.09	-0.67 to 0.41
0.5 or more	23	not summarised	not summarised

The categories with a rise of 0.5 mmol/L or more are pooled because the category with a rise of 1 mmol/L or more holds fewer than 10 pairs. In the category with an absolute change below 0.5 mmol/L, 12% of pairs changed by more than the pre-specified limit in either direction (5% would be expected if the limit applied to the change between two samples; it applies to two single measurements within one sample). The SD of the T-amplitude change not explained by the potassium change was 1.43 mm. A cross-sectional description between patients (one random sample per patient, 6,000 patients, adjusted for age, sex and heart rate) gave a slope of -0.03 mm per mmol/L (CI -0.19 to 0.13); between-patient and within-patient slopes can differ, and this is no evidence for the within-patient effect.

S5. Sensitivity analyses of the T-amplitude slope

Table S4. Slope of T amplitude per mmol/L potassium (mm; 95% CI); all analyses in this table were pre-specified

Analysis	Pairs	Slope (mm per mmol/L)	95% CI
Primary (heart-rate adjusted)	216	0.26	-0.04 to 0.63
Unadjusted for heart rate, same pairs	216	0.25	-0.05 to 0.61
Huber robust estimator	216	0.07	-0.11 to 0.29
Without influential pairs (DFBETAS)	205	above $2/\sqrt{n}$	
Initial potassium up to 3.5 mmol/L	24	not estimated (fewer than 30 pairs)	
Initial potassium 3.5 to 5.0 mmol/L	139	0.42	-0.03 to 0.85
Initial potassium above 5.0 mmol/L (below 6.0)	53	0.19	-0.34 to 0.90
Potassium rises only	63	-0.26	-0.91 to 0.24
Potassium falls only	132	0.33	-0.22 to 1.02
Initial potassium as covariate	216	0.19	-0.19 to 0.61
Change in bicarbonate as covariate	199	0.20	-0.14 to 0.58
Heart-rate change below 10 per minute	142	0.51	0.09 to 0.97
One pair per patient (first)	about 200	0.26	-0.04 to 0.63
Mixed model (random intercept per patient)	216	0.26	-0.01 to 0.54
Without documented treatment (no indication or unclear)	240	0.18	-0.10 to 0.52
All initial values, no treatment indication	324	-0.01	-0.12 to 0.10
Raw signed lead II amplitude, primary pairs	216	0.19	-0.11 to 0.54
Raw amplitude, all valid pairs including polarity changes	245	0.04	-0.28 to 0.40
ECG segment within 15 minutes of the order time at both ends	146	0.18	-0.22 to 0.70
All potassium sources (including point-of-care)	314	0.21	-0.03 to 0.49
Measurement error of potassium, SD 0.1 mmol/L (dilution factor 0.96)	216	0.28	-0.04 to 0.65
Measurement error of potassium, SD 0.2 mmol/L (dilution factor 0.83)	216	0.32	-0.05 to 0.75
Documented treatment (descriptive; median initial potassium 3.7 mmol/L, not a group of treated hyperkalaemia, see S11)	105	0.01	-0.61 to 0.44
Training split only	177	0.30	-0.02 to 0.73
Validation split and test split	19 and 20	not estimated (fewer than 30 pairs each)	

Subset analyses used 500 bootstrap draws, so their interval limits are less stable than those of the primary analysis (2,000 draws). The table holds 22 sensitivity analyses with an estimate; only the

analysis restricted to a heart-rate change below 10 per minute has an interval that excludes zero, and it is one of 22 and is not interpreted. The mixed model converged (second optimiser) with a warning that the estimate lies on the boundary of the parameter space (patient variance near zero); its sign equals that of the primary model. The measurement-error rows are a deterministic rescaling (slope divided by the dilution factor), not a new estimate. The analysis with all initial values adds 108 pairs with an initial potassium of 6.0 mmol/L or more and is the narrowest interval of the table. In these 108 pairs potassium almost always fell over a short interval, and the cause is unexamined (S11); regression to the mean and the different range of potassium can both contribute to the smaller estimate.

Reader check of influential pairs (post hoc; it examines the measurement). After the results were known, the same reader (the author, who specified the algorithm and knew the hypothesis) measured the T amplitude at both ends of the 11 influential pairs, identified by the pre-specified rule ($|\text{DFBETAS}|$ above $2/\sqrt{n}$; Table S4), together with the ends of 11 randomly chosen comparison pairs of the primary set, in random order and blind to potassium, treatment and the algorithm's values. The comparison pairs were not matched on T amplitude, so blinding to which strips came from influential pairs was only partial. The reader marked the baseline and the T peak on three beats per strip as in S9 and gave the QRS direction, which equalled the automatic polarity at all ends. Most of the 22 ends in each group could be read; the number of unreadable ends is not given because it is small. The algorithm's values are the QRS-relative T amplitudes of the locked analysis, compared with the reader's value signed by the QRS direction.

Table S5. Reader and algorithm at the ends of the influential and comparison pairs (post hoc), T amplitude in mV, difference reader minus algorithm

Ends	Mean, reader	Mean, algorithm	Mean difference (95% CI)	SD of the difference	Limits of agreement	ICC (1,1)	Spearman
Influential pairs (22 ends)	0.23	0.09	0.13 (0.02 to 0.25)	0.25	-0.36 to 0.63	0.62	0.67
Comparison pairs (22 ends)	0.36	0.32	0.05 (0.01 to 0.08)	0.08	-0.11 to 0.20	0.73	0.67
All ends read (most of 44)	0.29	0.20	0.09 (0.03 to 0.15)	0.19	-0.29 to 0.47	0.67	0.74
Change in T amplitude within pairs (second minus first end; pairs with both ends read, most pairs)	-0.01	-0.07	0.06 (-0.07 to 0.18)	0.26	-0.46 to 0.57	0.49	0.49

The median difference of the end values was 0.02 mV in both groups and the median absolute difference 0.06 mV at the influential and 0.03 mV at the comparison ends; the largest absolute difference was 0.76 mV at the influential and 0.21 mV at the comparison ends. The mean difference at the influential ends is therefore driven by a few large differences. For the change within pairs the median difference was 0.01 mV, the median absolute difference 0.07 mV and the largest

absolute difference 0.76 mV. No counts of ends or pairs within a tolerance, outside a range or with the same sign are given, because they would be small.

Reading of the measurement. At the influential ends the algorithm measured smaller T amplitudes than the reader (mean 0.09 against 0.23 mV), with a mean difference of 0.13 mV and a wider scatter of the difference than at the comparison ends (SD 0.25 against 0.08 mV); at the comparison ends agreement was closer, as in the 70 readings of S9. For the change in T amplitude within pairs, which is the level that matters for the slope, agreement was poorer than for single values (limits -0.46 to +0.57 mV, ICC 0.49), and the SD of the difference (0.26 mV) exceeds the SD of the observed change not explained by potassium (0.14 mV). The reader is not the truth: the reader also read higher than the algorithm at the comparison ends (0.05 mV) and in the 70 readings (0.04 mV), so the comparison measures a difference between two methods and not an error of the algorithm alone. Limits: post hoc; one reader, who is the author; 11 influential and 11 comparison pairs; comparison pairs not matched on T amplitude; the reader may have recognised the influential strips by their small amplitude; the comparison covers amplitude only.

Table S6. Slope of T amplitude per mmol/L potassium with reader values (post hoc; mm; 95% CI)

Analysis	Pairs	Slope (mm per mmol/L)	95% CI
Algorithm values at all pairs (primary)	216	0.26	-0.04 to 0.63
Without the influential pairs (algorithm values)	205	0.04	-0.10 to 0.23
Reader values at the influential pairs, algorithm values elsewhere	216	0.10	-0.09 to 0.29
Reader values at all pairs read, algorithm values elsewhere	216	0.06	-0.12 to 0.26

Where the reader could not read an end, the pair keeps the algorithm value; a variant that dropped such pairs gave a similar estimate (0.12 mm; -0.05 to 0.33). The slope is estimated with the locked model (heart-rate adjusted, 2,000 patient-cluster bootstrap draws); the reproduction of the locked estimate and of the estimate without influential pairs was a gate of the calculation and passed.

This re-estimation is biased towards a lower slope by construction and is not an estimate of the true slope. Reader values replaced algorithm values only at the pairs that were selected because their algorithm values drove the estimate (and at an equal number of comparison pairs), while all other pairs keep algorithm values with the same unknown error in both directions. Replacing the extreme pairs lowers the slope even if the algorithm is unbiased on average. The intervals treat the reader's values as exact (reading error is not included), no interval is given for the difference between the slopes, and some pairs remain influential by the same rule after the replacement (fewer than 10). The re-estimation does not show what the slope would be if the reader had meas-

ured all pairs. The pre-specified result remains “imprecise”: the interval includes zero and the margin of 0.1 mm per mmol/L.

S6. Controls

Table S7. Controls for the T-amplitude slope (pre-specified) and for steepness (post hoc)

Control	Amplitude (mm per mmol/L)	Steepness (mV/s per mmol/L; post hoc)
Time-shifted ECG windows (2 to 6 h from the order time, outside the search window), 118 pairs	-0.26 (-0.64 to 0.02)	-0.35 (-0.97 to 0.19)
Primary windows of the same 118 pairs	0.16	0.82
Contrast, primary minus shifted	0.42 (-0.01 to 0.90)	1.17 (0.09 to 2.58)
Permutation across patients (500 permutations): null mean, SD	0.00, 0.14	0.01, 0.34
Permutation: 2.5th and 97.5th percentile of the null	-0.30 and 0.28	-0.63 and 0.69
Permutation: observed slope, permutation p	0.26, 0.084	0.75, 0.034
Sodium, change per 5 mmol/L, raw (200 pairs)	-0.11 (-0.45 to 0.17)	-0.25 (-0.87 to 0.41)
Sodium, partial on the potassium change	-0.05 (-0.37 to 0.23)	-0.04 (-0.82 to 0.71)

The correlation between the changes in sodium and potassium was -0.33, so sodium is not a clean null exposure. The specification asked for a margin of ± 0.005 mV per mmol/L around zero to call a control “near zero”, and for the control to block interpretation if its interval excluded zero. No control was near zero, none excluded zero, and the permutation null of the amplitude is wider than that margin. The controls were therefore not blocking and not confirming. The time-shifted amplitude slope has the size of the main estimate with the opposite sign, which points to time structure or regression to the mean as possible contributors and not to a clean null. For steepness the time-shifted contrast has a lower limit above zero, which is compatible with a relation tied to the windows near the sample, but the interval is wide (118 pairs).

S7. T-wave steepness (secondary outcome; analyses beyond the locked slope are post hoc)

Pre-specified result. Slope 0.75 mV/s per mmol/L (CI 0.21 to 1.37; patient-cluster bootstrap with 2,000 draws; 216 pairs). The category against the margin of 0.2 mV/s was “positive above the margin” because the lower limit (0.209) lay just above it. That category is not stable (below). Steepness has no reference reading, and the steepness depends on the T-peak location and the fixed window after it.

Table S8. Post hoc sensitivity analyses of the steepness slope (mV/s per mmol/L); subset analyses used 500 bootstrap draws

Analysis	Pairs	Slope	95% CI
Locked primary model	216	0.75	0.21 to 1.37
Unadjusted for heart rate	216	0.72	0.15 to 1.34
Huber robust estimator	216	0.48	0.01 to 0.89
Without influential pairs	about 200	0.58	0.17 to 0.96
Heart-rate change below 10 per minute	142	1.03	0.41 to 1.75
Potassium rises only	63	0.35	-0.88 to 1.27
Potassium falls only	132	0.39	-0.68 to 1.47
One pair per patient	about 200	0.71	0.19 to 1.33
Initial potassium as covariate	216	0.65	0.10 to 1.16
Change in bicarbonate as covariate	199	0.58	0.01 to 1.19
Mixed model	216	0.75	0.12 to 1.38
Without documented treatment	240	0.57	0.12 to 1.19
All initial values	324	0.02	-0.21 to 0.29
All potassium sources	314	0.52	0.08 to 1.07
Documented treatment (median initial potassium 3.7 mmol/L, not a group of treated hyperkalaemia, see S11)	105	1.20	0.06 to 2.40
Initial potassium 3.5 to 5.0 mmol/L	139	1.27	0.43 to 2.17
Initial potassium above 5.0 mmol/L (below 6.0)	53	0.32	-1.02 to 1.46
ECG segment within 15 minutes of the order time	146	0.62	0.06 to 1.24
Training split only	177	0.59	-0.07 to 1.31
Validation and test split together	39	1.21	0.16 to 2.18

All 19 analyses below the locked primary model are shown. The slope was positive in all of them (0.02 to 1.27 mV/s per mmol/L); the interval excluded zero in 14 and included it in 5 (potassium rises only, falls only, all initial values, initial potassium above 5.0 mmol/L, training split). All analyses use the same pairs and are correlated, and lower limits close to zero (0.01 for the Huber estimator and for the bicarbonate covariate) rest on 500 draws.

Seeds. With the locked seed the lower interval limit was 0.209; with three other fixed seeds 0.195, 0.166 and 0.220; over 44 seeds the mean was 0.208 (SD 0.018, range 0.166 to 0.253), and “positive

above the margin” occurred in 33 of 44 repeats; with 20,000 draws the interval was 0.205 to 1.380. Each of the 44 repeats used 2,000 bootstrap draws with a different seed; the counts refer to repeats of the bootstrap, not to patients.

Heart rate. Heart rate explains little of the slope (change in potassium and in heart rate correlated at 0.03); steepness itself depends on the heart-rate change (-0.03 mV/s per beat per minute, CI -0.06 to -0.004).

Repeatability and scale. See Table S2. In units of the classical limit the steepness slope is 0.18 per mmol/L (CI 0.05 to 0.34), against 0.15 for the amplitude (CI -0.02 to 0.36); with the robust limits (post hoc) 0.36 and 0.50. To compare the two outcomes on one scale we expressed both slopes in four ways (bootstrap correlation of the two slopes 0.58, correlation of the paired changes 0.51):

Scale	Amplitude	Steepness	Difference (steepness minus amplitude), 95% CI
SD within, all samples (per mmol/L)	0.42	0.51	0.09 (-0.40 to 0.50)
SD within, ends of primary pairs	0.58	0.50	-0.07 (-0.75 to 0.46)
Robust SD within, all samples	1.39	1.01	-0.38 (-2.06 to 0.89)
SD of the residual change (per mmol/L)	0.18	0.23	0.04 (-0.16 to 0.23)

The sign of the difference changes with the scale and every interval includes zero. The data fit a response of steepness twice as strong as that of amplitude as well as a weaker one. A steepness that depends mechanically on amplitude (ratio of steepness to amplitude 20.8 per second in the intraoperative database, 21.4 in MC-MED) would give a slope of 0.55 mV/s per mmol/L; the observed 0.75 has an interval that includes 0.55. Adjusted for the change in amplitude, the steepness slope was 0.45 mV/s per mmol/L (CI 0.02 to 0.92); the amplitude slope adjusted for the change in steepness was 0.10 mm per mmol/L (CI -0.13 to 0.40). Both outcomes come from the same T wave, so these adjustments neither prove nor exclude additional information.

Steepness by category of potassium change.

Category of potassium change (mmol/L)	Pairs	Median change (mV/s)
≤ -1.0	33	-0.26
-1.0 to -0.5	28	-0.36
Absolute change below 0.5	132	-0.52
0.5 or more	23	0.24

The median change is at most 0.13 times the steepness repeatability limit in every category. The steepness is associated with the potassium change in the group, but in individual pairs its change is small against the noise between windows of the same sample. The steepness measure has not been checked against a reference reading.

S8. Timing of the ECG segments

Table S9. Distance of the selected ECG segment to the order and the result time (minutes)

End of the pair	Quantile	Segment midpoint minus order time	Segment midpoint minus result time
First	5th	0.0	-152
First	25th	0.0	-86
First	median	0.6	-62
First	75th	21	-45
First	95th	74	-15
Second	5th	-11	-200
Second	25th	0.0	-116
Second	median	0.0	-84
Second	75th	0.0	-62
Second	95th	0.3	-46

The segment lay before the order time in 14% of first ends and 22% of second ends. The selection rule picks the segment nearest to the order time, so most segments sit at the order time, and the result arrives about 1 to 1.5 hours later. The sampling time lies between the two and is not documented. Every positive statement carries the sentence: sampling time not documented (only order and result); the finding applies to the ECG nearest to the order time within the window from 15 minutes before the order to 15 minutes after the result. The analysis restricted to an ECG within 15 minutes of the order time at both ends (146 pairs; Table S4) gave 0.18 mm per mmol/L (-0.22 to 0.70); the two windows then provide the same measurement by construction.

S9. Reference reading

Material. 100 ten-second strips (25 mm/s, 10 mm per mV) drawn at random from the training and validation splits (the test split was not part of the selection), one sample per patient, 10 strips shown twice; the draw was stratified to about 70% strips with a positive and 30% with a negative automatic QRS direction and used only segments that qualify as ends of potassium pairs. The reader, an ECG-experienced person who is the author, who specified the algorithm and knew the hypothesis, marked for three consecutive beats the isoelectric baseline (PR or TP) and the T peak (the deepest end for a negative T) and gave the QRS direction. The reader did not know potassium, treatment or the algorithm's values. 89 strips were measured and 11 marked as not assessable (flat T wave, artefact). The algorithm read the 60-second window around the strip midpoint, which contains the whole strip; at the strip level the algorithm returns exactly the values of the main run.

Table S10. Agreement of the reader and the algorithm for T amplitude, difference reader minus algorithm

Set	Readings	Bias, mV (95% CI)	SD of the difference, mV	Limits of agreement, mV	ICC (1,1)	Spearman
Raw signed amplitude, all	70	+0.040 (+0.016 to +0.064)	0.100	-0.157 to +0.236	0.84	0.90
Raw, QRS positive	50	+0.041 (+0.008 to +0.075)	0.117	-0.188 to +0.271	0.81	0.89
Raw, QRS negative	20	+0.036 (+0.020 to +0.051)	0.033	-0.029 to +0.100	0.96	0.90
QRS-relative (endpoint), all	70	+0.019 (-0.006 to +0.045)	0.106	-0.189 to +0.228	0.94	0.945
Raw, one value per strip	more than 60	+0.034 (+0.012 to +0.057)	0.090	-0.142 to +0.211	0.87	0.91

The 70 readings come from more than 60 strips (strips shown twice enter twice in the main rows); the 89 measured strips are the starting point and 19 had no valid algorithm result. Approximate 95% confidence intervals of the limits of agreement for the raw amplitude (Bland and Altman formula, own arithmetic, ignoring the repeated strips) are -0.20 to -0.12 mV for the lower and +0.20 to +0.28 mV for the upper limit. The intraclass correlation is ICC (1,1); for two fixed methods ICC (A,1) or a concordance coefficient would be more usual, and the QRS-relative row is inflated by the variance between the two polarity groups. Mean raw amplitude was 0.290 mV for the reader and 0.250 mV for the algorithm; the median difference was +0.029 mV and the median absolute difference 0.036 mV (0.36 mm); 60% of the strips lay within ± 0.05 mV. The reader's direction of the QRS equalled the automatic polarity in all strips in which both gave one. The reader's own repeat (strips shown twice; fewer than 10 pairs measurable twice, so orienting only) had a within-reader SD of 0.019 mV. The scatter of the difference is more than five times that, so reading noise alone does not explain it. In fewer than 10 strips, all with a positive QRS, the difference exceeded 0.15 mV, and the algorithm had the opposite sign in some of them; these few strips determine the limits of agreement.

The algorithm was invalid in 19 of the 89 measured strips (21%), mostly because of an irregular rhythm in the 60-second window. These strips were readable for the reader. The selection of the

primary pairs therefore runs through rhythm regularity as well as signal quality. In the 11 strips that the reader marked as not assessable, the algorithm's T amplitudes were small (median about 1.0 mm, against 2.6 mm in the measured strips with a valid algorithm result). The comparison covers amplitude only; there is no reference reading for steepness, and one reader, one session and 10-second strips against a 60-second median beat mean that the two methods do not measure the same beats. The reader is the author, who also specified the algorithm: the comparison is a description, not an independent validation. There was no second reader and no repeat reading after an interval.

S10. Other questions of the specification

The locked specification contained two further questions, which we report for completeness.

QRS duration against bicarbonate at potassium of 6.5 mmol/L or more (serum, first sample per patient, 194 patients). The slope of QRS duration on bicarbonate, adjusted for potassium and age, was 0.23 ms per mmol/L (CI -0.75 to 1.07). The expectation, from an earlier analysis in another database, was a negative slope of -0.58 ms per mmol/L, with a margin of 0.29 ms. The mean QRS duration was 104 ms in the lowest bicarbonate tertile (67 patients), 103 ms in the middle (66) and 107 ms in the highest (61). The result is "imprecise". Bicarbonate is a surrogate and not proof of acidosis.

Persistently raised against rapidly normalising potassium (initial value of 6.0 mmol/L or more without documented treatment). The persistently raised group had fewer than 30 patients (23) against 135 in the rapidly normalising group, below the pre-specified minimum of 30 per group. The question was therefore not assessable, and no inference is drawn.

S11. Characteristics of primary and other pairs (post hoc, descriptive)

Post hoc, after the locked results were seen; descriptive, without tests or intervals. The tables describe patients: one row per patient and group, the first pair of the patient in that group (ordered by the order time of the first value); a patient can appear in several groups. Sampling time is not documented (only order and result); ECG information refers to the ECG nearest to the order time within the window from 15 minutes before the order to 15 minutes after the result, and the interval between samples is the interval between order times. The test split was read, because all splits belong to the locked primary set. Triage, disposition, length of stay and diagnoses were loaded from the source file of the visits for groups A, B0, C* and D* only.

Groups.

Group	Definition
A	primary pairs, the 216 pairs of the locked primary selection
B	valid pairs outside the primary selection (5,585 minus 216 = 5,369)
C	pairs with documented treatment (category “documented” of the specification), regardless of ECG; lies in B
D	pairs with an initial potassium of 6.0 mmol/L or more, regardless of ECG and treatment; lies in B
C*	analysis set “documented treatment” (initial potassium below 6.0 mmol/L, full measurement; 105 pairs)
D*	analysis set “initial potassium of 6.0 mmol/L or more without treatment indication” (full measurement; the 108 pairs by which “all initial values” differs from the primary set)
B0	B without C and D: initial potassium below 6.0 mmol/L and no documented treatment (unclear treatment included), without the full measurement

C and D overlap; C and D lie in B; A is disjoint from C and D; C* lies in C and D* in D.

Table S11. Characteristics of the groups (patient level)

Characteristic	A primary pairs	B valid pairs without primary selection	C documented treatment (all valid pairs)	D initial potassium 6.0 or more (all valid pairs)	C* documented treatment (analysis set)	D* initial potassium 6.0 or more (analysis set)	B0 B without C and D
Pairs (n)	216	5,369	1,660 (rounded down)	1,070 (rounded down)	105	108	3,090 (rounded down)
Patients (n, rounded down)	200	4,050	1,290	900	100	100	2,520
Age, years: median (Q1 to Q3)	60 (39 to 73)	64 (49 to 77)	63 (49 to 75)	66 (52 to 78)	56 (38 to 67)	60 (42 to 78)	64 (49 to 77)
Male sex	61%	52%	52%	52%	55%	56%	52%
Initial potassium, mmol/L: median (Q1 to Q3)	4.2 (3.8 to 5.1)	4.4 (3.8 to 5.7)	4.4 (3.4 to 6.0)	6.4 (6.2 to 7.1)	3.7 (2.9 to 4.7)	6.4 (6.1 to 7.1)	4.2 (3.9 to 4.7)
Potassium change, mmol/L: median (Q1 to Q3)	-0.20 (-0.53 to 0.10)	-0.20 (-0.80 to 0.20)	-0.10 (-0.70 to 0.40)	-1.70 (-2.40 to -0.80)	0.10 (-0.40 to 0.80)	-1.90 (-2.60 to -1.50)	-0.10 (-0.40 to 0.20)
Potassium change, mmol/L: SD	0.70	1.05	1.05	1.26	0.87	1.12	0.61
Potassium rise	28%	32%	40%	5%	50%	<10%	33%
Potassium fall	62%	60%	53%	95%	45%	>90%	56%
Potassium unchanged	10%	8%	7%	<2%	<10%	<10%	10%
Interval between samples (order times), hours: median (Q1 to Q3)	2.6 (1.7 to 3.8)	4.0 (2.4 to 6.3)	4.0 (2.7 to 6.0)	2.1 (1.5 to 3.1)	3.3 (2.2 to 4.0)	1.7 (1.4 to 2.1)	4.7 (2.9 to 6.9)
Heart rate at the first end available	100% (by definition)	43%	50%	59%	100% (by definition)	100% (by definition)	35%
Heart rate at the first end, per minute: median (Q1 to Q3)	91 (78 to 104)	87 (73 to 104)	89 (73 to 105)	82 (69 to 99)	98 (81 to 109)	83 (73 to 101)	88 (74 to 104)
QRS duration at the first end available	79%	21%	24%	27%	73%	84%	17%
QRS duration at the first end, ms: median (Q1 to Q3)	93 (78 to 116)	92 (78 to 120)	94 (78 to 118)	94 (78 to 122)	92 (76 to 116)	86 (78 to 112)	92 (80 to 125)
ECG segment at both ends within the window	100% (by definition)	16%	20%	35%	100% (by definition)	100% (by definition)	8%
Valid measurement at both ends	100% (by definition)	8%	12%	19%	100% (by definition)	100% (by definition)	2%

Percentages refer to the patients of the group in whole percent. “<x%” or “>y%” means that the numerator or its complement is below 10 patients; where such a share appears in a split (rise, fall, unchanged), the other shares of that split are rounded to 5%. Patient counts are rounded down to tens. “Unchanged” is a change of absolute size below 1e-9. QRS duration is an addition and exists

only where the measurement at the first end was valid. Heart rate at the first end is not a measure of coverage: it exists only where the first end was measured, and its medians in B, C, D and B0 describe that measured subset. Characteristics with fewer than 10 patients are not reported.

Table S12. Main reason for exclusion in group B (first unmet step of the flow, pairs)

First unmet step	Pairs	Share of B
No ECG segment at both ends within the window	4,615	86.0%
ECG present, measurement invalid at one end	378	7.0%
Polarity uncertain or different	70	1.3%
Overlapping windows, missing heart rate, or documented or unclear treatment (combined as in Table S1)	198	3.7%
Initial potassium of 6.0 mmol/L or more (otherwise without exclusion)	108	2.0%

Table S13. Treatment classes in C and C* (pairs with at least one administration or order of the class in the treatment window; classes overlap)

Treatment class	C documented treatment (all valid pairs)	C* documented treatment (analysis set)
Potassium-lowering (insulin, glucose, bicarbonate, albuterol, binders)	60%	60%
Intravenous calcium	35%	<15%
Potassium supplement	40%	60%
Dialysis order	5%	<10%

Shares are rounded to 5% because the number of pairs in C* (105) is public.

Table S14. Triage, disposition, length of stay and diagnoses in A, B0, C* and D* (patient level; source file of the visits)

Characteristic	A primary pairs	B0 B without C and D	C* documented treatment	D* initial potassium 6.0 or more
Triage 1 (resuscitation)	<5%	2%	<10%	<10%
Triage 2 (emergent)	45%	39%	45%	30%
Triage 3 (urgent)	50%	57%	50%	65%
Triage 4 and 5 or missing	<5%	1%	<10%	<10%
Discharged home	35%	12%	25%	45%
Admitted to a ward	40%	59%	40%	30%
Intensive care	10%	9%	15%	<10%
Observation or missing	20%	21%	20%	20%
Length of stay available	>95%	95%	>90%	>90%
ED length of stay (unit of the source file, consistent with hours): median (Q1 to Q3)	7.7 (6.0 to 9.5)	10.6 (7.6 to 16.7)	8.0 (6.3 to 10.7)	6.4 (5.0 to 7.8)
ED diagnosis chronic kidney disease or dialysis (N18, Z99.2)	<5%	<1%	<10%	<10%
ED diagnosis acute kidney injury (N17)	<5%	3%	<10%	<10%
ED diagnosis diabetes (E10 to E14)	<5%	1%	16%	<10%
ED diagnosis heart failure (I50)	<5%	3%	<10%	<10%
ED diagnosis hyperkalaemia (E87.5)	<5%	<1%	<10%	<10%

In the small groups A, C* and D* the splits of triage and disposition are rounded to 5%. Diagnoses are the emergency department diagnoses of the visit (mostly one main diagnosis); they are no history of comorbidity, so the shares are lower bounds of disease frequency, and a missing diagnosis counts as absent (fewer than 1% of patients). Observation is counted with missing disposition. The length of stay is missing in a part of the patients of B0. Groups B, C and D were not supplemented.

Reading. 1. The primary selection is small and determined by the availability of the monitor ECG: of the 5,369 other valid pairs, 4,615 lacked an ECG segment at one or both ends within the window (Table S12). 2. Compared with B0, the primary pairs were younger (median 60 vs 64 years), more often male (61% vs 52%), had shorter intervals between samples (median 2.6 vs 4.7 hours), were admitted less often (about 40% vs 59%), discharged more often (about 35% vs 12%) and stayed shorter in the emergency department (median 7.7 vs 10.6). Median initial potassium (4.2 mmol/L) and median potassium change were close (-0.20 vs -0.10 mmol/L). Whether these differences change the slope was not examined. 3. C* (documented treatment) is not a group of hyperkalaemia treatments: the median initial potassium was 3.7 mmol/L (2.9 to 4.7), half of the patients had a potassium rise, and by the treatment classes about 60% of the pairs had a potassium supplement

and about 60% a potassium-lowering drug class (insulin, glucose, bicarbonate, albuterol, binders); intravenous calcium about 10%. It is mainly a group of low potassium values with potassium given, and insulin, glucose or albuterol given for other reasons. Diabetes as an ED diagnosis in about 16% and intensive care in about 15% fit this mixture, not a hyperkalaemia group (hyperkalaemia as ED diagnosis below 10%). The slope in this group must not be read as an effect of hyperkalaemia treatment. 4. D* (initial potassium 6.0 mmol/L or more without treatment indication; median initial potassium 6.4 mmol/L) almost always shows a potassium fall (above 90%; median -1.9 mmol/L) over a short interval (median 1.7 hours); in A potassium fell in 62%. Only about 10% of the pairs with a high initial potassium reached the measurement (D* of D), and entry into the analysis depended on the ECG. The cause of the fall (undocumented treatment, haemolysis or spontaneous course) is unexamined. D* was relatively mild (discharge about 45%, triage 3 about 65%, shortest stay, median 6.4); the queried diagnoses occur in fewer than 10% each. 5. Diagnoses are single main diagnoses of the visit; low shares for kidney disease and heart failure say nothing about the history of the patients.

S12. Reporting

The study followed STROBE and RECORD [12,13]; the completed checklists will be uploaded at submission as the research checklist. Repeatability is reported according to GRRAS [19] where applicable. No patient identifiers, signals or sample-level data are shown. Analysis code will be released as described in the main text.