Supplemental Digital Content

Supplemental Methods

Study Population

The inclusion and exclusion criteria for cohorts are specified in Fig. S1. Our focus was COVID-19 patients that were admitted to the hospital floor. Consistent with clinical practice, COVID-19 diagnosis was based on a 'positive' or 'inconclusive' result from an RT-PCR test. . Patients were included in the study cohorts even if COVID-19 diagnosis occurred after the decision to admit the patient (but still within the hospitalization) because the model could still make predictions on patients under investigation prior to RT-PCR confirmation. To build and assess the models, we omitted those directly admitted to the ICU because the focus of this work was clinical progression among patients not immediately requiring ICU-level care. In addition, patients who were treated in surgical care units were removed to avoid confounding signals related to the procedure itself (e.g., a patient admitted to the ICU for serial monitoring post surgery instead of clinical management; see Tables S4 and S5 for documented procedures of omitted surgical patients). Finally, we excluded a small number of patients that were admitted to the floor but were later transferred to the ICU without meeting criteria for acute respiratory failure (e.g., a COVID-19 patient admitted to the floor transferred to the ICU due to need for temporary dialysis or frequent neurologic assessments but with stable respiratory status). There are no step-down units at UW. Step-down units in the Mid-Atlantic and Pneumonia cohorts were treated as non-ICU locations.

The Pneumonia cohort includes patients who presented to the emergency department, were admitted with a diagnosis of pneumonia, and for whom viral or bacterial pneumonia was also listed as a diagnosis at the time of discharge. This cohort represents a population of hospitalized patients at risk for ARF due to respiratory infections other than SARS-CoV-2.

For each included encounter, we extracted a comprehensive collection of vital signs and laboratory test results from the electronic health record. Tables S1, S2, and S3 include a complete list of the markers that we consider in our analysis. In the Mid-Atlantic cohort we did not have chloride and glucose measurements. In the Pneumonia cohort, we did not have lymph %, CRP, LDH, proBNP, or ferritin. We used all observations of these markers between the time of arrival (including time in the emergency department prior to admission) up until the time of discharge. We also collected all respiratory therapies that each patient received, including the type of device (nasal cannula, non-invasive positive pressure ventilation device, and mechanical ventilator) and the level of supplemental oxygen when applicable. Finally, we included the time of discharge or death.

We randomly split the Mid-Atlantic patients into four mutually exclusive groups: a train set (438 encounters, 63 with ARF outcome), tuning set (430 encounters, 55 with ARF outcome), development set (440 encounters, 60 with ARF outcome), and test set (433 encounters, 55 with ARF outcome). We did not split the UW patients into subsets and used that cohort only for validation. Finally, we randomly split the Pneumonia cohort into three distinct groups: a train set (1159 encounters, 48 with ARF outcome), tuning set (1157 encounters, 43 with ARF outcome), and test set (1159 encounters, 47 with ARF outcome).

Outcomes and Predictors

The World Health Organization's criteria for severe COVID-19 includes patients that are ever treated with 15 liters/minute or more of supplemental oxygen, a non-invasive positive pressure device, or mechanical ventilator (15). In our cohorts, many patients were placed on 15 liters/min of oxygen during brief sessions with non-MD providers (e.g., physical therapy). To mitigate false positives associated with the routine brief use of escalated therapy for potentially strenuous activities, we required at least one of the following criteria subsequent to the time when the patient is treated with 15 liters/min or more of oxygen: (1) the patient is treated with 15 liters/minute or more of oxygen for more than 8 consecutive hours; (2) the patient is subsequently escalated to 30 liters/minute of oxygen or more; (3) the patient is escalated to an NIPPV device or mechanical ventilator. If the patient was transiently treated with 15 liters/min or more of oxygen and at least one of the above occurred, we considered the onset time of ARF to be when supplemental oxygen was raised to 15 liters/min or more. If none of the additional criteria were met before oxygen was reduced to less than 15 liters/min, but the additional criteria are met later in the same encounter, we considered the onset time to be the later time supplemental oxygen was raised to 15 liters/min or more or they required NIPPV or mechanical ventilation. The Pneumonia cohort does not contain supplemental oxygen levels, so we used the first time a patient is placed on an NIPPV device or a mechanical ventilator as the onset time of respiratory failure. The timing of changes in supplemental oxygen, including NIPPV devices and mechanical ventilation, was included in data collected from EMR flowsheets and procedures information entered into the EMR.

During model development, we used the onset time of ARF to define labels for each hour. For patients that did not develop ARF, each hour was labeled as a negative sample. For patients that do develop ARF, we labeled all hours within 24 hours prior to onset time as positive

samples and all prior hours as negative samples. We refer to the labels assigned to each hour as "hourly labels".

For all markers, we calculated hourly features starting from time of admission using the complete history of each marker up until that time point. For each marker, we calculated the first value recorded during that encounter, the minimum value prior to that timepoint, the maximum value prior to that timepoint, and the latest value. When a marker was not recorded prior to the timepoint, the values of all trajectory features related to that marker were assigned a distinct non-numeric value indicating missingness. We refer to the predictors computed at each hour as "hourly predictors".

Model Development

We fit all models in this paper using gradient boosted decision trees as implemented in the Python xgboost package (17). Gradient boosted trees require a number of hyperparameters to be set prior to learning. Important hyperparameters include: the number of decision trees to include in the ensemble, the depth of each decision tree, the learning rate (this determines the influence that each new tree in the ensemble has on the final output), and the minimum number of samples required to form new leaves during the splitting operation. We used 10-fold cross validation on the train splits and a grid search to determine the hyperparameter configuration that gave best performance. We used the tuning sets to monitor AUC as the model trained. We used "early stopping" to stop the boosting process once the AUC on the tuning set stopped improving for 10 rounds. After training, we performed error analysis and refined our approach, including outcome definition, choice of model inputs, feature representation of inputs, model architecture,

and tuning parameters, using the Mid-Atlantic and Pneumonia development sets. We did not examine the test sets until generating the final results presented.

We constructed training samples by subsampling each encounter. We found that including all samples from each encounter caused the model to overfit (the samples collected from a single encounter are strongly correlated). For each negative encounter, we randomly selected a single hour from the encounter and used the hourly predictors at that time as model input and the hourly label as the supervised output. For each positive encounter, we randomly selected a single positive hour (i.e. one within 24 hours of ARF onset), and, if the patient was hospitalized for more than a day, a single negative hour (i.e. one outside of the 24 hour window prior to ARF).

For each marker, we fit three different models. First, we fit a model that uses hourly predictors for all features in the Mid-Atlantic train and tuning sets. Second, we fit a model that uses only the latest value of the marker as input on the Mid-Atlantic train and tuning sets. Finally, to assess the performance and distinguishing characteristics of predictive models specifically derived in patients with COVID-19 compared with general models of respiratory infection, we used hourly predictors for all features, but trained on the Pneumonia train and tuning sets. We refer to these models as the full models, latest-value models, and pneumonia models respectively. We also trained three models that use hourly predictors for all markers: a model trained on the Pneumonia cohort, a model trained on the Mid-Atlantic cohort using markers available in the Pneumonia cohort, and the final ARC model using all markers available in the Mid-Atlantic cohort. Tuning parameters and cross validation performance for the ARC model are shown in Table S8.

Prognostic Value of Markers and Trajectory Data

We evaluated the discriminative power of each individual marker using the area under the receiver operating characteristic curve (ROC AUC) at the encounter level. We used the models to make hourly predictions, which are scores between 0 and 1 (with values closer to 1 indicating higher risk). For patients that did not develop ARF, we calculated the maximum score for the full encounter. For encounters that did develop ARF, we calculated the maximum score up to 1 hour before the onset of respiratory failure (that is, all true positive encounters were truncated 1 hour prior to onset of ARF). We used the maximum scores for all encounters to compute the ROC AUC. The AUC point estimates and standard errors were computed using bootstrap resampling of encounters (after the models have been fit).

For all validation experiments, we excluded encounters where the patient was still in the hospital at the time of data collection and did not experience ARF outcome (because we could not rule out that the patient experienced ARF later in their hospitalization). We also excluded encounters where the first hourly prediction after admission was already less than 1 hour before ARF onset or after the patient was already discharged from the hospital. After applying these additional exclusions, the pooled validation cohort included 408 encounters from the Mid-Atlantic test set (50 with ARF outcome) and 261 encounters from the UW cohort (25 with ARF outcome).

To demonstrate the importance of incorporating the history and trajectory of each marker instead of the latest value only, we also calculated the ROC AUC for each marker using the latest-value models. We used the same procedure as the full models to compute the maximum score for each encounter.

Comparison to ARF in Other Types of Pneumonia

To characterize similarities and differences in the specific markers predictive of progression to ARF in COVID-19 versus other forms of respiratory infection, we compared the discriminative power of each marker using the COVID-19 trajectory models and the pneumonia models. In this analysis, we did not include Lymphocyte %, CRP, LDH, Pro-BNP, or ferritin because they were not available in the Pneumonia cohort. We validated both models using the Mid-Atlantic test data, and calculated AUC with bootstrapped standard error estimates using the same encounter-level procedure as above. We also developed two combined models using all markers; the first was trained on the Mid-Atlantic train and tune splits, and the second was trained on the Pneumonia train and tune splits. We validated both models on the Mid-Atlantic test data and the Pneumonia test data by measuring ROC AUC and area under the precision recall curve (PR AUC). Predictions were made for all patients in the Pneumonia cohort, not just for patients who were diagnosed with pneumonia prior to admission.

Assessing Marker Alert Timing Relative to ARF

To assess timeliness of each marker, we used the output of the full models to define a clinical alert and measured the time between the first alert and the onset of ARF. To ensure sufficient samples, we evaluated timeliness only for markers with an AUC above 0.60 (as measured by the encounter-wise AUC calculations above). We only measured timeliness on true positive patients that did not have an alert at admission because these patients may have already developed severe COVID-19 at time of presentation.

To define an alert for each trajectory model, we chose a score threshold that correctly excluded 75% of the encounters that did not develop ARF (i.e. a 75% specificity). For all true positive encounters (defined as encounters where a given model score has a maximum score

above the 75% specificity threshold) without an alert on admit, we calculated the *alert onset* as the first time at which the score crosses the threshold. We calculated the time between alert onset and onset of ARF to assess timeliness. This method did not require knowledge of when the maximum risk score was reached during an encounter, thus only used data collected prior to alert onset time.

Prognostic Value of Combining Marker Trajectories

To assess the benefit gained by jointly considering the trajectory of multiple markers simultaneously to determine risk of ARF, we fit the ARC model using the trajectory predictors for all markers as input. To evaluate ARC, we calculated the maximum score for each encounter (using the entire stay for patients without ARF, and the time prior to onset for patients that develop ARF). We plotted the ROC curve and precision-recall curve, and calculated the area under each curve as a summary of performance. We report performance on an encounter-level because the overall performance of the model in predicting ARF ahead of ARF onset is the most clinically relevant performance. Reporting performance of the individual hourly predictions would require unnecessarily complex statistical analysis to account for dependencies between predictions.

To compare to an existing model, we also plotted the ROC and precision-recall curves using the maximum Modified Early Warning Score (MEWS) for each encounter (18). In order to calculate MEWS, we converted the Glasgow Coma Score (GCS) in our data to alert/verbal/painful/unresponsive (AVPU) score using the following conversion: A = GCS 15, V = GCS 11-14, P = GCS 5-10, U = GCS 3-4 (19–21).

We hypothesized that patients who do not meet the criteria for ARF but who receive moderate supplemental oxygen may appear to be at high risk, and so may have a higher rate of false positives than among patients who receive little to no supplemental oxygen. To investigate this hypothesis, we also calculated the ROC AUC using a sequence of increasingly inclusive definitions of patients without ARF. In the least inclusive case, we calculated ROC AUC using patients that develop ARF and those that received no supplemental oxygen. We hypothesized that these two sets of patients will appear most clinically distinct. We then calculated ROC AUC by including patients that received higher levels of supplemental oxygen, but did not meet our criteria for ARF. We hypothesized that patients on higher levels of supplemental oxygen would be increasingly severe and therefore appear more clinically similar to patients that develop ARF.

We used the Python shap package to estimate the effect size of predictors, which allows us to estimate the importance of individual features for each prediction. The score estimates the degree to which a feature contributes to the predicted risk based on concepts from collaborative game theory (22), and has been applied to other applications in biomedicine (23). Using the estimated effect size, we studied the relative importance of markers and their trajectory features in the combined model. We determined the importance of each individual predictor in the model using the average absolute magnitude of the estimated effect size on the test data and visualized the estimated effect sizes of the ten most important predictors.

To characterize the potential clinical impact of ARC, we calculated the times of important events for true positive patients. We calculated alert times using the same technique as when evaluating the timeliness of each marker (i.e. the first time that the model's score rises above a threshold that achieves 75% specificity), the time of ARF onset, and the time of first ICU transfer (if applicable).

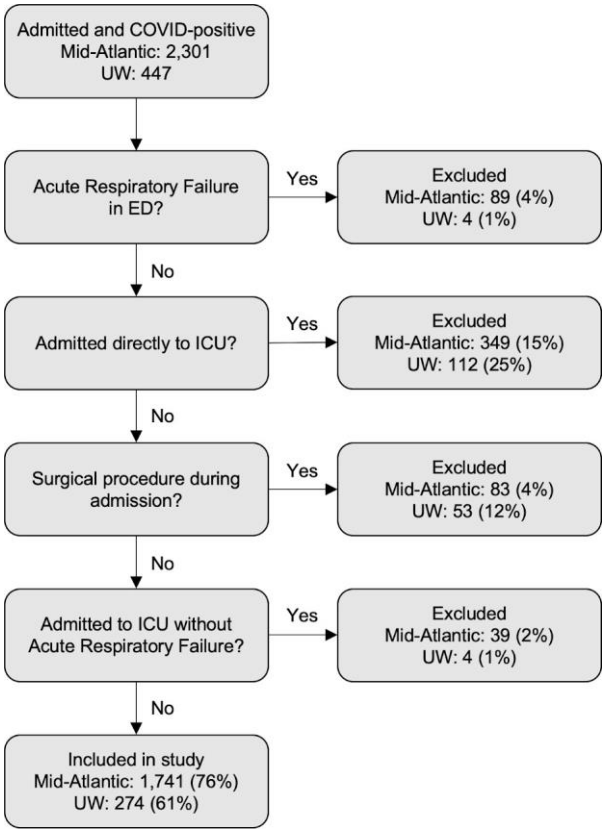

Fig S1. Inclusion criteria for Mid-Atlantic and UW cohorts. Flowchart depicts the criteria for defining the Mid-Atlantic cohorts, including the number (percent) of patient encounters excluded by each criterion.

Table S1. Marker summary statistics for Mid-Atlantic cohort

Marker	Fraction Observed	Mean	Std. Dev.
BP (Systolic)	0.96	123.4	23.07
BP (Diastolic)	0.96	68.56	13.52
Resp rate	0.96	23.93	8.15
Heart rate	0.96	88.45	18.63
Temperature	0.96	98.74	1.47
SpO2	0.96	95.73	4.33
GCS	0.54	10.91	4.24
BUN	0.95	26.46	20.88
Calcium	0.95	8.47	0.72
Chloride			
CO2	0.95	24.85	4.68
Creatinine	0.95	1.48	1.71
Glucose			
Potassium	0.95	4.13	0.59
Sodium	0.95	140.04	6.04
WBC	0.96	9.75	5.76
Lymph %	0.93	14.71	10.63
Platelets	0.96	266.62	140.84
Arterial pH	0.31	7.37	0.08
PaCO2	0.31	49.85	12.99
PaO2	0.12	84.87	32.8
Lactate	0.53	2.22	2.27
Troponin	0.79	0.25	1.59
D-dimer	0.83	4.81	8.97
CRP	0.63	9.51	9.14
LDH	0.64	460.1	393.77
proBNP	0.54	2936.67	10078.17
Ferritin	0.8	1385.92	4584.45

Table S2. Marker summary statistics for UW cohort

Marker	Fraction Observed	Mean	Std. Dev.
BP (Systolic)	0.99	119.54	22.79
BP (Diastolic)	0.99	68.99	14.82
Resp rate	0.99	23.43	7.67
Heart rate	0.99	91.72	18.84
Temperature	0.99	98.46	1.19
SpO2	0.99	94.3	4.94
GCS	0.79	11.55	4.25
BUN	0.85	31.11	22.3
Calcium	0.85	8.51	0.73
Chloride	0.85	101.3	5.82
CO2	0.85	27.5	5.4
Creatinine	0.85	1.48	1.66
Glucose	0.85	142.81	55.97
Potassium	0.85	4.07	0.6
Sodium	0.85	137.12	5.39
WBC	0.86	10.8	6.67
Lymph %	0.73	16.06	11.44
Platelets	0.85	260.95	140.32
Arterial pH	0.1	7.38	0.08
PaCO2	0.1	52.05	13
PaO2	0.1	85.33	58.29
Lactate	0.48	1.87	1.54
Troponin	0.14	0.96	4.38
D-dimer	0.34	3.16	4.66
CRP	0.42	9.73	9.38
LDH	0.42	392.8	232.45
proBNP	0.46	3394.06	8421.49
Ferritin	0.33	1091.95	2063.17

Table S3. Marker summary statistics for Pneumonia cohort

Marker	Fraction Observed	Mean	Std. Dev.
BP (Systolic)	1	125.38	23.25
BP (Diastolic)	1	68.14	14.07
Resp rate	1	20.43	6.78
Heart rate	1	89.08	18.91
Temperature	1	98.42	1.2
SpO2	1	96.19	3.77
GCS	0.54	13.36	2.79
BUN	1	27.03	21.52
Calcium	1	8.51	0.76
Chloride	1	100.74	6.33
CO2	1	25.59	4.91
Creatinine	1	1.59	1.67
Glucose	1	137.03	62.91
Potassium	1	4.12	0.62
Sodium	1	138.72	5.19
WBC	1	11.03	6.39
Lymph %			
Platelets	1	240.16	129.35
Arterial pH	0.24	7.37	0.09
PaCO2	0.25	47.07	13.99
PaO2	0.11	105.52	59.89
Lactate	0.65	2.2	2.03
Troponin	0.65	0.35	3.36
D-dimer	0.08	4.85	7.06
CRP			
LDH			
proBNP			
Ferritin			

Table S4. Surgical procedures performed for patients excluded due to surgery from the Mid-Atlantic cohort

Procedure	Number of encounters
EGD (ESOPHAGOGASTRODUODENOSCOPY)	5
CREATION, TRACHEOSTOMY, CREATION, GASTROSTOMY, PERCUTANEOUS, ENDOSCOPIC	4
HEMIARTHROPLASTY, HIP, BIPOLAR	3
EGD, WITH PEG TUBE OR PEJ TUBE INSERTION	3
SALPINGECTOMY, LAPAROSCOPIC	3
REPLACEMENT, TRACHEOSTOMY TUBE	2
EGD (ESOPHAGOGASTRODUODENOSCOPY), COLONOSCOPY	2
INCISION AND DRAINAGE, LOWER EXTREMITY	2
ERCP (ENDOSCOPIC RETROGRADE CHOLANGIOPANCREATOGRAPHY)	2
LAPAROTOMY, EXPLORATORY	2
CHOLECYSTECTOMY, LAPAROSCOPIC	2
REPAIR, LACERATION, FACE, ORIF, FRACTURE, HUMERUS, ORIF, WRIST	1
LAPAROTOMY EXPLORATORY WITH LYSIS OF ADHESIONS	1
ORIF, FRACTURE, FEMUR	1
AMPUTATION, FOOT, INCISION AND DRAINAGE, LOWER EXTREMITY	1
IRRIGATION AND DEBRIDEMENT, UPPER EXTREMITY	1
APPENDECTOMY, LAPAROSCOPIC	1
HYSTERECTOMY, TOTAL, ABDOMINAL, ROBOT-ASSISTED, WITH BILATERAL SALPINGO-OOPHORECTOMY, ROBOT X PROCEDURE, CYSTOSCOPY	1
VITRECTOMY, USING 25-GAUGE INSTRUMENTS	1
WOUND EXPLORATION	1
INSERTION, ELECTRODE LEAD, TEMPORARY TRANSVENOUS PACING	1
BRONCHOSCOPY, RIGID	1
REPAIR, ANEURYSM, ARTERY, FEMORAL	1
LOWER EXTREMITY IRRIGATION AND DEBRIDEMENT, SKIN GRAFT SPLIT THICKNESS	1
TRACHEOSTOMY, LARYNGOSCOPY DIRECT, BRONCHOSCOPY FLEXIBLE	1
INSERTION, TROCHANTERIC NAIL, FEMUR, PROXIMAL	1
EGD, WITH ENTERAL STENT INSERTION	1

AMPUTATION, FOOT, TRANSMETATARSAL, ANGIOGRAM, LOWER EXTREMITY, WITH ANGIOPLASTY AND STENT INSERTION	1
ORIF, FRACTURE, CLAVICLE	1
REVISION, ARTIFICIAL URINARY SPHINCTER, CYSTOSCOPY	1
EXTREMITY FREE FLAP MICROVASCULAR	1
CYSTOSCOPY, WITH URETERAL STENT INSERTION	1
INCISION AND DRAINAGE, LUMBAR REGION	1
LAPAROTOMY EXPLORATORY	1
ORIF, ANKLE, EXTERNAL FIXATION, LOWER EXTREMITY, IRRIGATION AND DEBRIDEMENT, LOWER EXTREMITY	1
CHOLECYSTECTOMY LAPAROSCOPIC	1
CREATION, TRACHEOSTOMY, CREATION, GASTROSTOMY, PERCUTANEOUS, ENDOSCOPIC, INSERTION, GASTROSTOMY TUBE, LAPAROSCOPIC	1
RECTAL ABSCESS IRRIGATION & DEBRIDEMENT (I&D)	1
AMPUTATION, ABOVE KNEE	1
ARTERIOVENOUS FISTULOGRAM WITH VENOUS ANGIOPLASTY	1
MICROLARYNGOSCOPY, SUSPENSION, SIMPLE, BRONCHOSCOPY, RIGID, DILATION, LARYNX AND TRACHEA	1
LAMINECTOMY, DECOMPRESSIVE, SPINE, LUMBAR, WITH FUSION USING INSTRUMENTATION	1
INCISION AND DRAINAGE, WOUND, TORSO	1
THORACENTESIS INTERVENTIONAL	1
IRRIGATION AND DEBRIDEMENT, SHOULDER	1
FISTULOGRAM, ARTERIOVENOUS	1
CLOSED REDUCTION, HIP, WITH PERCUTANEOUS PINNING	1
BRONCHOSCOPY, BRONCHOSCOPY, WITH BRONCHOALVEOLAR LAVAGE, BRONCHOSCOPY, WITH TRANSBRONCHIAL BIOPSY	1
ULTRASOUND, UPPER GI TRACT, ENDOSCOPIC	1
ULTRASOUND, UPPER GI TRACT, ENDOSCOPIC, WITH ANESTHETIC AGENT INJECTION AROUND CELIAC PLEXUS OR CELIAC PLEXUS NEUROLYSIS, FINE NEEDLE ASPIRATION BIOPSY, WITH US GUIDANCE	1
SIGMOIDOSCOPY, FLEXIBLE	1
ENDARTERECTOMY, FEMORAL	1
BRONCHOSCOPY	1
PROSTATECTOMY, RETROPUBIC, RADICAL	1
INSERTION, CATHETER, HEMODIALYSIS, TUNNELED	1
DILATATION AND EVACUATION (D&E)	1

WOUND CLOSURE	1
MICROLARYNGOSCOPY, SUSPENSION	1
FREE FLAP REVISION/RETURN, FOOT DEBRIDEMENT, SKIN GRAFT SPLIT THICKNESS	1
CREATION, TRACHEOSTOMY	1
ORIF, FRACTURE, PHALANX, FINGER	1
ANGIOGRAM, VISCERAL VESSEL, SELECTIVE	1
AMPUTATION, TOE, INCISION AND DEBRIDEMENT, FOOT	1
BRONCHOSCOPY, WITH TRACHEOSTOMY CREATION AND PEG TUBE INSERTION	1
INCISION AND DEBRIDEMENT, HAND, ORIF, FRACTURE, HAND	1
CORPECTOMY, SPINE, CERVICAL, ANTERIOR APPROACH, WITH INSTRUMENTATION	1
COLONOSCOPY CASE CART	1
CREATION, CHOLECYSTOSTOMY, PERCUTANEOUS	1
FASCIOTOMY, DECOMPRESSIVE, FOR COMPARTMENT SYNDROME	1
CREATION, TRACHEOSTOMY, LARYNGOSCOPY, DIRECT, BRONCHOSCOPY, FIBEROPTIC	1
HEMICOLECTOMY, RIGHT, LAPAROTOMY	1
COLECTOMY, LAPAROSCOPIC, EXAM UNDER ANESTHESIA, RECTUM, INSERTION, SETON STICH, FOR ANAL FISTULA, INCISION AND DRAINAGE, ABSCESS, PERIRECTAL	1
LAPAROTOMY	1
COLONOSCOPY	1
INCISION AND DRAINAGE, UPPER EXTREMITY	1
CLOSED REDUCTION, HIP	1
REMOVAL, CATHETER, HEMODIALYSIS, TUNNELED	1
EYE RUPTURED GLOBE REPAIR	1
HYSTERECTOMY, TOTAL, ABDOMINAL	1
ARTHROPLASTY, KNEE, TOTAL	1
HYSTERECTOMY, TOTAL, ABDOMINAL, ROBOT-ASSISTED, WITH STAGING AND LYMPHADENECTOMY, ROBOT X PROCEDURE	1
HYSTERECTOMY, TOTAL, ABDOMINAL, LAPAROTOMY, EXPLORATORY, PEDIATRIC	1
EGD (ESOPHAGOGASTRODUODENOSCOPY) CASE CART	1
PARATHYROIDECTOMY	1
SIGMOIDOSCOPY FLEXIBLE, EGD (ESOPHAGOGASTRODUODENOSCOPY)	1
CLOSED REDUCTION, ELBOW	1

DILATION AND EVACUATION, UTERUS, EXAM UNDER ANESTHESIA	1
REPLACEMENT, GASTROSTOMY TUBE	1
REPAIR, ARTERY, FEMORAL	1
AMPUTATION, BELOW KNEE	1
AMPUTATION, TOE	1
DISCECTOMY, DECOMPRESSIVE, SPINE, CERVICAL, POSTERIOR APPROACH, WITH FUSION	1
EXAM UNDER ANESTHESIA, CLOSURE, WOUND, ABDOMINAL	1
RIGHT HEART CATHETERIZATION	1
INSERTION, GASTROSTOMY TUBE, PERCUTANEOUS	1
EXPLORATION, WOUND, LOWER EXTREMITY	1
INCISION AND DEBRIDEMENT, FOOT, SKIN GRAFT SPLIT THICKNESS	1
ORIF, WRIST	1
EGD, WITH PEG TUBE INSERTION	1

Table S5. Surgical procedures performed for patients excluded due to surgery from the UW cohort

Procedure	Number of encounters
CYSTOSCOPY EVAC FULGARATION	1
CYSTOLITHOLAPAXY / LASER LITHOTRIPSY / SUPRAPUBIC TUBE EXCHANGE (CYSTOSCOPY/ URETEROSCOPY/ LASER LITHOTRIPSY (ENDO)	1
RIGHT UPPER EXTREMITY FISTULAGRAM via FEMORAL VEIN PUNCTURE (ANGIOGRAM/ AORTOGRAM/ ANGIOPLASTY (VASC))	1
MICRO DIRECT LARYNGOSCOPY, BRONCHOSCOPY (MICROLARYNGOSCOPY, BRONCHOSCOPY/TRACH)	1
LEFT ANKLE ORIF	1
RIGHT TIBIAL NAIL (TIB/FIB IM NAIL (INS/EX/DYNAM))	1
L1 CORPECTOMY (SPINE THORAC-LUMB LATERAL ANTERIOR APPROACH (ORTHO)	1
RIGHT TRANSMETATARSAL AMPUTATION (FOOT/ANK AMPUTATION (W/ STSG))	1
RIGHT LEG I D (TIB/FIB SOFT TIS/I&D/STSG/ VAC I&D WOUND LOWER EXTREMITY)	1
REVISION LOOP COLOSTOMY / POSS ENDOSCOPIC COLOSTOMY + LAPAROSCOPY + LAPAROTOMY (LAPAROSCOPIC COLECTOMY)	1
C2 PSIF + C1 LAMINECTOMY (SPINE CERVICAL POSTERIOR (NEURO) w/ & w/o USS)	1
UNILATERAL INTERHEMISPHERIC COLLOID CYST RESECTION APPROACH / POSS EXTERNAL VENTRICULAR DRAIN (CRANIOTOMY / RESECT TUMOR)	1
CRPP POSTERIOR PELVIC RING (ACET/PELV/HIP PINS / SI SCREWS / EXT FIX ACETABUL)	1
LEFT CAROTID ENDARTERECTOMY (CAROTID ENDARTERECTOMY)	1
PARS PLANA VITRECTOMY/ TUBE EXPLANT (PARS PLANA VITRECTOMY /MEMBRANE/SILICONE (incl COR)	1
T11 - L3 PSIF / T12 - L1 LAMINECTOMY POSS COSTOTRANSVERS (SPINE THORAC-LUMB POST without USS (ORTHO))	1
C2-T5 PSIF REVISION (SPINE THORAC-LUMB POST w/ FUSION / USS (ORTHO))	1
APPENDECTOMY LAPAROSCOPY# (APPENDECTOMY LAPAROSCOPY)	1
OPEN GASTROSTOMY (EXPLORATORY LAPAROTOMY)	1
I D RIGHT LEG + VAC CHANGE (TIB/FIB SOFT TIS/I&D/STSG/ VAC I&D WOUND LOWER EXTREMITY)	1
LIVER TRANSPLANT (LIVER TRANSPLANT)	1
LEFT BKA STUMP WASHOUT WITH POSSIBLE REVISION (AMPUTATION--U.E./L.E. (VASC))	1
ORIF RIGHT HIP/ POSS ORIF RIGHT HUMERUS (HUMERUS--DISTAL--ORIF / PERC	1

PIN / BONE GRAFT)	
LEFT ABOVE-KNEE AMPUTATION (AMPUTATION--U.E./L.E. (VASC))	1
LEFT LOWER EXTREMITY FEMORAL TIBIA BYPASS WITH VEIN (BYPASS GRAFT FEMORAL-TIBIAL (VASC))	1
FEMUR IM NAIL / RIGHT (FEMUR IM NAIL (INS/EXCH/DYNAM))	1
EXPLORATION OF FUSION / L2-L3 TLIF / EXTENSION OF FUSION TO T11 / T10 VERTEBROPLASTY (SPINE THORAC-LUMB POST w/ FUSION / USS (NEURO))	1
I D RIGHT KNEE WITH POLY EXCHANGE (KNEE/PATELLA SOFT TIS/ I&D/VAC)	1
AAP-- PEG TUBE (AAP--ENDOSCOPY--TO BE DONE IN 8MB ENDOSCOPY SUITE)	1
MICRODIRECT LARYNGOSCOPY W/TRACH STOMA REVISION (MICROLARYNGOSCOPY, BRONCHOSCOPY/DILITATION)	1
LEFT LEG ANGIOGRAM WITH INTERVENTION (ANGIOGRAM/ AORTOGRAM/ ANGIOPLASTY (VASC))	1
RIGHT THIGH DPC (FEMUR SOFT TIS/ I&D/ STSG/ VAC/ QUADRIC EPSPLASTY)	1
CYSTOSCOPY; CLOT EVACUATION; BLADDER FULGERATION (BLADDER TUMOR TRANSURETHRAL RESECTION)	1

Table S6. Sampling density for key markers

	Mid-Atlantic	UW
Median time between measurements (hours)		
Supplemental oxygen	2.7	3.7
SpO2	3.8	3.7
Respiratory rate	3.8	4.1
Median time between new data and model prediction (min)	25	29

Table S7. Distribution of time between first alert and ARF onset for true positive encounters

Time interval (h)	Number of true positive alerts (%)
≤ 6	8 (14%)
6-12	3 (5%)
12-24	14 (24%)
24-48	13 (22%)
48-96	14 (24%)
> 96	6 (10%)