Sampling Frequency for Machine Learning to Separate Monitoring Artifact from Instability

Anthony Wertz, MS[1] ~ **Marilyn Hravnak**, RN, PhD, ACNP-BC, FAAN, FCCM[2]
Artur Dubrawski, PhD, MEng[1] ~ **Lujie Chen**, MS[1] ~ **Tiffany Pellathy**, MS, ACNP-BC[2]
Gilles Clermont, MD[3] ~ **Michael R. Pinsky**, MD, MCCM[3]

1. Auton Lab, Robotics Institute, Carnegie Mellon University
2. School of Nursing, University of Pittsburgh
3. School of Medicine, University of Pittsburgh
Pittsburgh, PA, USA

Disclosures

- Funding: NIH R01NR013912
- No commercial conflict of interest

Motivation

We know: Vital sign (VS) data collected every 20s can be used to adjudicate alerts, classifying them as either artifacts or real instabilities.

We asked: Would sampling VS data less frequently impair ability to define real vs. artifact alerts?

Hypothesis:

Models using data sampled less frequently can be used to adjudicate alerts.

Motivation

We know: Vital sign (VS) data collected every 20s can be used to adjudicate alerts, classifying them as either artifacts or real instabilities.

We asked: Would sampling VS data less frequently impair ability to define real vs. artifact alerts?

Hypothesis:
Models using data sampled less frequently can be used to adjudicate alerts.

Why evaluate lower sampling frequencies?

- 20s resolution is often not available in deployed systems.
    - *How far can we downsample without losing clinical utility?*
- Higher data frequencies entail higher collection and storage costs.
- In retrospective analysis existing data repositories may have only low frequency data.

Importance of Artifact Detection

For Clinicians:

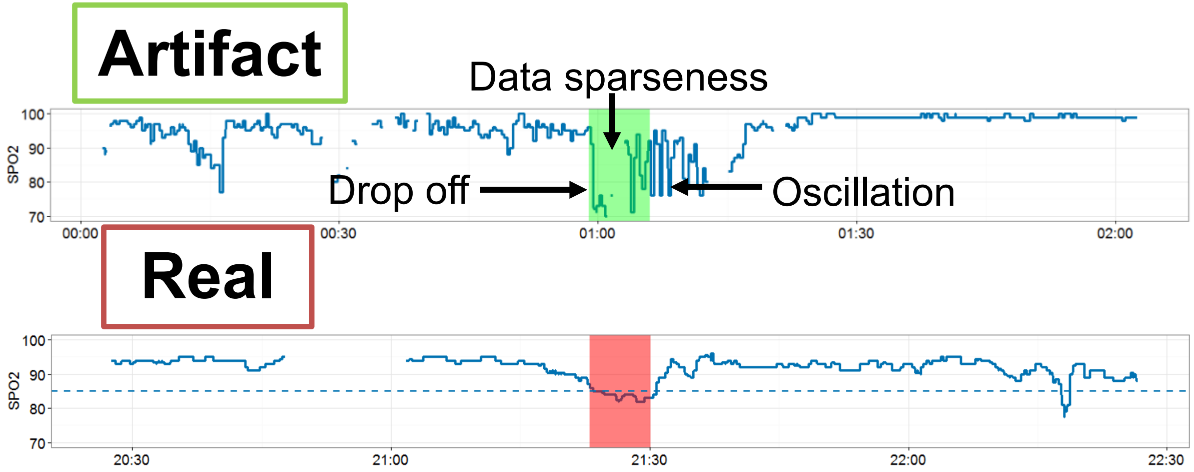
- Treat artifact vs instability differently.
- Delayed response time due to alarm fatigue.

  Bonafide et al. J Hosp Med. 2015; 10(6):345-51

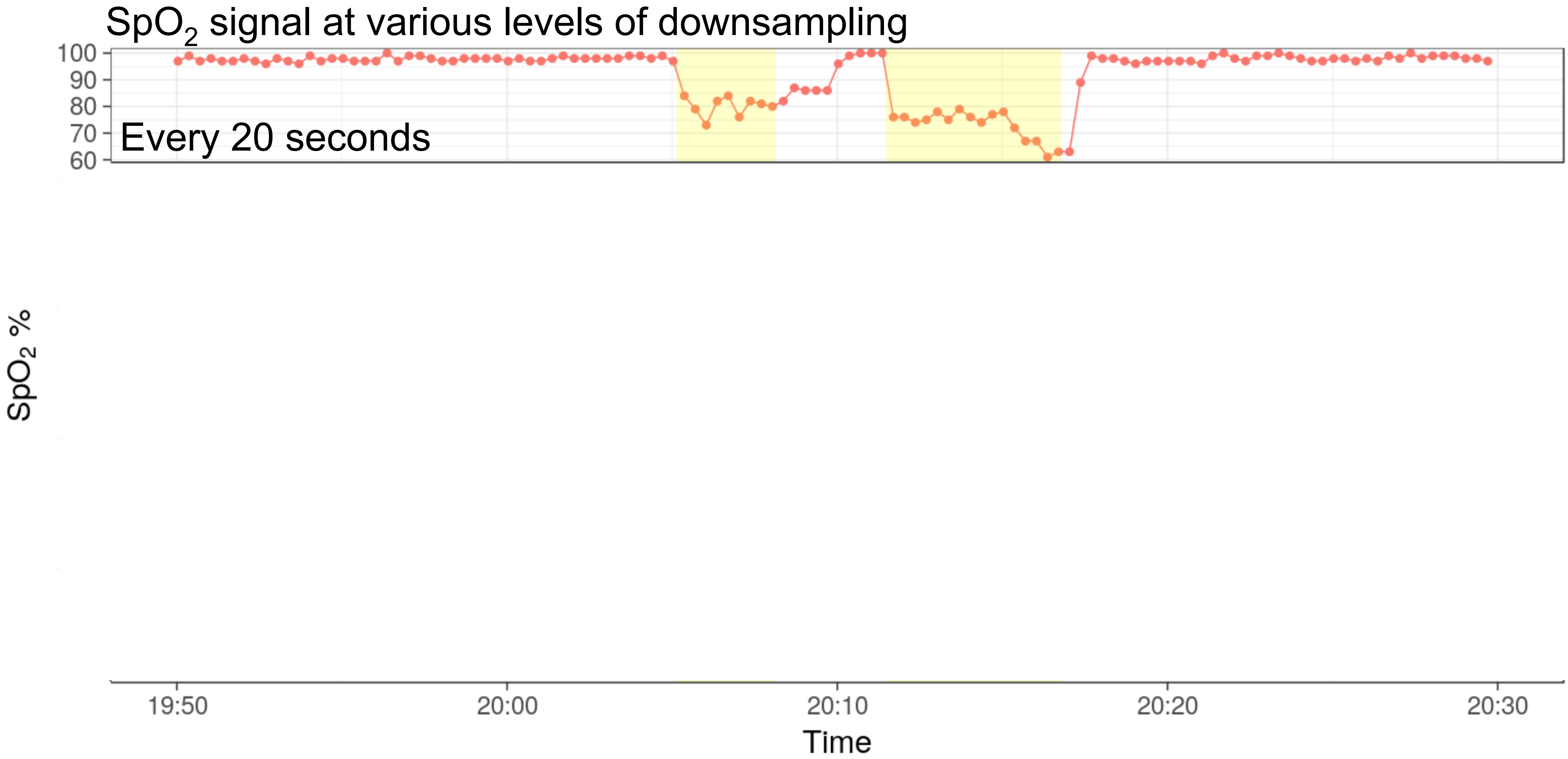
For Modelling Instability:

- More difficult in the presence of artifacts.
- Might end up modeling the artifact instead

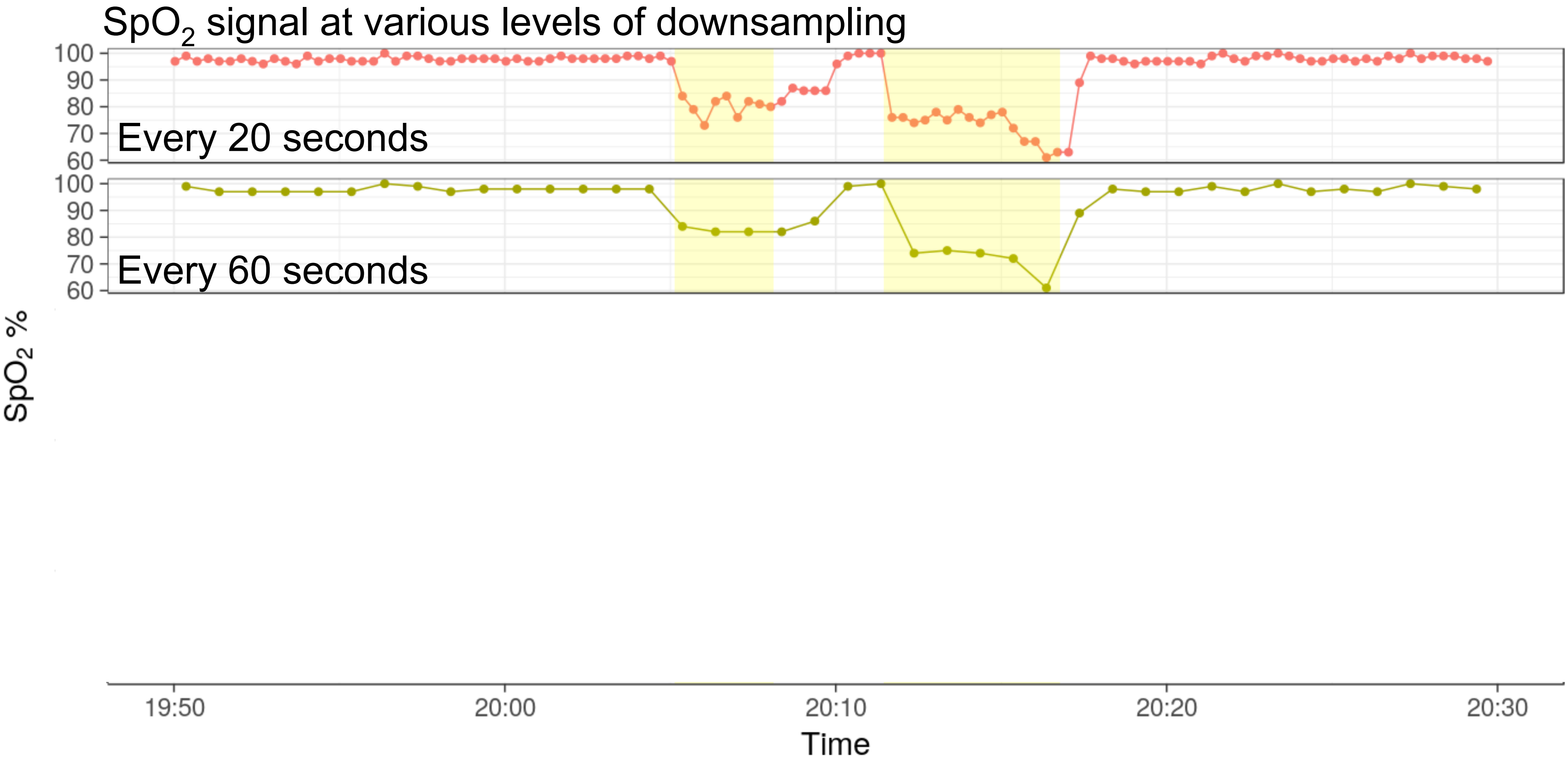
Alerts versus Artifacts



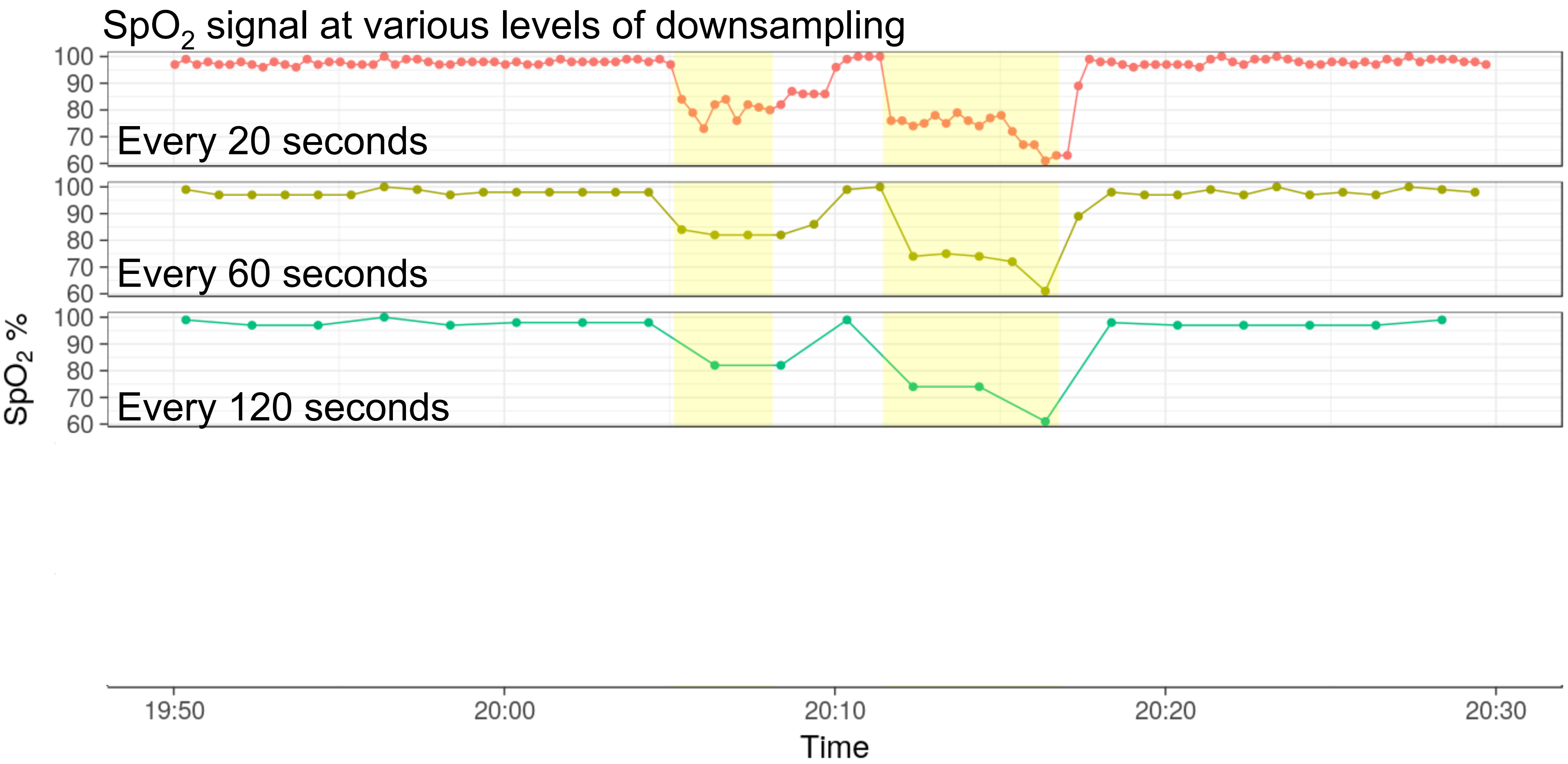
Original Vital Sign Data Collected Every 20s



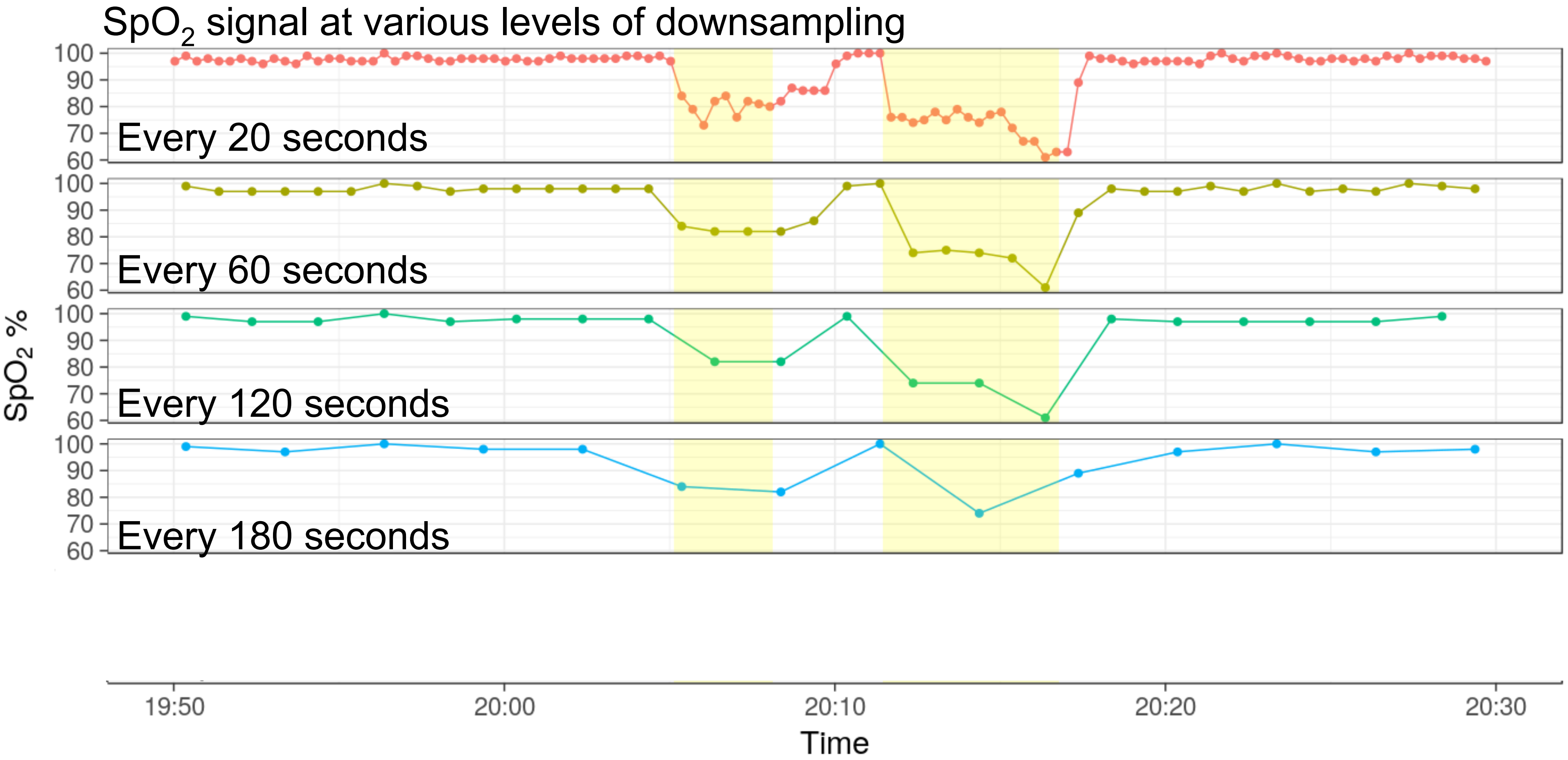
Downsampling Reduces Number of Observations



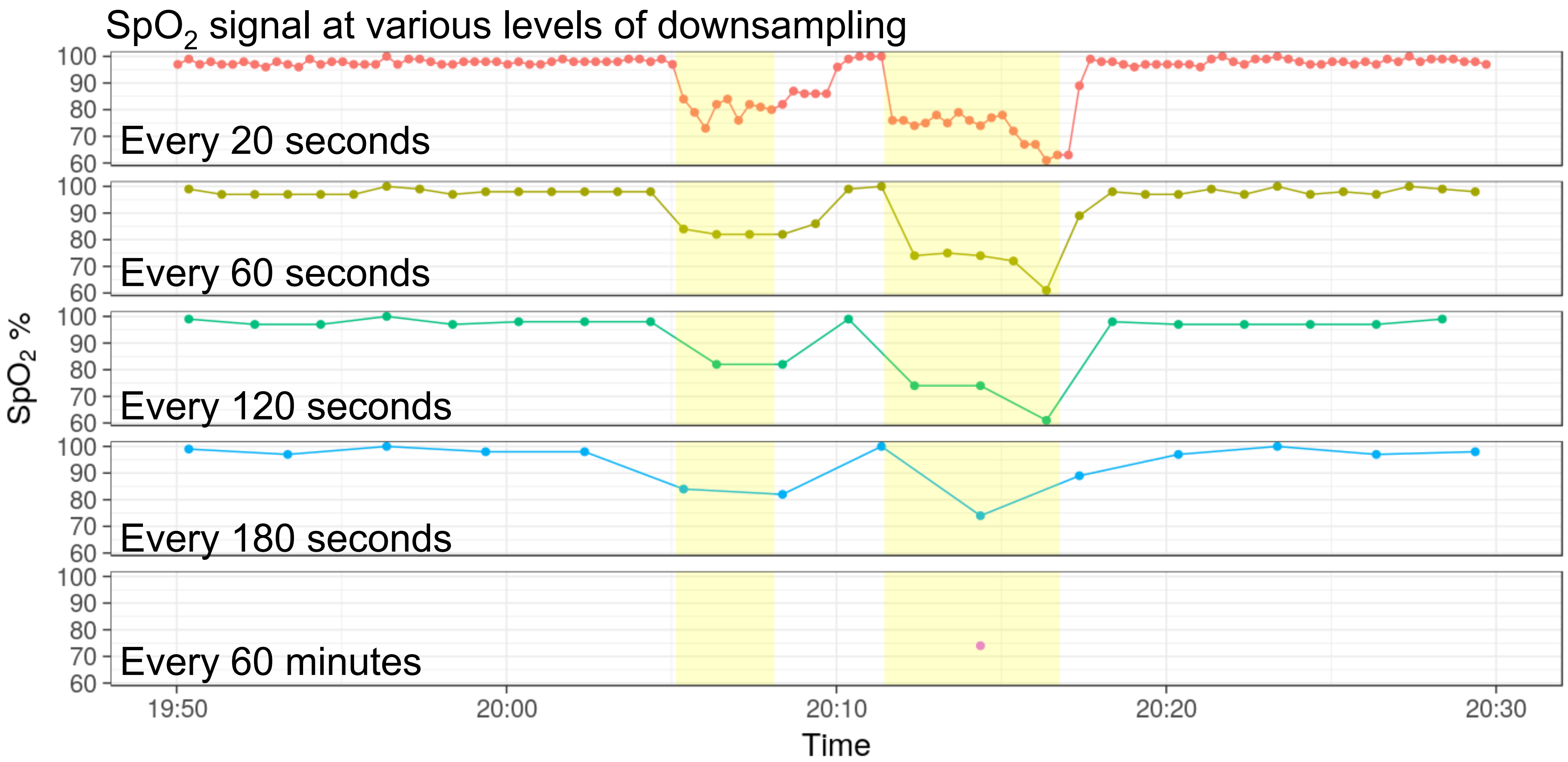
Downsampling Reduces Number of Observations



Downsampling Reduces Number of Observations



Excessive Downsampling May Leave Too Little Data



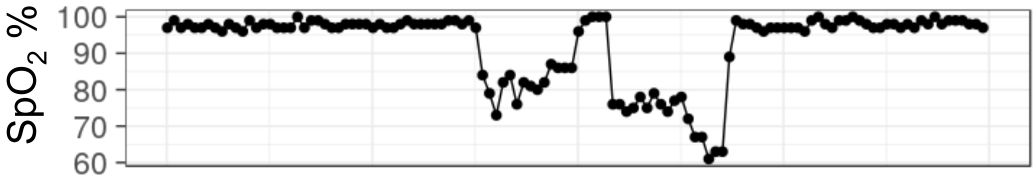
Dataset Collected from Step Down Unit Patients

- Collected continuous vital sign (VS) data streams from 200 step-down unit (SDU) patients.
  - Heart rate (HR) - 3-lead ECG.
  - Respiratory rate (RR) - bioimpedance signaling.
  - Pulse O_2 saturation (SpO_2) - pulse oximeter.
  - Intermittent noninvasive blood pressure (BP) - sphygmomanometer.
- Cardiorespiratory instability (CRI) alerts generated when VS signals are outside defined thresholds.
- Alerts labeled "real" or "artifact" by clinical investigators as previously reported.

Hravnak et al. J Clin Monit Comp 2016; 30:875-88

Alerts Generated as Vital Signs Move Outside Thresholds

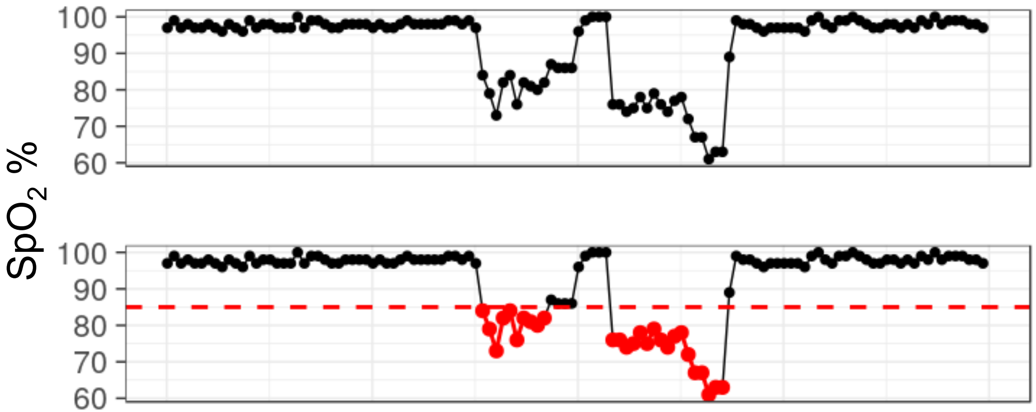
- Alerts *generated* using established process. *



* Hravnak et al. J Clin Monit Comp 2016; 30:875-88

Alerts Generated as Vital Signs Move Outside Thresholds

- Alerts *generated* using established process. *
- Mark observation "out of bounds" if it is outside specified thresholds.

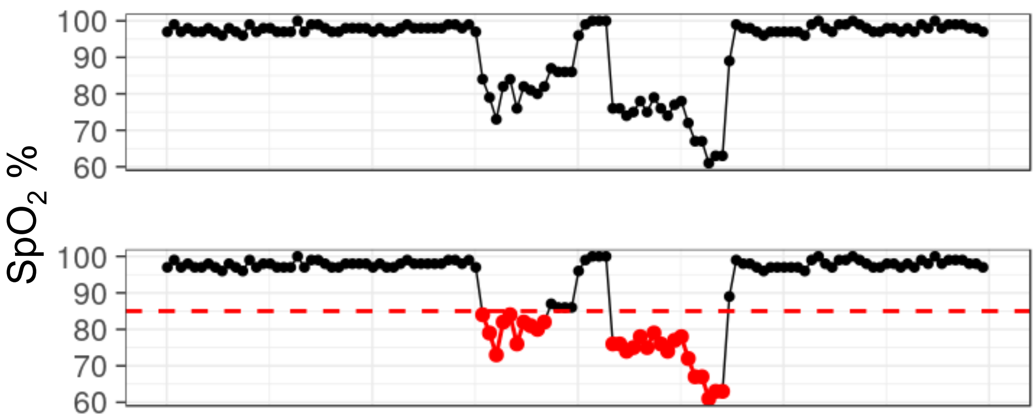


* Hravnak et al. J Clin Monit Comp 2016; 30:875-88

Alerts Generated as Vital Signs Move Outside Thresholds

- Alerts *generated* using established process. *
- Mark observation "out of bounds" if it is outside specified thresholds.
- Alert thresholds: **

HR	<	40	*OR*	HR	>	140
RR	<	9	*OR*	RR	>	36
SpO_2	<	85				
Sys	<	80	*OR*	Sys	>	200
Dia	>	110				



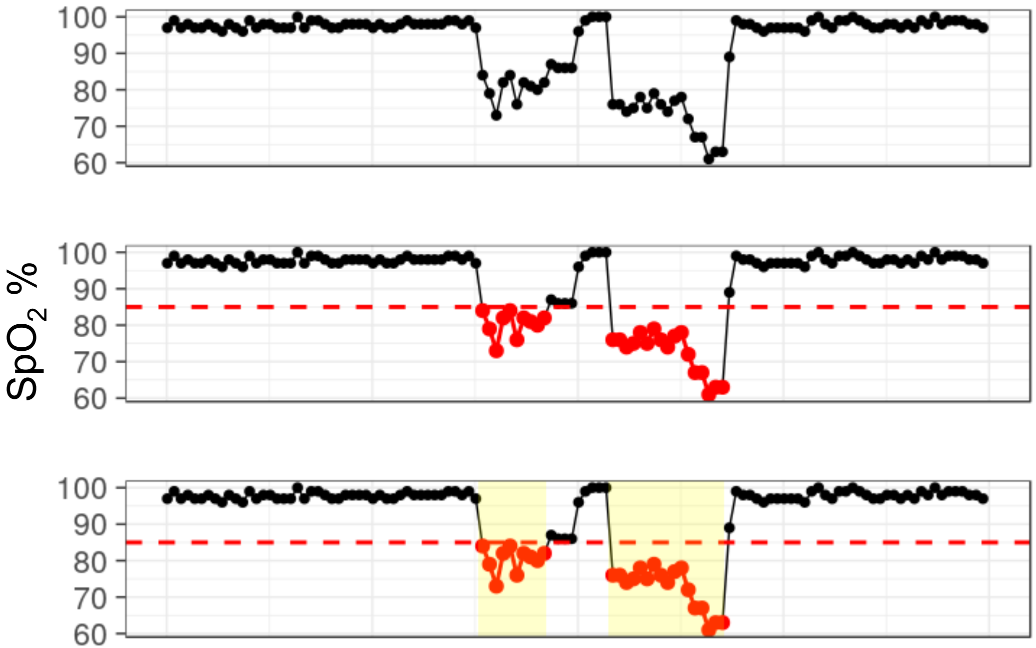
* Hravnak et al. J Clin Monit Comp 2016; 30:875-88
** Chen et al. Critical care medicine 2016; 44(7):e456-e463

Alerts Generated as Vital Signs Move Outside Thresholds

- Alerts *generated* using established process. *
- Mark observation "out of bounds" if it is outside specified thresholds.
- Alert thresholds: **

HR	<	40	*OR*	HR	>	140
RR	<	9	*OR*	RR	>	36
SpO_2	<	85				
Sys	<	80	*OR*	Sys	>	200
Dia	>	110				

- Alert if out of bounds for at least 3 minutes.

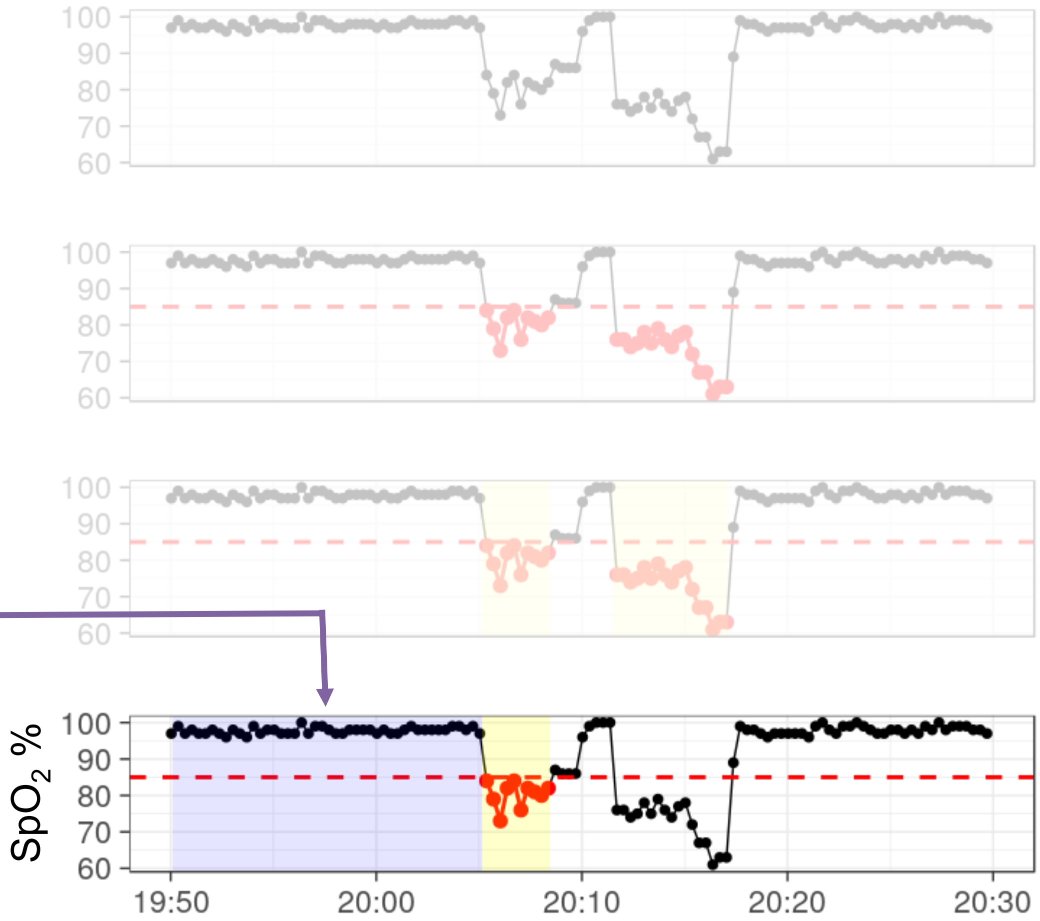


* Hravnak et al. J Clin Monit Comp 2016; 30:875-88

** Chen et al. Critical care medicine 2016; 44(7):e456-e463

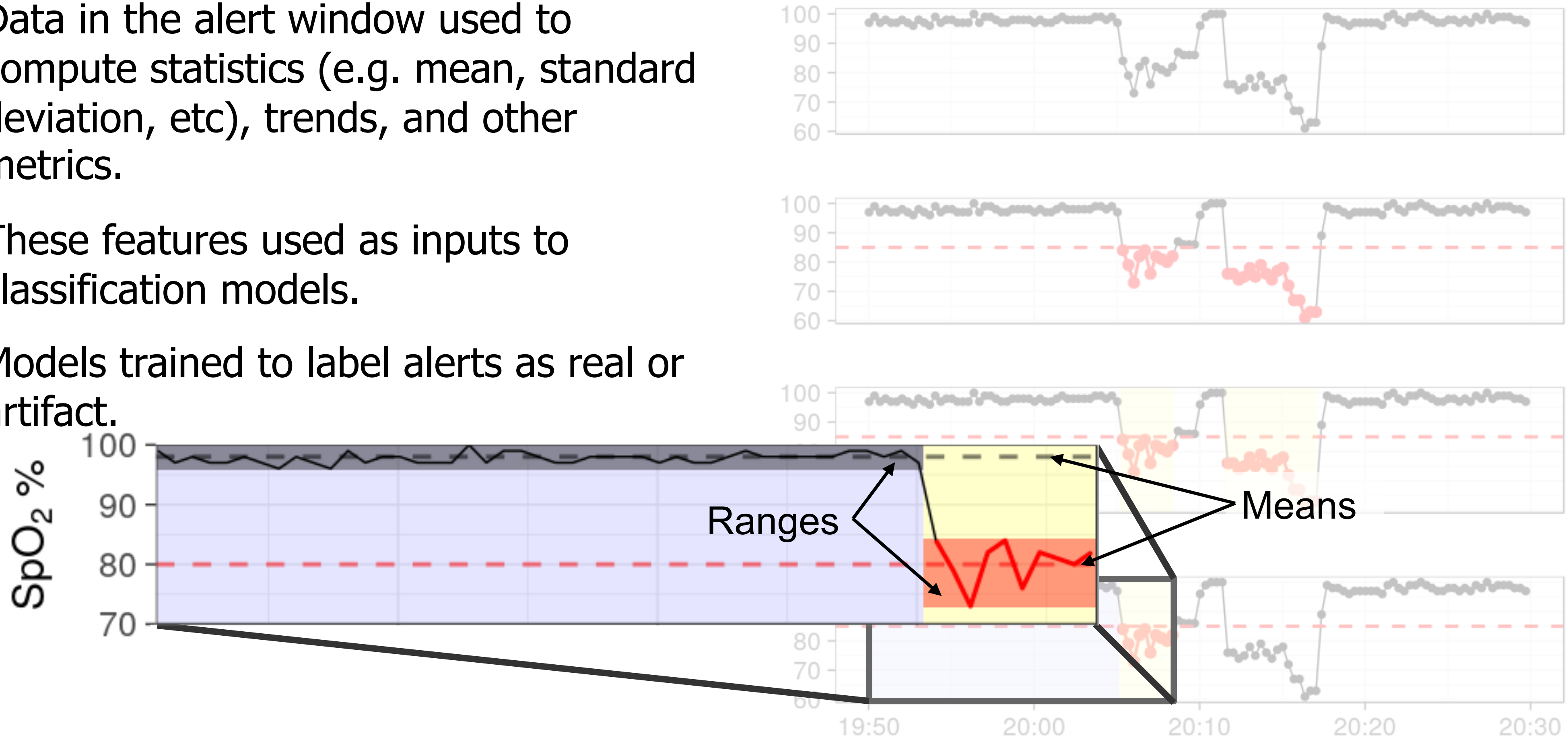
Features Generated from Downsampled Vital Sign Data

- VS data downsampled to one observation every 20 seconds (original), 60s, 120s, 180s, and 60 minutes.
- Features generated using data in alert window (featurization).
    - Feature: Secondary variable generated from original signal.
- 15 minutes prior to the alert used as a baseline.
- Raw VS along with generated features are the inputs to the models.



Featurizing Time Series by Evaluating Data Near Event

- Data in the alert window used to compute statistics (e.g. mean, standard deviation, etc), trends, and other metrics.
- These features used as inputs to classification models.
- Models trained to label alerts as real or artifact.



Models Evaluated through Cross Validation and ROCs

- Models evaluated in a leave-one-patient-out cross validation framework.

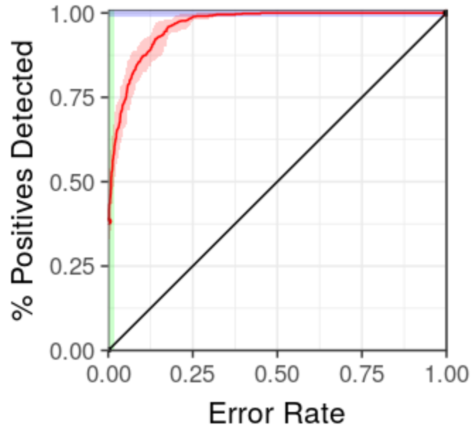
Models Evaluated through Cross Validation and ROCs

- Models evaluated in a leave-one-patient-out cross validation framework.
- Random forest models trained (Auton Lab variant).
  - Handles missing values.
  - Builds explainable models.
  - Supports non-linear decision boundaries.
  - Successfully used in many other similar projects.

Models Evaluated through Cross Validation and ROCs

- Models evaluated in a leave-one-patient-out cross validation framework.
- Random forest models trained (Auton Lab variant).
  - Handles missing values.
  - Builds explainable models.
  - Supports non-linear decision boundaries.
  - Successfully used in many other similar projects.
- Performance compared using receiver operator characteristic (ROC) curves.

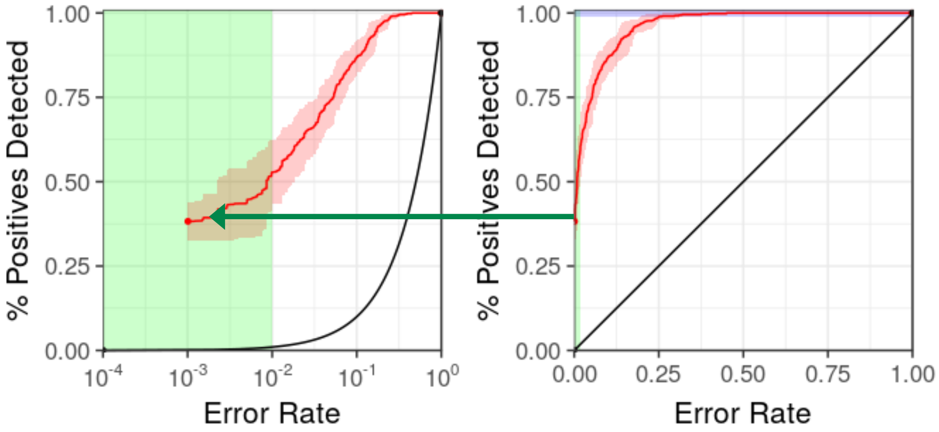
Positive class: Artifact
Negative class: Real instability



Models Evaluated through Cross Validation and ROCs

- Models evaluated in a leave-one-patient-out cross validation framework.
- Random forest models trained (Auton Lab variant).
  - Handles missing values.
  - Builds explainable models.
  - Supports non-linear decision boundaries.
  - Successfully used in many other similar projects.
- Performance compared using receiver operator characteristic (ROC) curves.

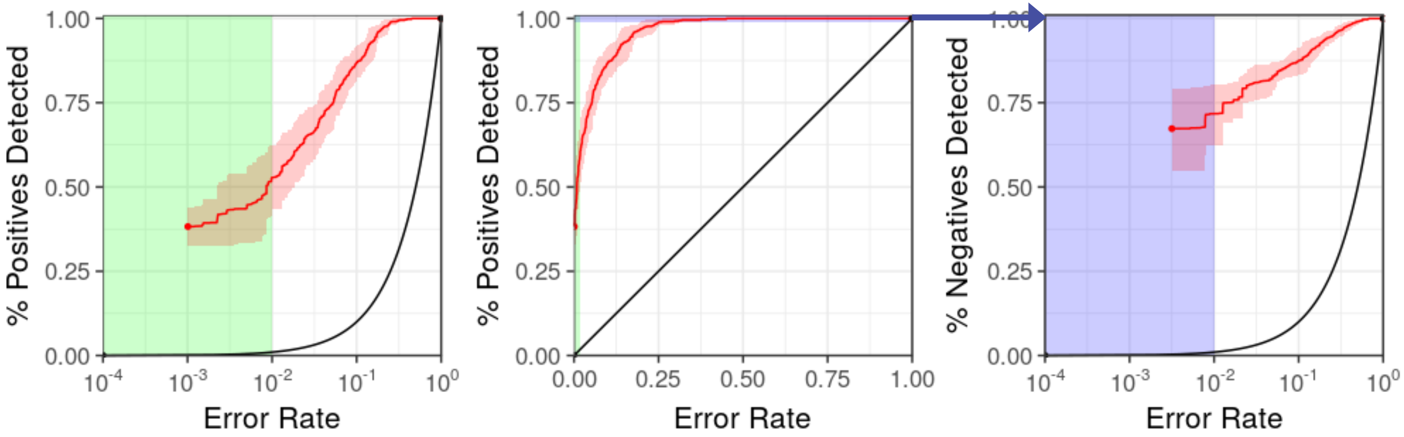
Positive class: Artifact
Negative class: Real instability



Models Evaluated through Cross Validation and ROCs

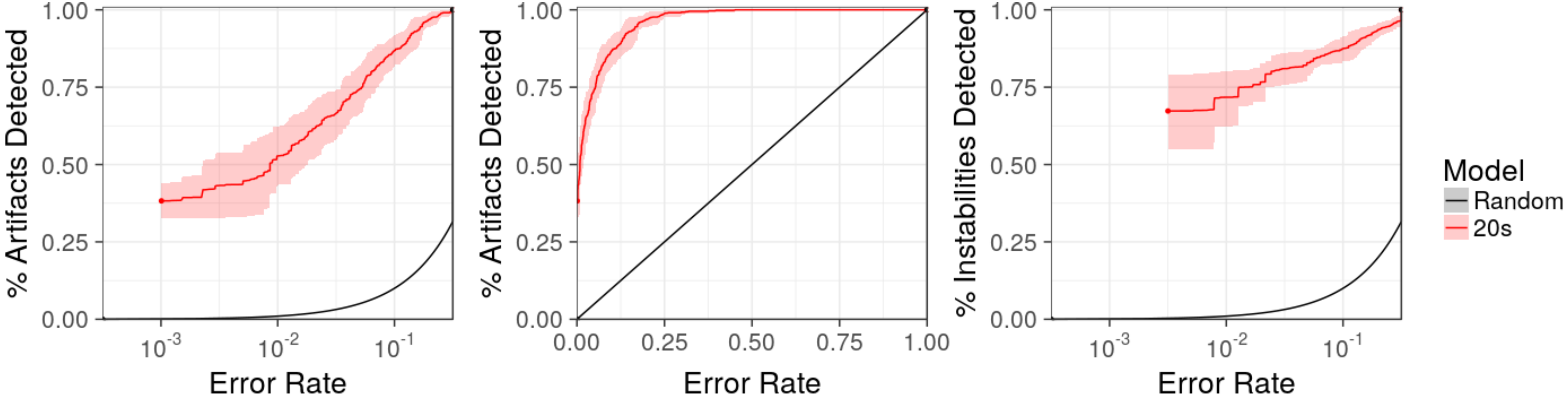
- Models evaluated in a leave-one-patient-out cross validation framework.
- Random forest models trained (Auton Lab variant).
  - Handles missing values.
  - Builds explainable models.
  - Supports non-linear decision boundaries.
  - Successfully used in many other similar projects.
- Performance compared using receiver operator characteristic (ROC) curves.

Positive class: Artifact
Negative class: Real instability



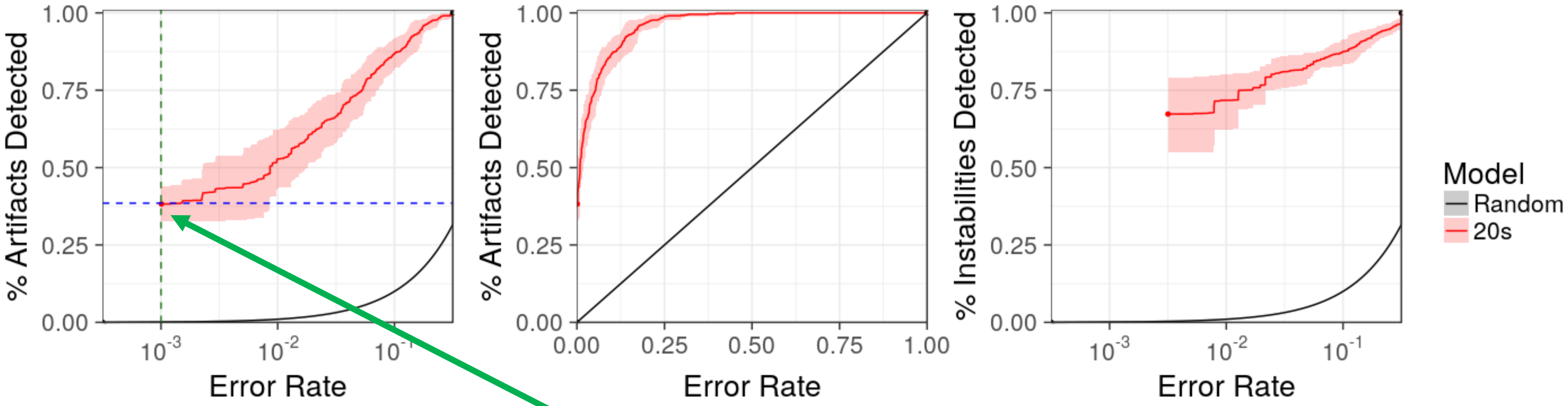
Performance of Original 20s Model

Positive class: Artifact
Negative class: Real instability



Detect 40% of Artifacts

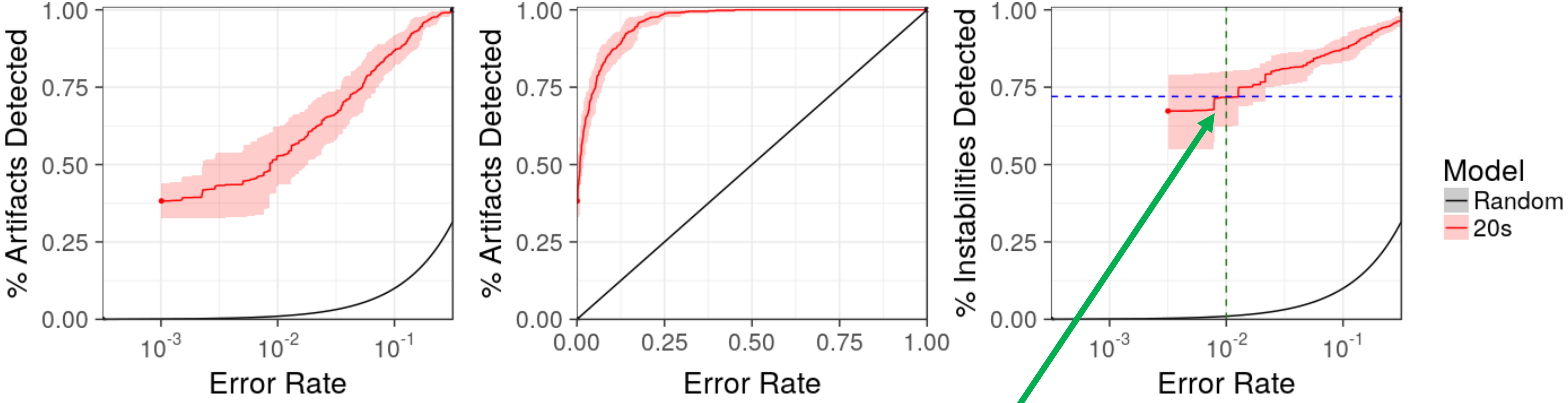
Positive class: Artifact
Negative class: Real instability



Using this threshold we correctly detect 40% of artifacts with only 1 error in 1000 decisions.

Detect 72% of Real Instabilities

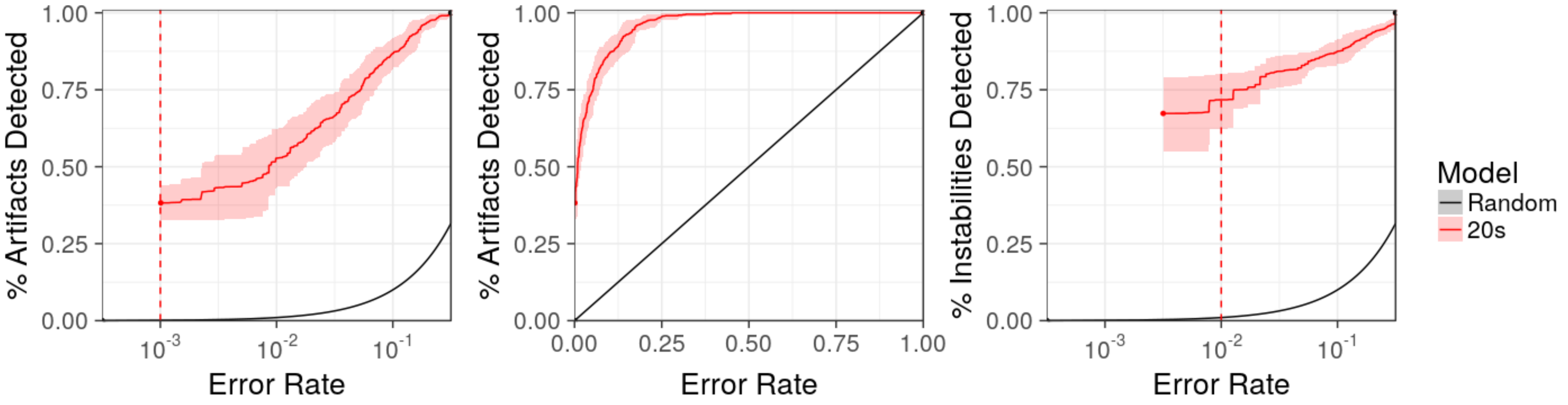
Positive class: Artifact
Negative class: Real instability



Using this threshold we correctly detect 72% of artifacts with only 1 error in 100 decisions.

For Uncertain Predictions Fall Back To Standard Practice

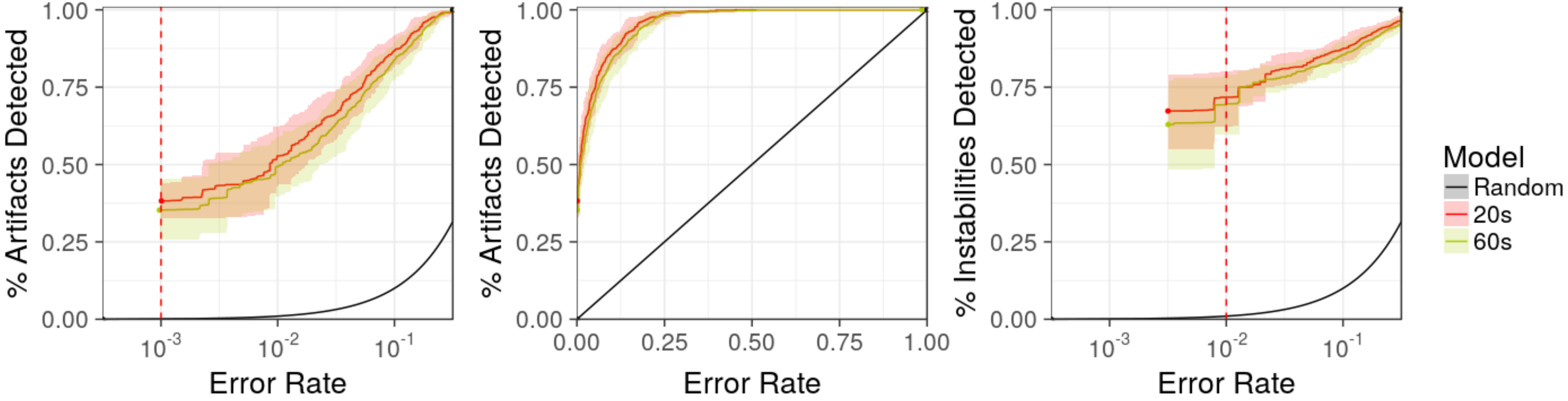
Positive class: Artifact
Negative class: Real instability



For predictions in between the model is less certain.
Fall back on current standard practice.

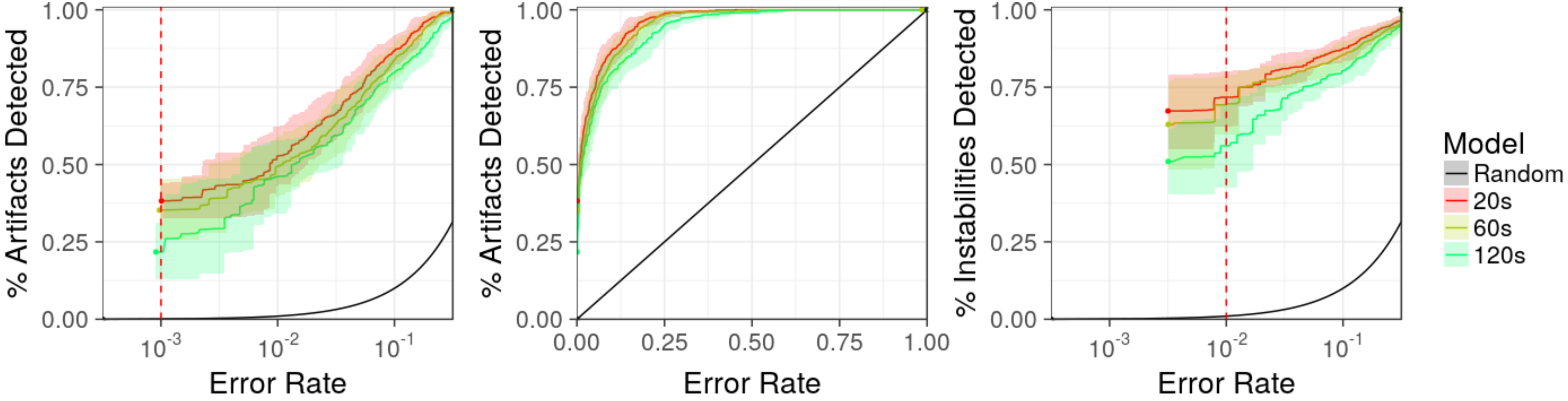
Nearly Identical Performance Sampling Every 60s

Positive class: Artifact
Negative class: Real instability



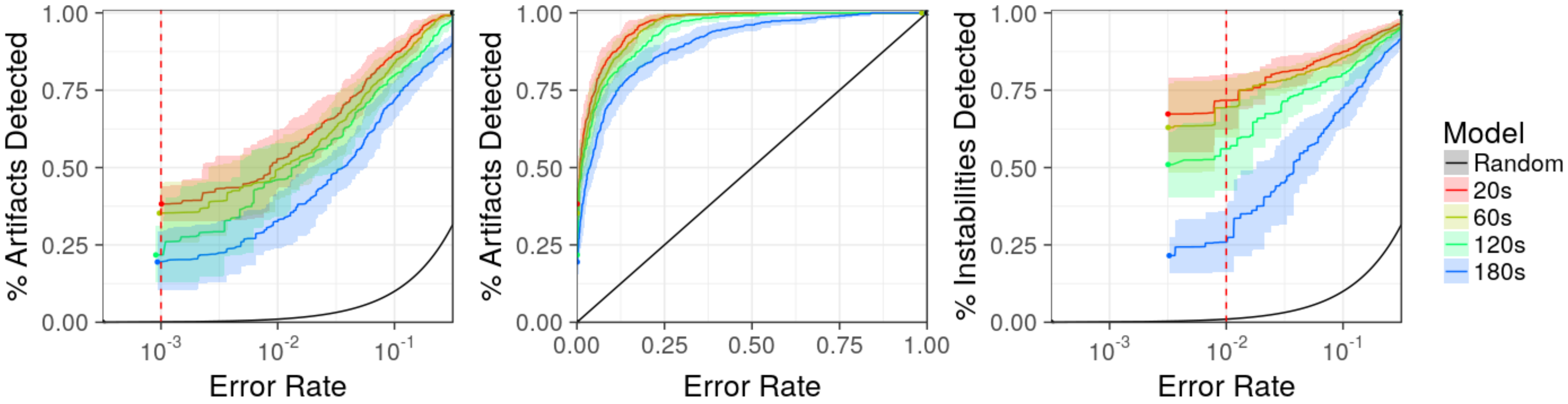
120s Model Still Differs Insignificantly

Positive class: Artifact
Negative class: Real instability



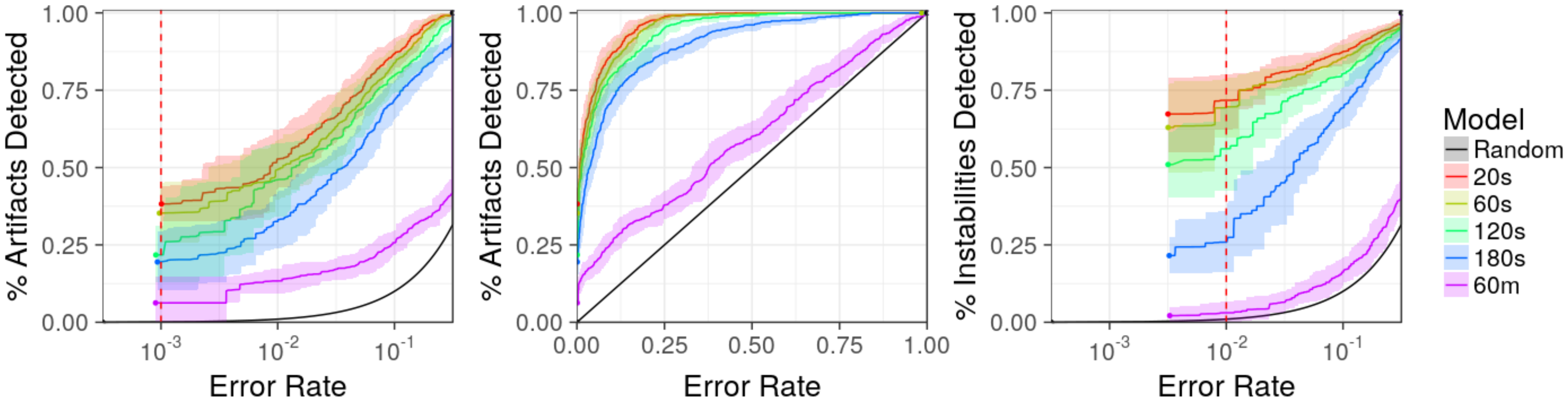
180s Model Degrades Significantly

Positive class: Artifact
Negative class: Real instability



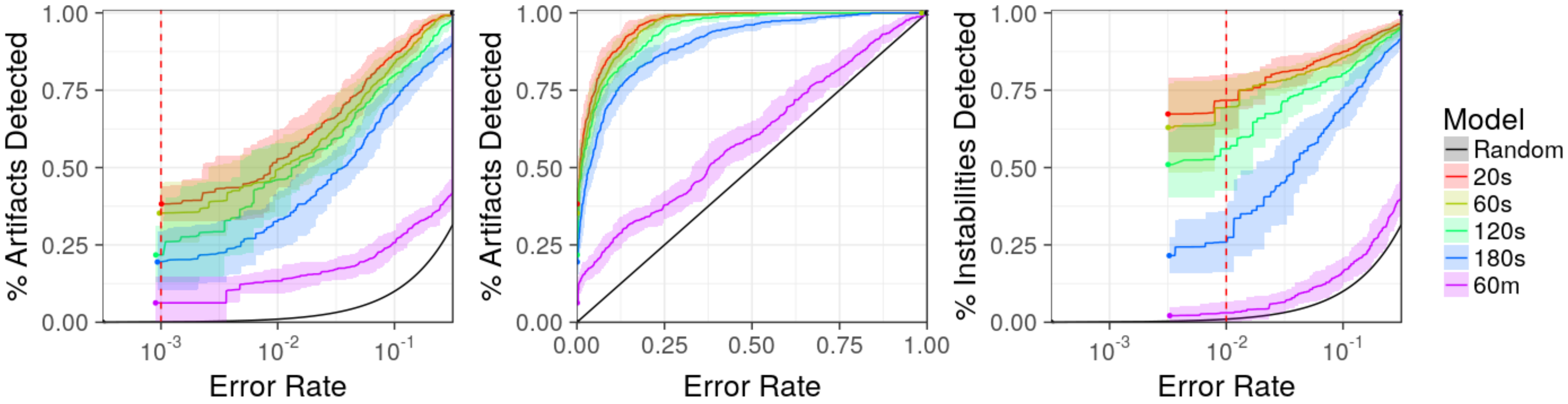
Sampling Every 60m is Detrimental to Performance

Positive class: Artifact
Negative class: Real instability



Good Model Performance if Sampling Every 1 or 2 Minutes

Positive class: Artifact
Negative class: Real instability



Key Finding: Data sampled every one or two minutes can be used to adjudicate alerts.

Lower Frequency Data Can Be Used to Classify Alerts

Key Finding

Vital sign data collected as infrequently as every 120 seconds can be used to adjudicate alerts without significantly sacrificing model performance.

Lower Frequency Data Can Be Used to Classify Alerts

Key Finding

Vital sign data collected as infrequently as every 120 seconds can be used to adjudicate alerts without significantly sacrificing model performance.

Impact (Why we care)

Analysis applicable to a wider range of existing systems which sample at lower rates.

Helps understand trade offs between sampling frequency and clinical utility of the models.

Lower Frequency Data Can Be Used to Classify Alerts

Key Finding

Vital sign data collected as infrequently as every 120 seconds can be used to adjudicate alerts without significantly sacrificing model performance.

Impact (Why we care)

Analysis applicable to a wider range of existing systems which sample at lower rates.

Helps understand trade offs between sampling frequency and clinical utility of the models.

Next Steps

Does sampling more frequently improve performance?

Work on similar projects suggests we can derive more descriptive features when higher frequency data is available.