Learning Multi-granular Models of Physiology for Detection of Bleeding

Anthony Wertz
Research Analyst
Auton Lab
Carnegie Mellon University

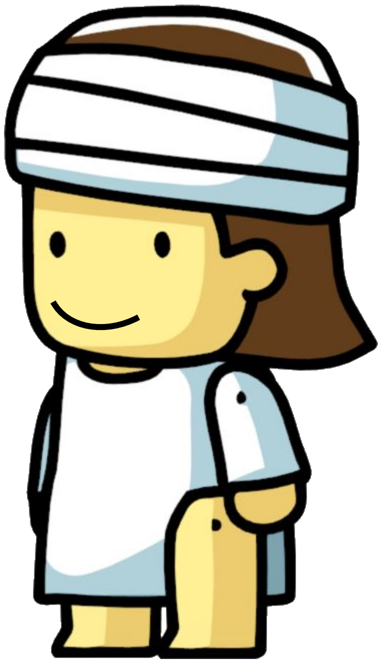
Carnegie Mellon University
Dr. Artur Dubrawski
Dr. Mathieu Guillame-Bert

University of Pittsburgh
Dr. Michael Pinsky
Dr. Gilles Clermont

October 3rd, 2017

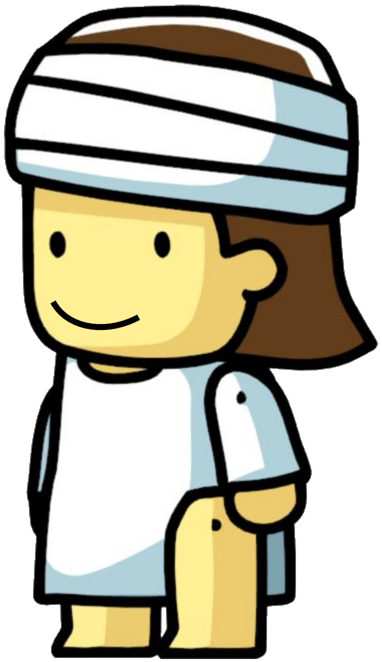
Funded by: R01GM126811

Questions



- Given observations of a patient’s vitals can we determine whether or not the patient is bleeding?

Questions

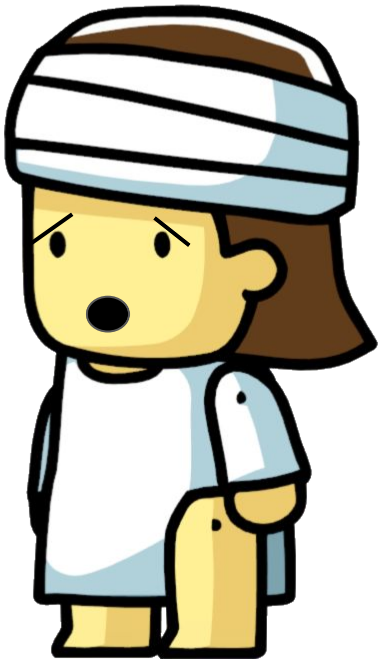


- Given observations of a patient's vitals can we determine whether or not the patient is bleeding?
- If so,
  - How quickly?
  - How often will we get false alarms?
  - How much data do we need?
  - How does the *a priori* knowledge of a patient's normal vitals affect our ability to detect bleeding?

Questions

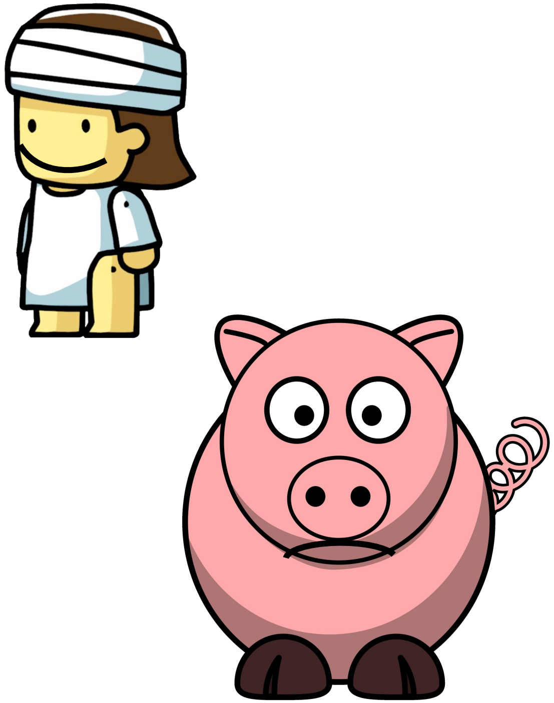
- Given observations of a patient's vitals can we determine whether or not the patient is bleeding?
- If so,
  - How quickly?
  - How often will we get false alarms?
  - How much data do we need?
  - How does the *a priori* knowledge of a patient's normal vitals affect our ability to detect bleeding?
- Can we design an experiment to collect the hemodynamic data from patients before and while controlled bleeding takes place to evaluate these questions?

Questions



- Given observations of a patient's vitals can we determine whether or not the patient is bleeding?
- If so,
  - How quickly?
  - How often will we get false alarms?
  - How much data do we need?
  - How does the *a priori* knowledge of a patient's normal vitals affect our ability to detect bleeding?
- Can we design an experiment to collect the hemodynamic data from patients before and while controlled bleeding takes place to evaluate these questions?
- It turns out, we can!

Questions



- Given observations of a patient's vitals can we determine whether or not the patient is bleeding?
- If so,
  - How quickly?
  - How often will we get false alarms?
  - How much data do we need?
  - How does the *a priori* knowledge of a patient's normal vitals affect our ability to detect bleeding?
- Can we design an experiment to collect the hemodynamic data from patients before and while controlled bleeding takes place to evaluate these questions?
- It turns out, we can! With pigs*

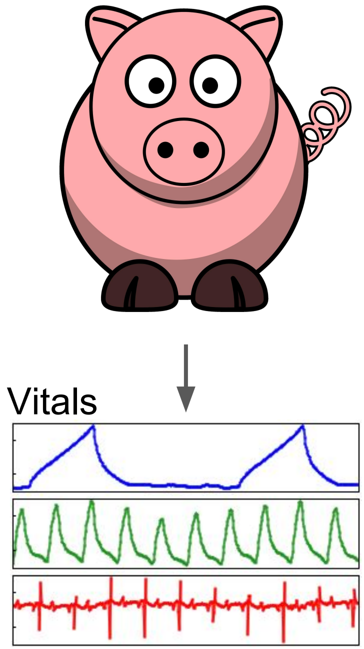
(* Ethical restrictions limit our ability to bleed humans, even in the name of science.)

Experimental Design

- Pigs are anesthetized and connected to various sensors for data collection…

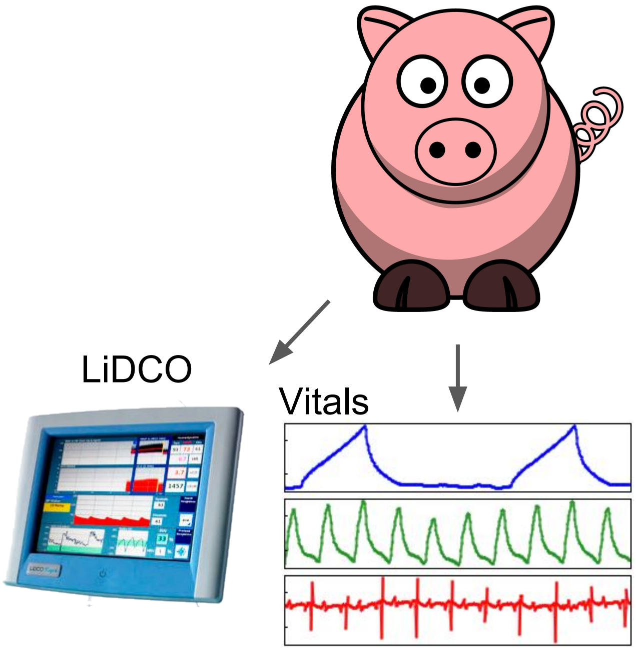
Experimental Design

- Pigs are anesthetized and connected to various sensors for data collection, including:
    - Vital sensor data (arterial, central venous, and pulmonary artery pressure, ECG, plethysmograph, SpO_2) at 250Hz and SvO_2 once every two seconds.



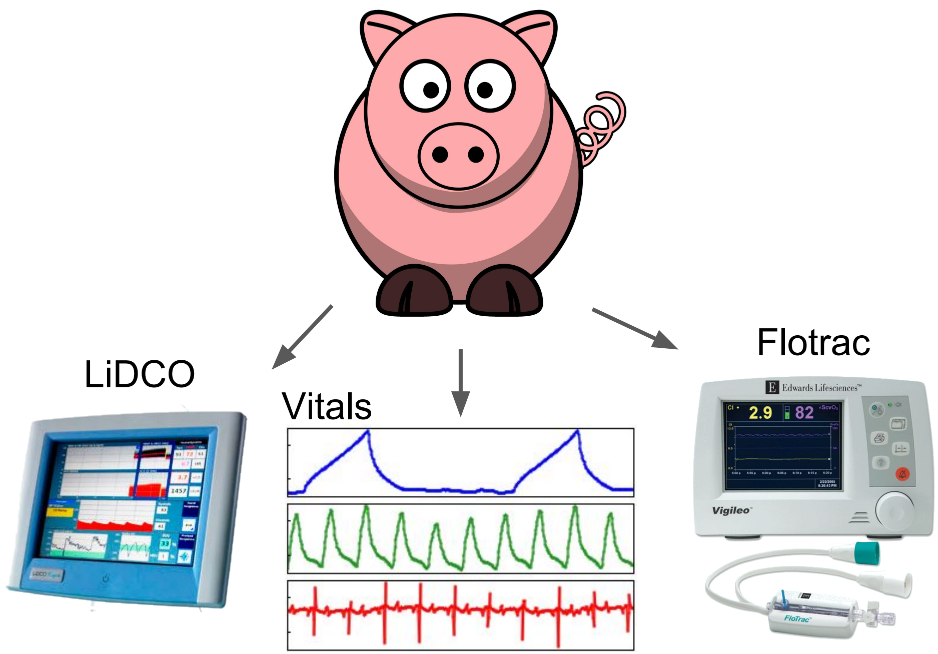
Experimental Design

- Pigs are anesthetized and connected to various sensors for data collection, including:
  - Vital sensor data (arterial, central venous, and pulmonary artery pressure, ECG, plethysmograph, SpO_2) at 250Hz and SvO_2 once every two seconds.
  - Beat-to-beat LiDCO data.



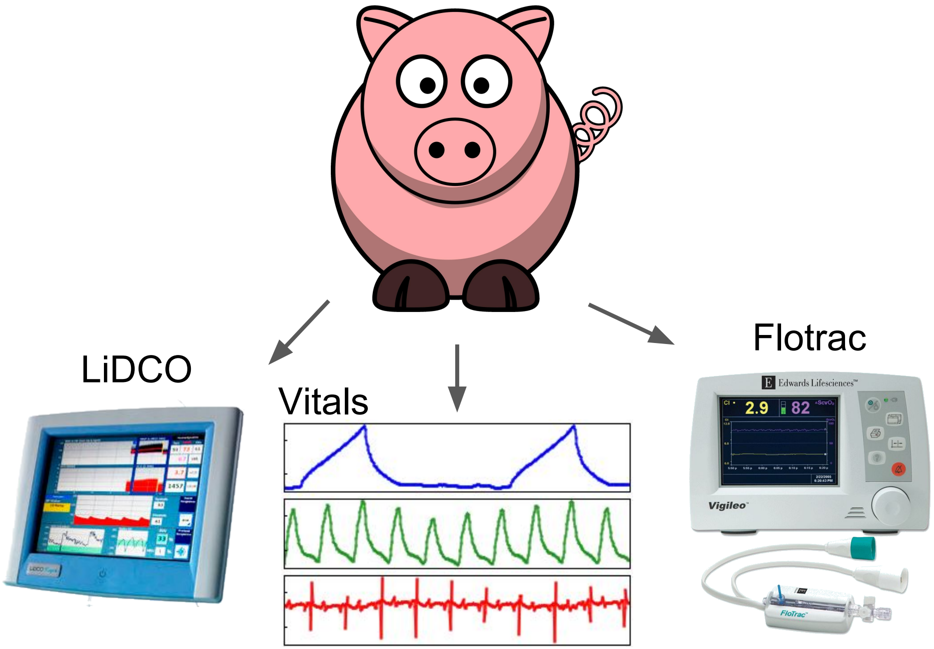
Experimental Design

- Pigs are anesthetized and connected to various sensors for data collection, including:
  - Vital sensor data (arterial, central venous, and pulmonary artery pressure, ECG, plethysmograph, SpO_2) at 250Hz and SvO_2 once every two seconds.
  - Beat-to-beat LiDCO data.
  - Flotrac data every 20 seconds.



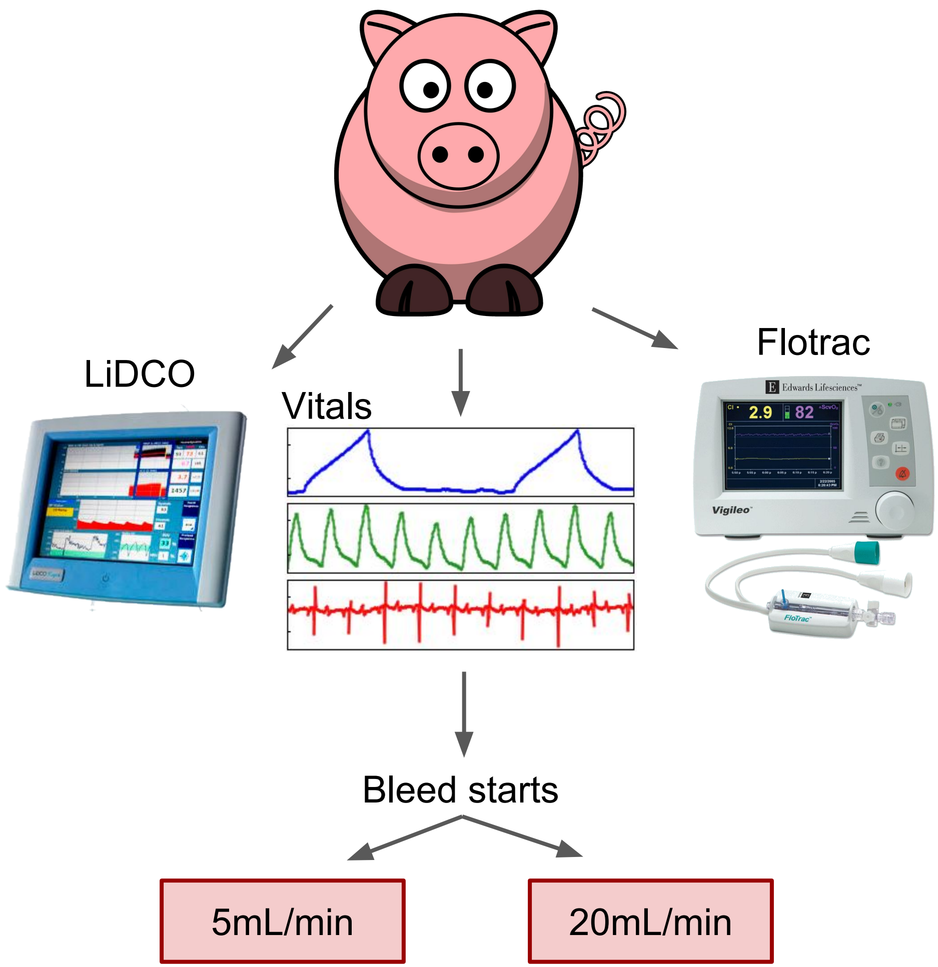
Experimental Design

- Pigs are anesthetized and connected to various sensors for data collection, including:
  - Vital sensor data (arterial, central venous, and pulmonary artery pressure, ECG, plethysmograph, SpO_2) at 250Hz and SvO_2 once every two seconds.
  - Beat-to-beat LiDCO data.
  - Flotrac data every 20 seconds.
- They are left to rest for 30 minutes while baseline data is collected.



Experimental Design

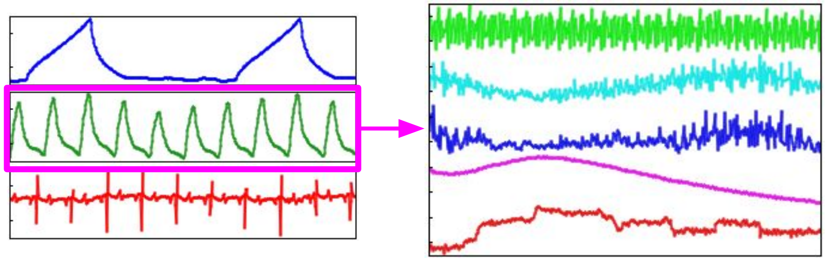
- Pigs are anesthetized and connected to various sensors for data collection, including:
  - Vital sensor data (arterial, central venous, and pulmonary artery pressure, ECG, plethysmograph, SpO_2) at 250Hz and SvO_2 once every two seconds.
  - Beat-to-beat LiDCO data.
  - Flotrac data every 20 seconds.
- They are left to rest for 30 minutes while baseline data is collected.
- The pigs are then bled at a constant rate, either:
  - 5mL/min until mean arterial pressure drops below 40mmHg, or
  - 20mL/min until mean arterial pressure drops below 30mmHg.



Computational Experimental Design

- The data is featurized and those features are split into different (though not mutually exclusive) groups.

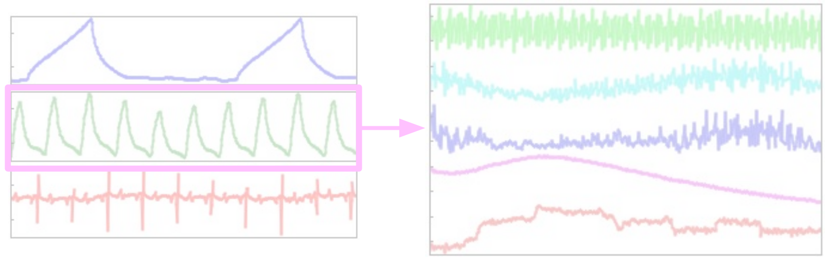
  (Featurization and the groups will be discussed soon.)



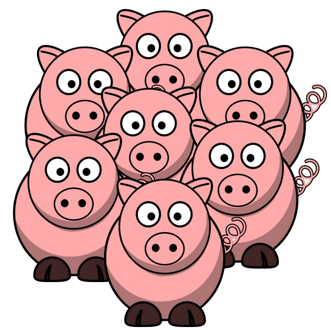
Computational Experimental Design

- The data is featurized and those features are split into different (though not mutually exclusive) groups.
- Those feature sets are used to validate random forest models that classify a pig as bleeding or not in a leave - one - pig - out cross validation framework.

(These will also be discussed briefly.)



Training Set

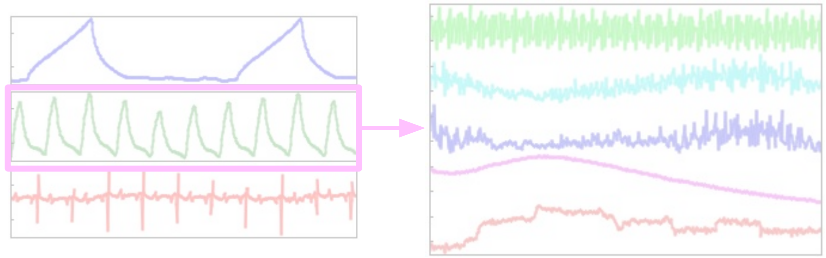


Testing Set

Computational Experimental Design

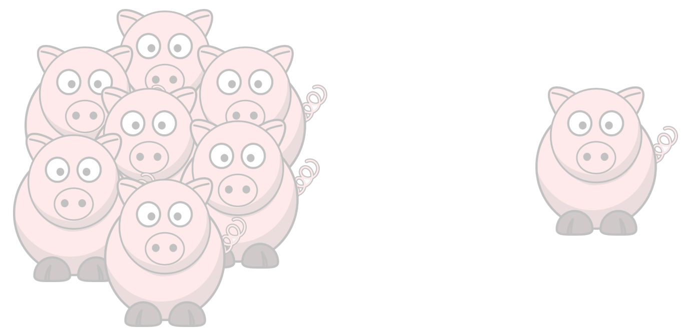
- The data is featurized and those features are split into different (though not mutually exclusive) groups.
- Those feature sets are used to validate random forest models that classify a pig as bleeding or not in a leave-one-pig-out cross validation framework.
- The detection results are evaluated by means of Receiver Operator Characteristic (ROC) and Activity Monitoring Operator Characteristic (AMOC) curves.

(These will be described in a bit more detail since they're necessary to understand the results.)

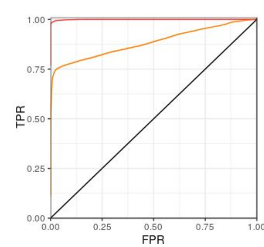


Training Set

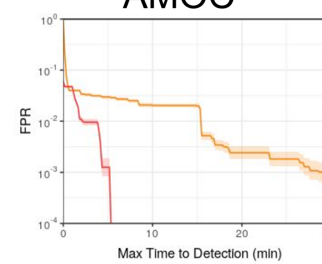
Testing Set



ROC

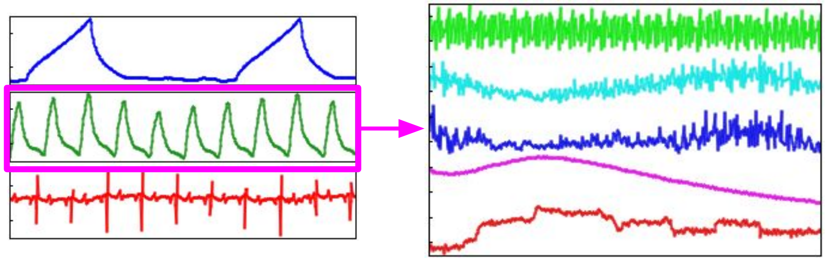


AMOC



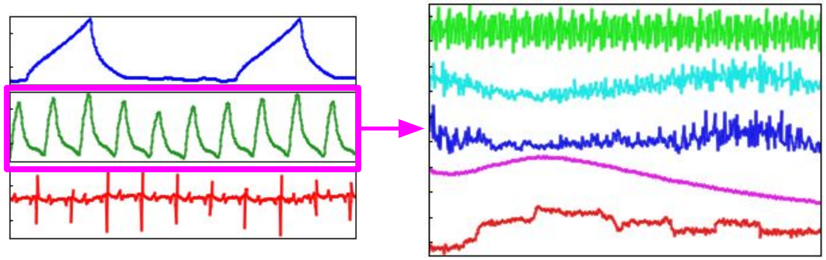
Featurizations and Groups

- A featurization is, simply put, a transformation of the original input data.
    - E.g. given a time series of blood pressures the mean is computed every five minutes.

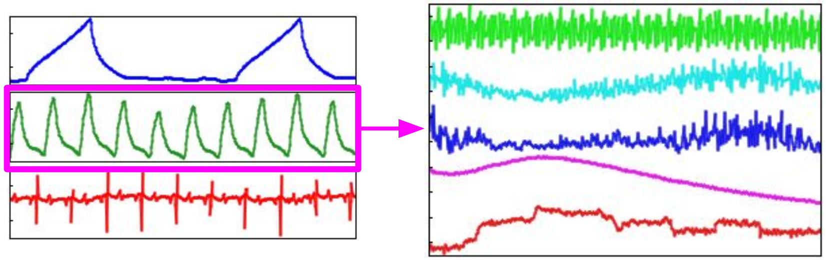


Featurizations and Groups

- A featurization is, simply put, a transformation of the original input data.
  - E.g. given a time series of blood pressures the mean is computed every five minutes.
- We are especially interested in time series featurizations since these are directly impacted by data granularity.



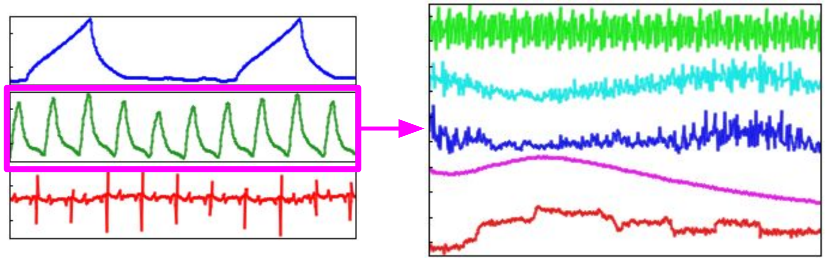
Featurizations and Groups



- A featurization is, simply put, a transformation of the original input data.
  - E.g. given a time series of blood pressures the mean is computed every five minutes.
- We are especially interested in time series featurizations since these are directly impacted by data granularity.
- To assess the impact of data granularity we grouped the featurizations in four blocks:

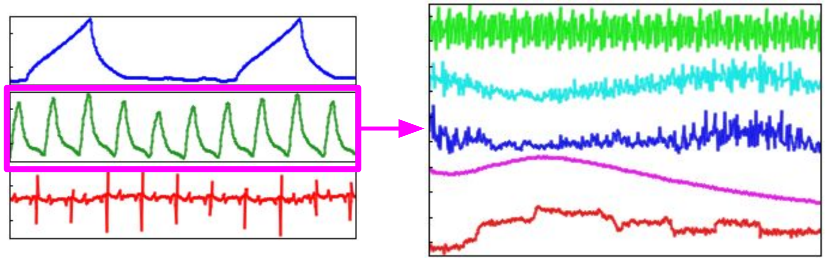
Featurizations and Groups

- A featurization is, simply put, a transformation of the original input data.
  - E.g. given a time series of blood pressures the mean is computed every five minutes.
- We are especially interested in time series featurizations since these are directly impacted by data granularity.
- To assess the impact of data granularity we grouped the featurizations in four blocks:
  - Low Frequency (LF)



Low Frequency (LF) - 7 Features
Instantaneous vital data only. Assuming a low frequency of data collection (around one observation every two minutes) useful time featurizations are not possible without incurring very long detection latencies.

Featurizations and Groups



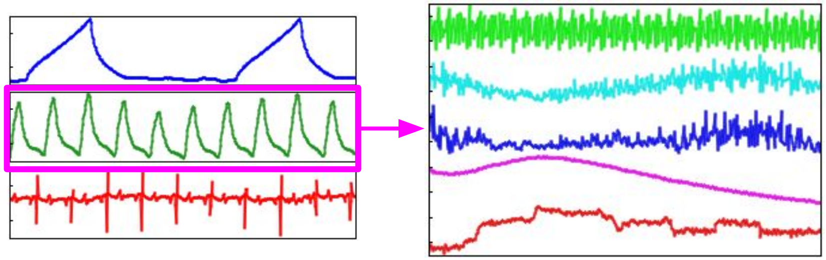
- A featurization is, simply put, a transformation of the original input data.
  - E.g. given a time series of blood pressures the mean is computed every five minutes.
- We are especially interested in time series featurizations since these are directly impacted by data granularity.
- To assess the impact of data granularity we grouped the featurizations in four blocks:
  - Low Frequency (LF)
  - Beat-to-Beat (B2B)

Beat-to-Beat (B2B) - 158 Features
Data from LiDCO device every heart beat, and data from Flotrac device every 20 seconds. Flotrac featurizations include additional features provided by Flotrac group that are not ordinarily available.

Low Frequency (LF) - 7 Features
Instantaneous vital data only. Assuming a low frequency of data collection (around one observation every two minutes) useful time featurizations are not possible without incurring very long detection latencies.

Featurizations and Groups

- A featurization is, simply put, a transformation of the original input data.
  - E.g. given a time series of blood pressures the mean is computed every five minutes.
- We are especially interested in time series featurizations since these are directly impacted by data granularity.
- To assess the impact of data granularity we grouped the featurizations in four blocks:
  - Low Frequency (LF)
  - Beat-to-Beat (B2B)
  - Beat-to-Beat + Low Frequency (B2B+LF)



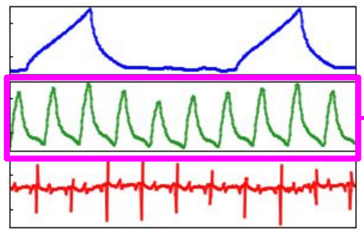
Beat-to-Beat + Low Frequency (B2B+LF)

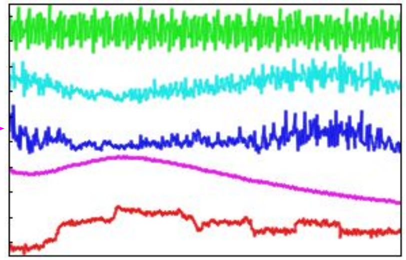
Beat-to-Beat (B2B) - 158 Features
Data from LiDCO device every heart beat, and data from Flotrac device every 20 seconds. Flotrac featurizations include additional features provided by Flotrac group that are not ordinarily available.

Low Frequency (LF) - 7 Features
Instantaneous vital data only. Assuming a low frequency of data collection (around one observation every two minutes) useful time featurizations are not possible without incurring very long detection latencies.

Featurizations and Groups

- A featurization is, simply put, a transformation of the original input data.
  - E.g. given a time series of blood pressures the mean is computed every five minutes.
- We are especially interested in time series featurizations since these are directly impacted by data granularity.
- To assess the impact of data granularity we grouped the featurizations in four blocks:
  - Low Frequency (LF)
  - Beat-to-Beat (B2B)
  - Beat-to-Beat + Low Frequency (B2B+LF)
  - High Frequency





High Frequency (HF) - 323 Features
Includes various featurizations of vital waveforms along with featurizations of B2B data.

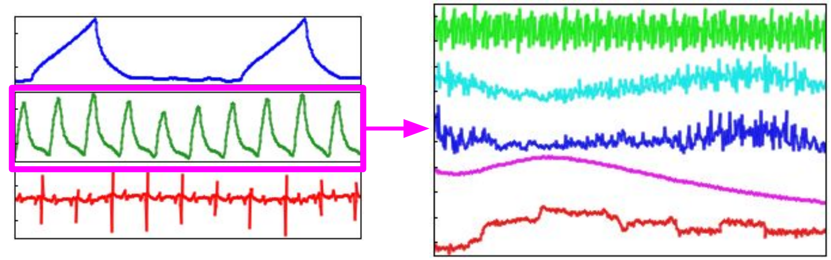
Beat-to-Beat + Low Frequency (B2B+LF)

Beat-to-Beat (B2B) - 158 Features
Data from LiDCO device every heart beat, and data from Flotrac device every 20 seconds. Flotrac featurizations include additional features provided by Flotrac group that are not ordinarily available.

Low Frequency (LF) - 7 Features
Instantaneous vital data only. Assuming a low frequency of data collection (around one observation every two minutes) useful time featurizations are not possible without incurring very long detection latencies.

Featurizations and Groups

- A featurization is, simply put, a transformation of the original input data.
  - E.g. given a time series of blood pressures the mean is computed every five minutes.
- We are especially interested in time series featurizations since these are directly impacted by data granularity.
- To assess the impact of data granularity we grouped the featurizations in four blocks:
  - Low Frequency (LF)
  - Beat-to-Beat (B2B)
  - Beat-to-Beat + Low Frequency (B2B+LF)
  - High Frequency
- We also compare models with and without baseline normalization.



High Frequency (HF) - ~~323~~ 639 Features
Includes various featurizations of vital waveforms along with featurizations of B2B data.

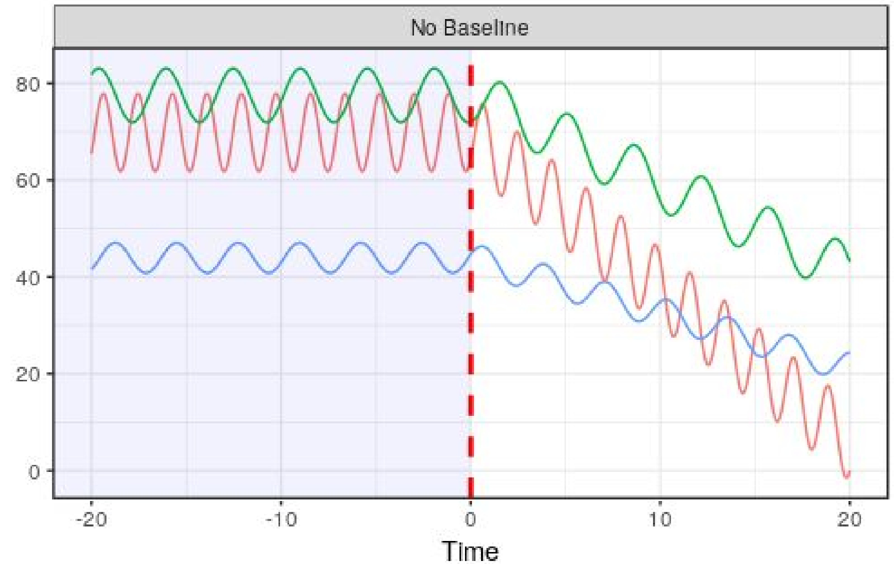
Beat-to-Beat + Low Frequency (B2B+LF)

Beat-to-Beat (B2B) - ~~158~~ 312 Features
Data from LiDCO device every heart beat, and data from Flotrac device every 20 seconds. Flotrac featurizations include additional features provided by Flotrac group that are not ordinarily available.

Low Frequency (LF) - ~~7~~ 14 Features
Instantaneous vital data only. Assuming a low frequency of data collection (around one observation every two minutes) useful time featurizations are not possible without incurring very long detection latencies.

Individual Baseline Normalization

- Patients can be very different when stable.

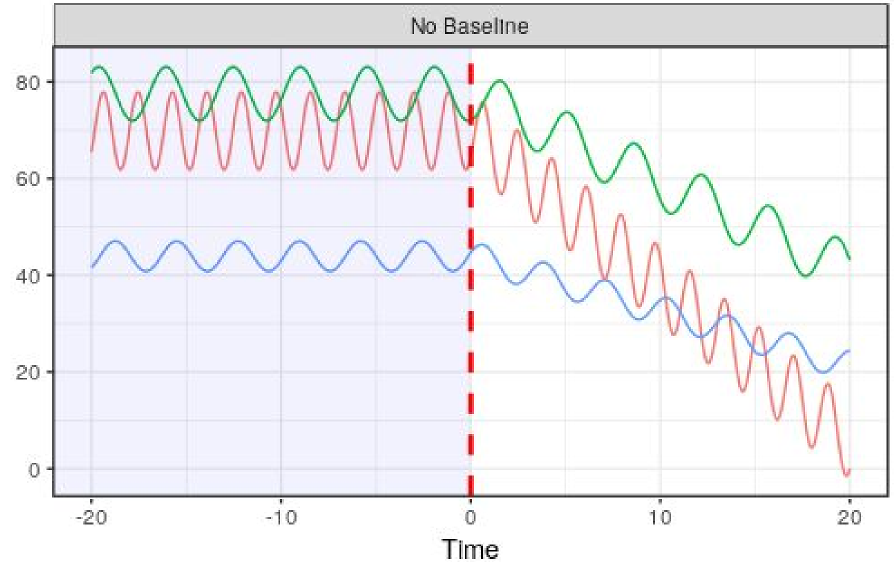


Stable before event.

After t=0 something interesting starts to happen to the signals.

Individual Baseline Normalization

- Patients can be very different when stable.
- What threshold yields fast detection of event at t=0 and few false alarms for all patients?

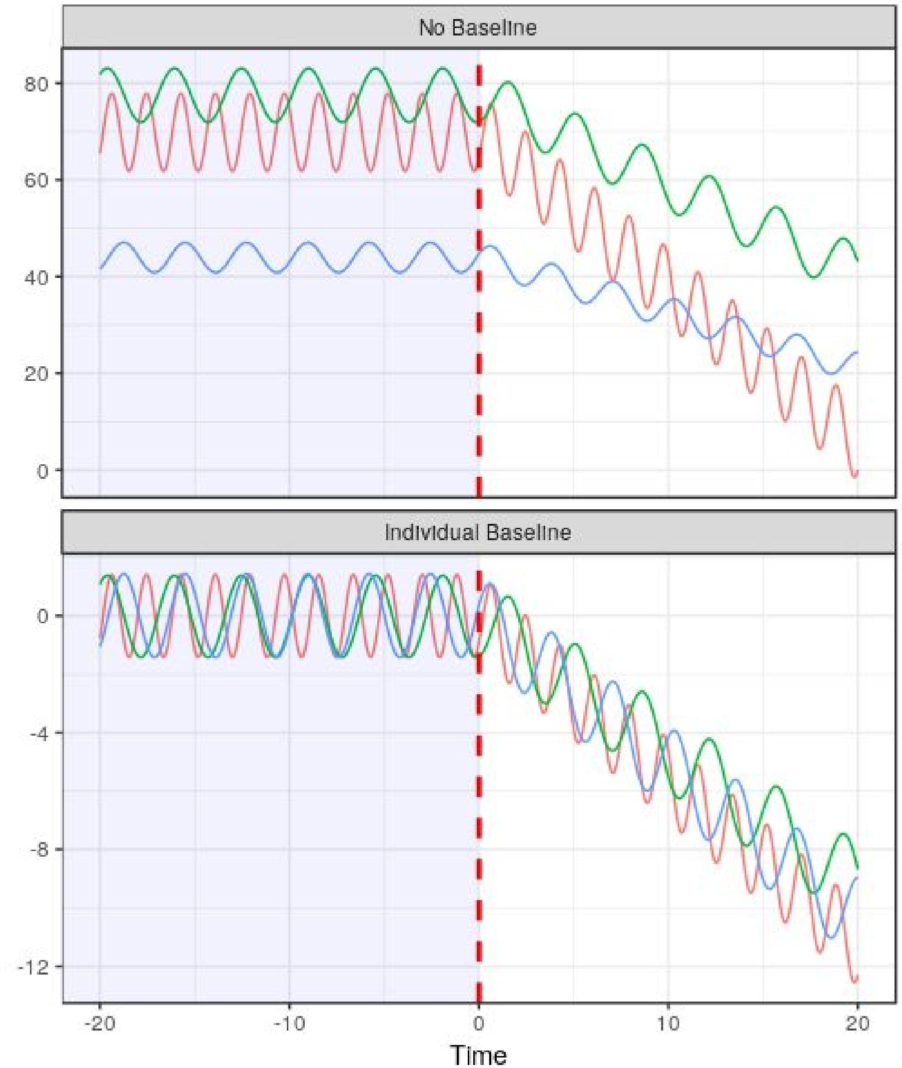


Stable before event.

After t=0 something interesting starts to happen to the signals.

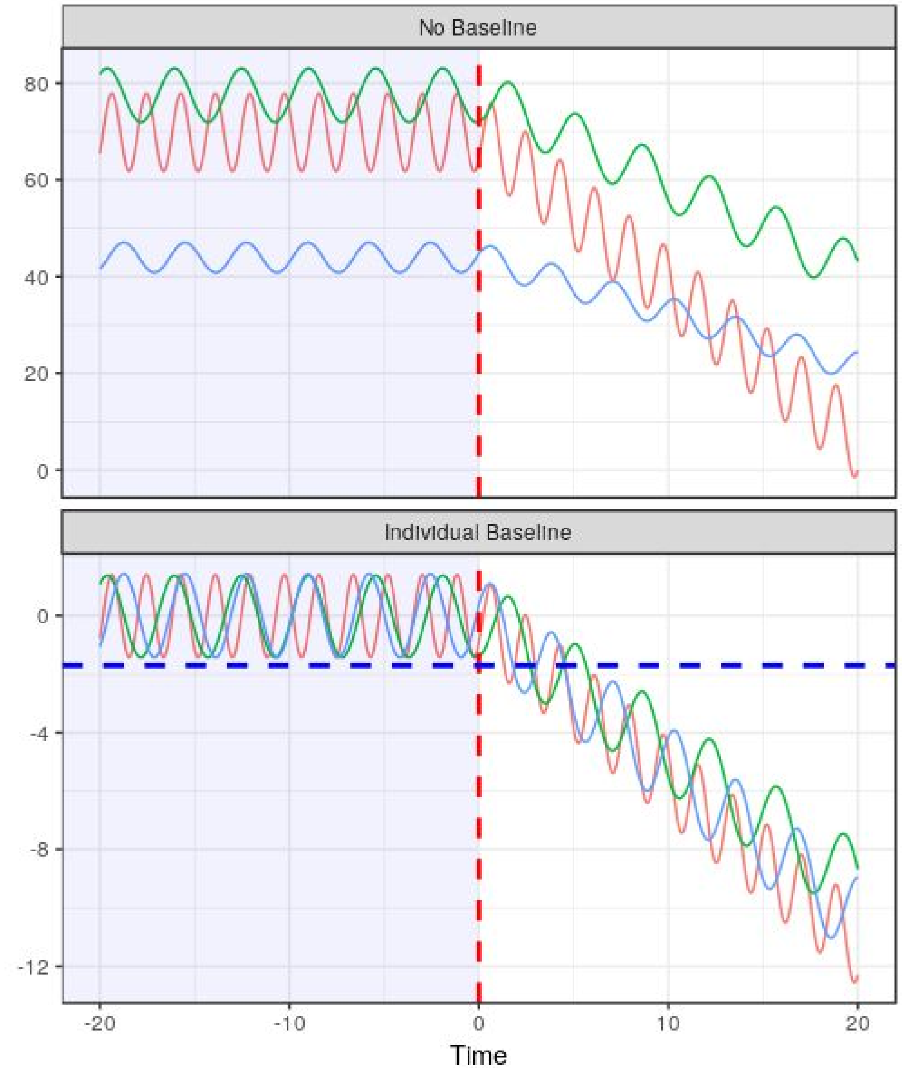
Individual Baseline Normalization

- Patients can be very different when stable.
- What threshold yields fast detection of event at t=0 and few false alarms for all patients?
- Assume some regularity in the baseline period:
  - Center on the mean.
  - Scale by its standard deviation.



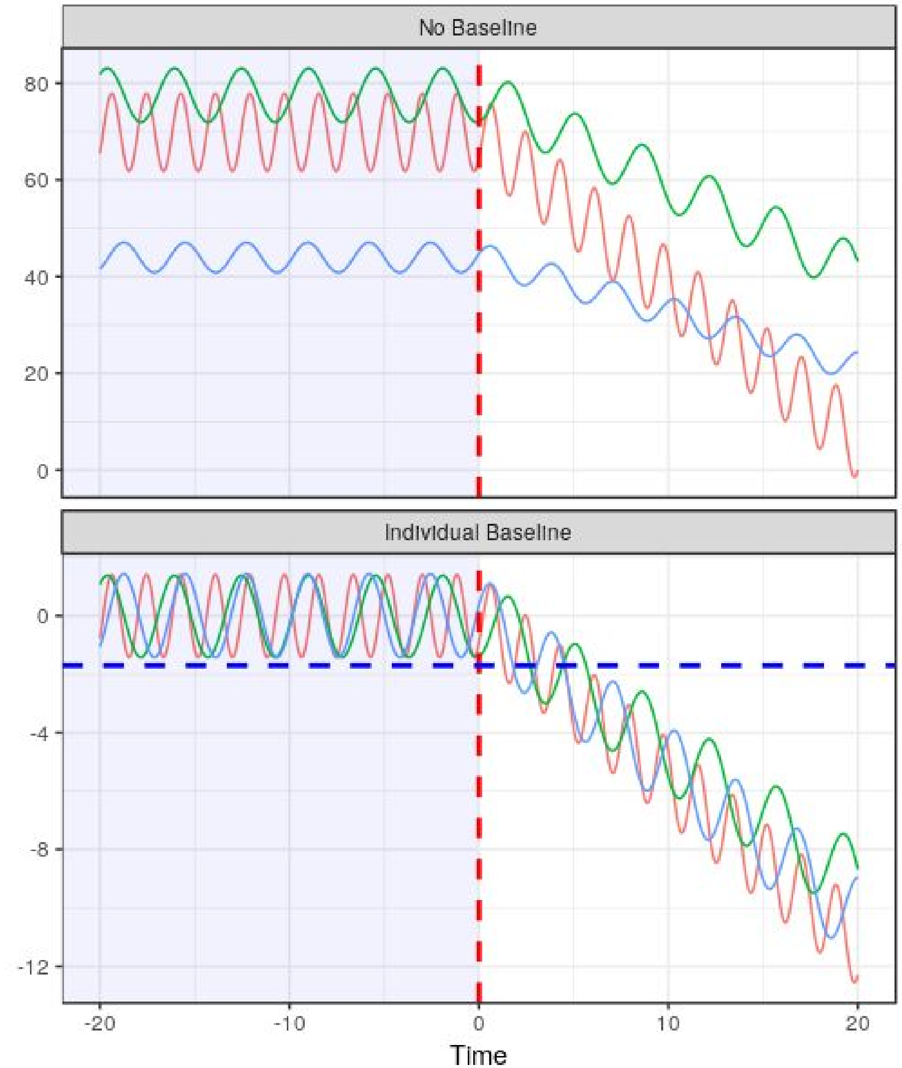
Individual Baseline Normalization

- Patients can be very different when stable.
- What threshold yields fast detection of event at t=0 and few false alarms for all patients?
- Assume some regularity in the baseline period:
  - Center on the mean.
  - Scale by its standard deviation.
  - Now we can find a threshold for this data the yields fast detections and few false positives.



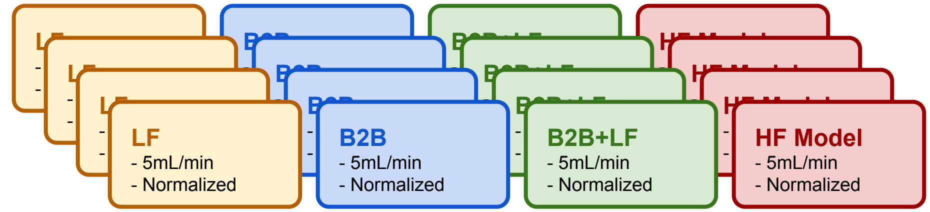
Individual Baseline Normalization

- Patients can be very different when stable.
- What threshold yields fast detection of event at t=0 and few false alarms for all patients?
- Assume some regularity in the baseline period:
  - Center on the mean.
  - Scale by its standard deviation.
  - Now we can find a threshold for this data the yields fast detections and few false positives.
- For this to work we need to collect data when we know the patient is stable.
  - Not available for every patient.
  - But can be captured for patients prior to, for example, surgery.



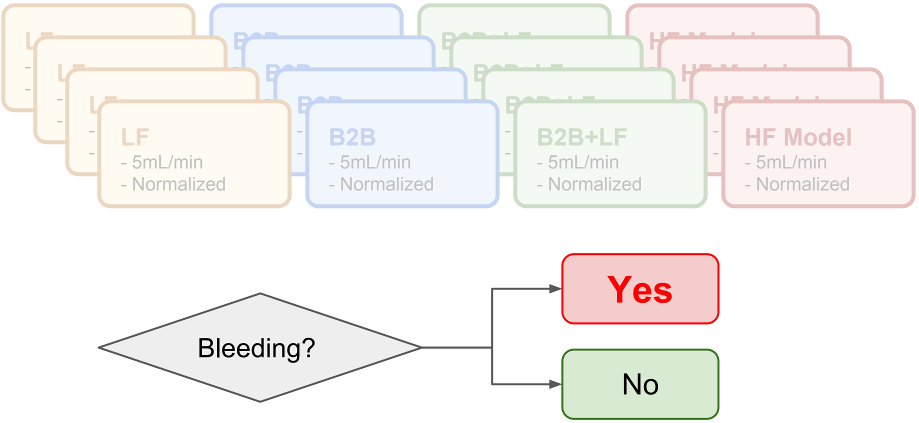
Evaluation Methodology

- Separate models are trained for combinations of
    - 5mL/min (n=14) or 20mL/min (n=46)
    - LF, B2B, B2B+LF, HF
    - No normalization or Individual baseline normalized
    - 16 different models in total.



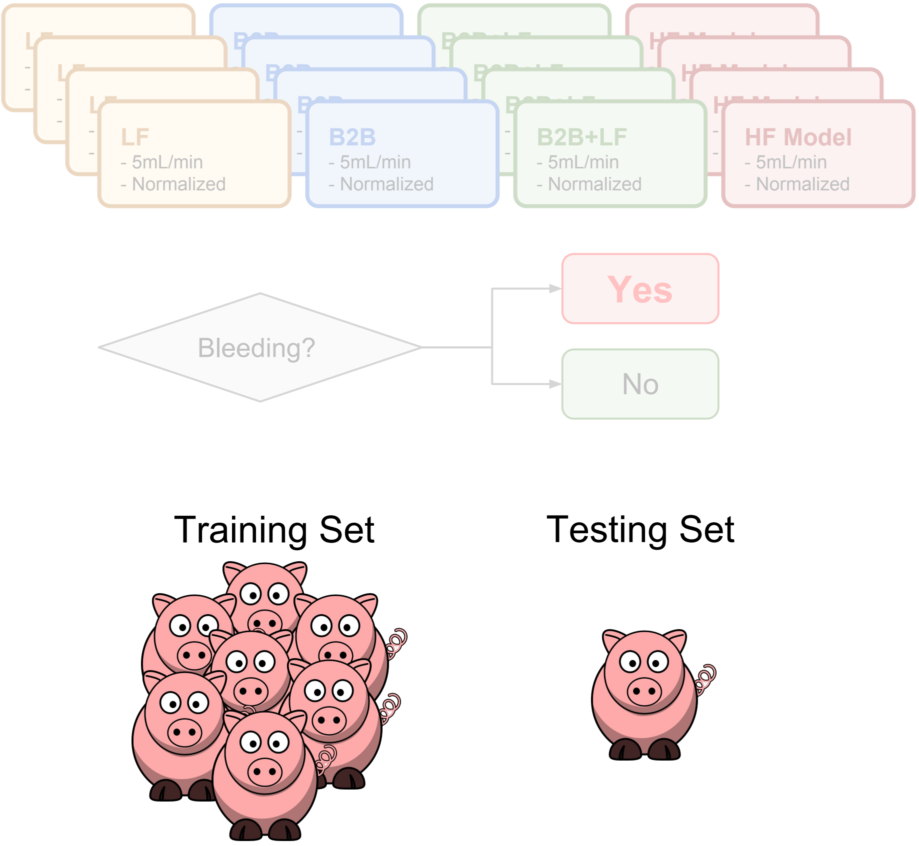
Evaluation Methodology

- Separate models are trained for combinations of
  - 5mL/min (n=14) or 20mL/min (n=46)
  - LF, B2B, B2B+LF, HF
  - No normalization or Individual baseline normalized
  - 16 different models in total.
- Each model is a random forest classifier built to distinguish non-bleeding instances (before t=0) from bleeding instances (after t=0).



Evaluation Methodology

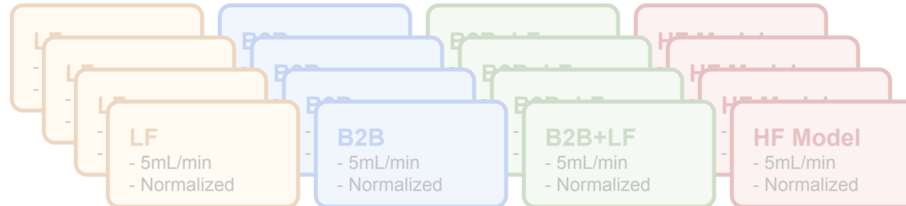
- Separate models are trained for combinations of
  - 5mL/min (n=14) or 20mL/min (n=46)
  - LF, B2B, B2B+LF, HF
  - No normalization or Individual baseline normalized
  - 16 different models in total.
- Each model is a random forest classifier built to distinguish non-bleeding instances (before t=0) from bleeding instances (after t=0).
- Models are evaluated in a leave-one-pig-out cross validation framework.

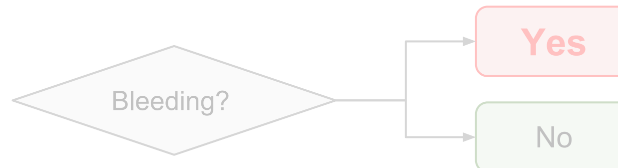


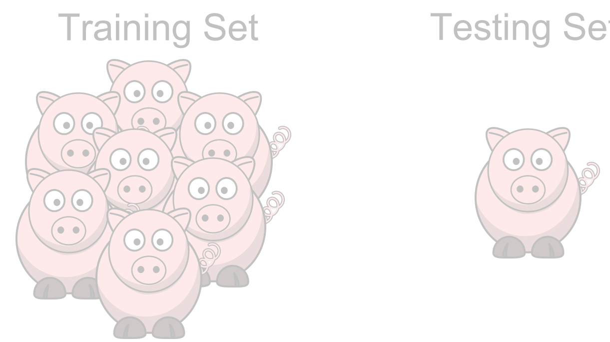
3 October 2017

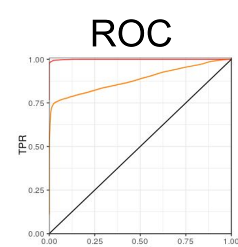
Evaluation Methodology

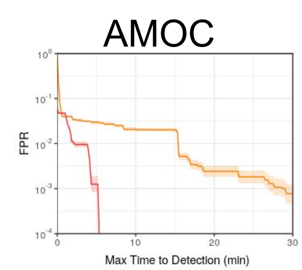
- Separate models are trained for combinations of
  - 5mL/min (n=14) or 20mL/min (n=46)
  - LF, B2B, B2B+LF, HF
  - No normalization or Individual baseline normalized
  - 16 different models in total.
- Each model is a random forest classifier built to distinguish non-bleeding instances (before t=0) from bleeding instances (after t=0).
- Models are evaluated in a leave-one-pig-out cross validation framework.
- Model performance is evaluated using ROC and AMOC curves.











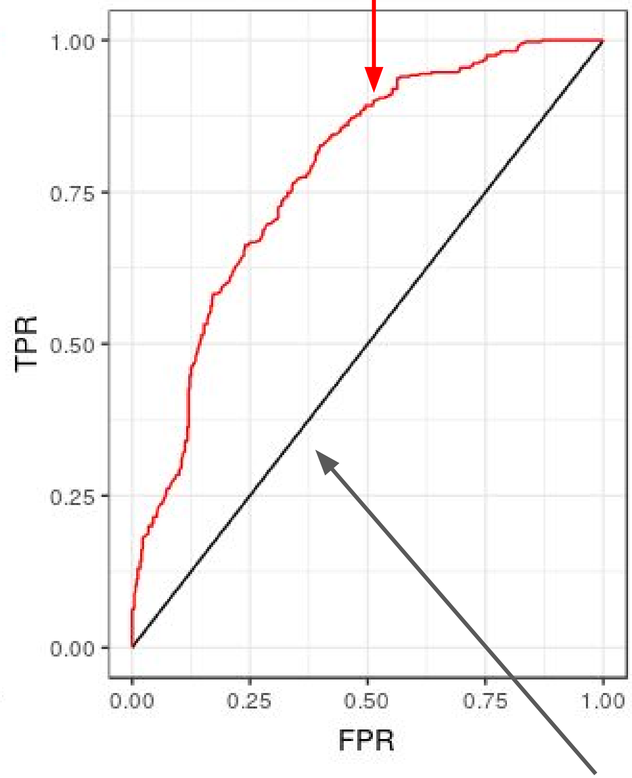
3 October 2017

Purpose of the Receiver Operating Characteristic (ROC) Curve

An ROC curve characterizes the performance tradeoffs made when tuning a classifier threshold.

We generally include at least a **random choice model** and one or more **other models** we want to compare.

This is the model we want to compare with random guessing.



This model (call it "Random") chooses a class at random with uniform probability.

Purpose of the ROC Curve

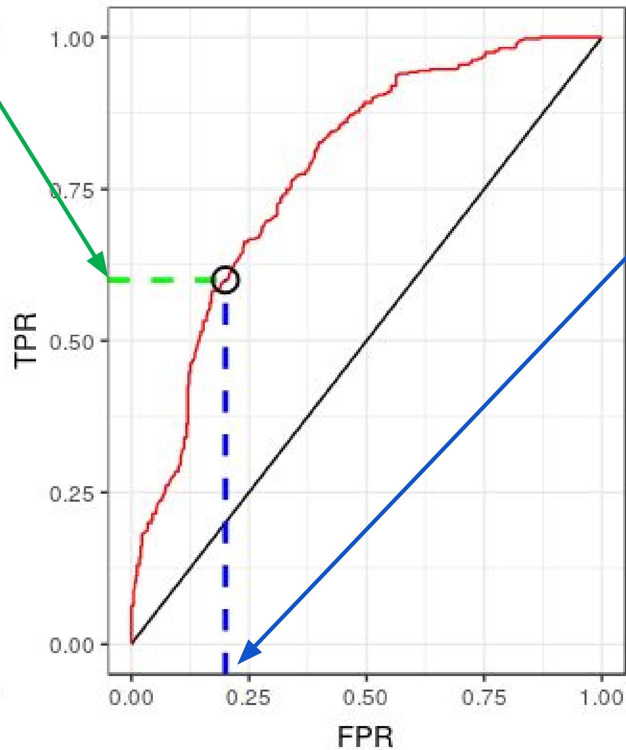
TPR: What fraction of the positive cases did we correctly identify?

An ROC curve characterizes the performance tradeoffs made when tuning a classifier threshold.

We generally include at least a **random choice model** and one or more **other models** we want to compare.

This curve characterizes the tradeoff between improving true positive rate (TPR) or false positive rate (FPR).

For a given FPR we can lookup the expected TPR.



FPR: How often are we incorrectly alerting of a condition that is not really present?

(Or: How much do the nurses hate the new monitor?)

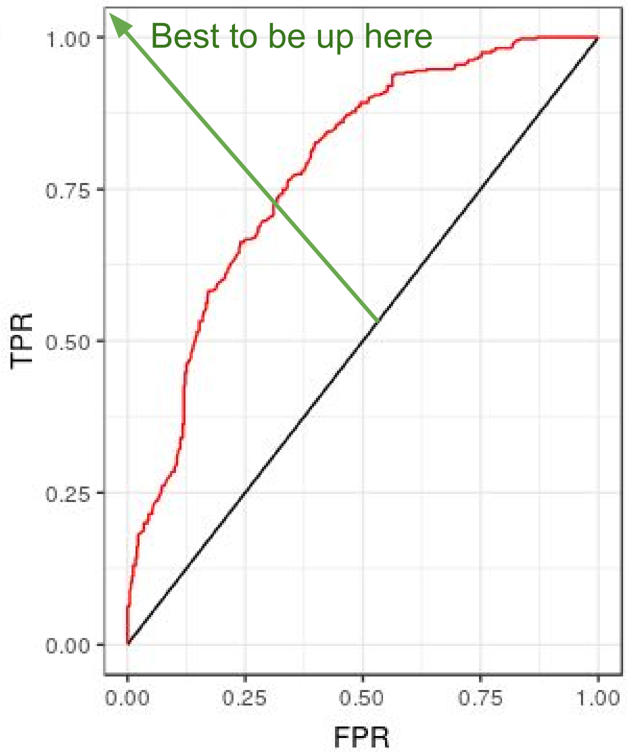
Evaluating an ROC Curve

An ROC curve characterizes the performance tradeoffs made when tuning a classifier threshold.

We generally include at least a **random choice model** and one or more **other models** we want to compare.

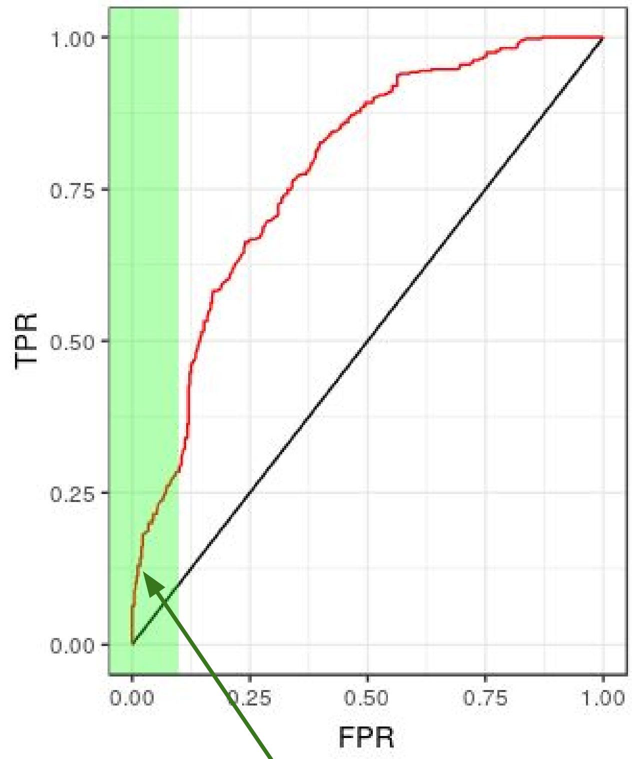
This curve characterizes the tradeoff between improving true positive rate (TPR) or false positive rate (FPR).

For a given FPR we can lookup the expected TPR.



A better performing classifier will tend to move the curve toward the top left corner (i.e. more positive detections made with fewer false detections).

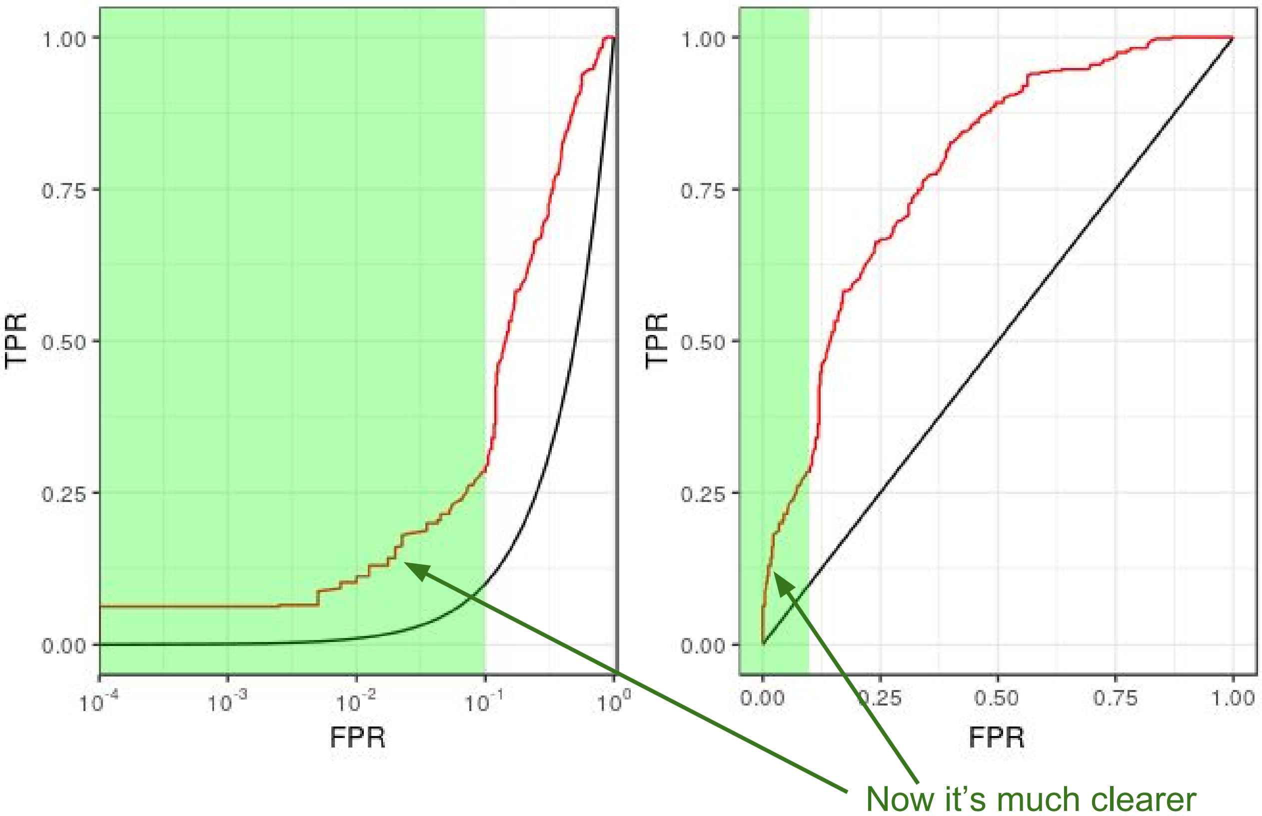
Low False Positive Rates on an ROC Curve



We are often most interested in the low FPR range in operation...

We want to look at this region.
Low FPR = Fewer false alarms.

Low False Positive Rates on an ROC Curve

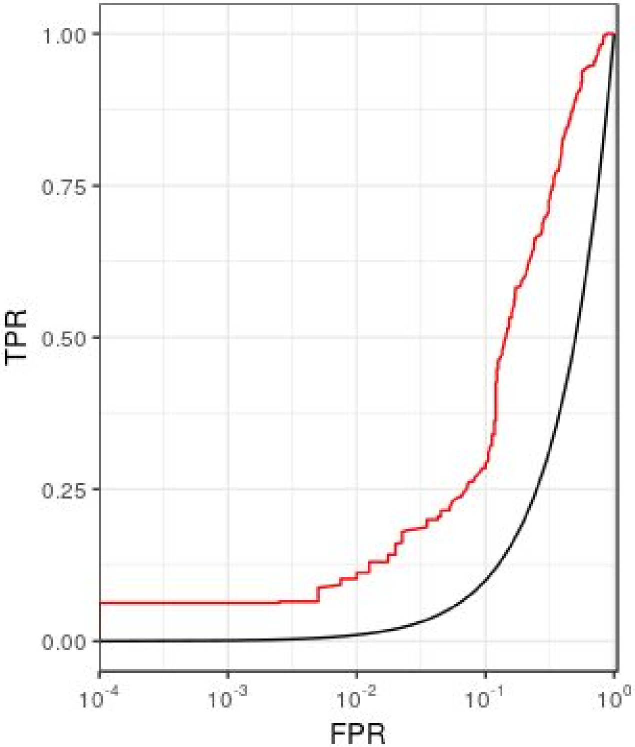


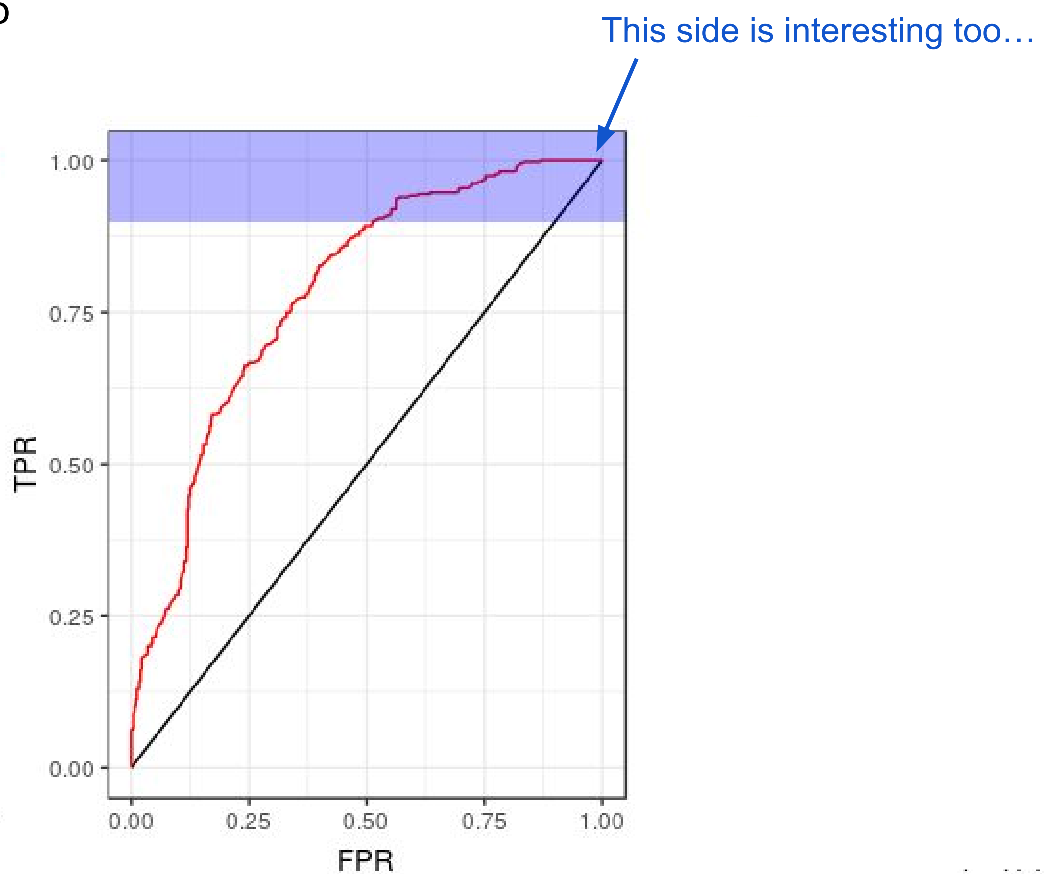
We are often most interested in the low FPR range in operation…

...so we plot FPR on the log scale to zoom in to the smaller values.

Low False Negative Rates on an ROC Curve

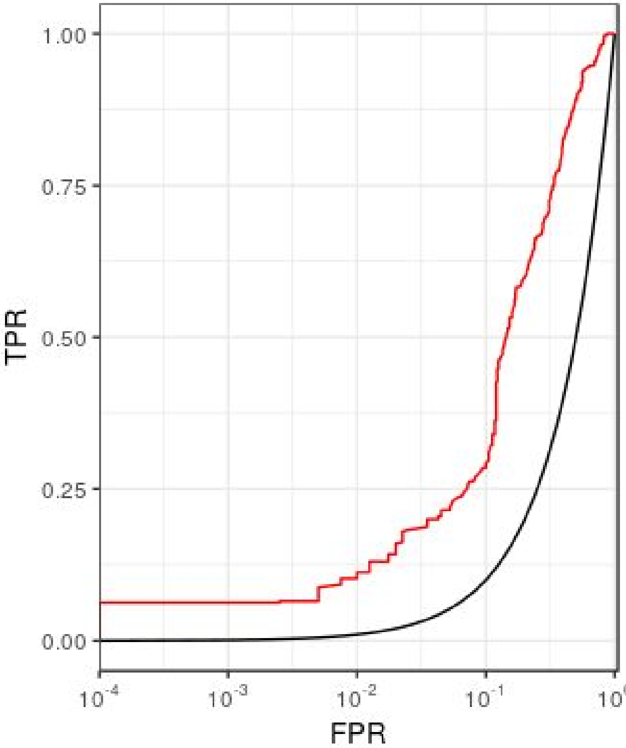
The other end of the ROC is also interesting, so we want to zoom there too…



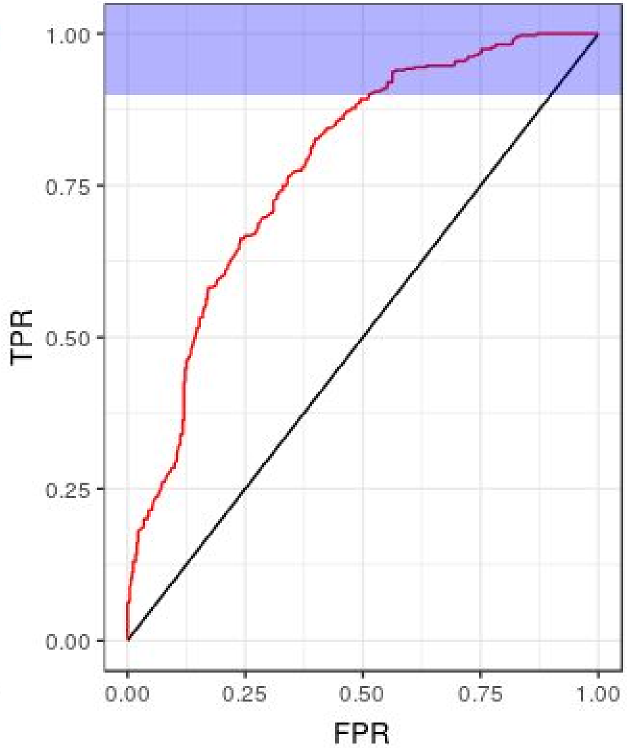


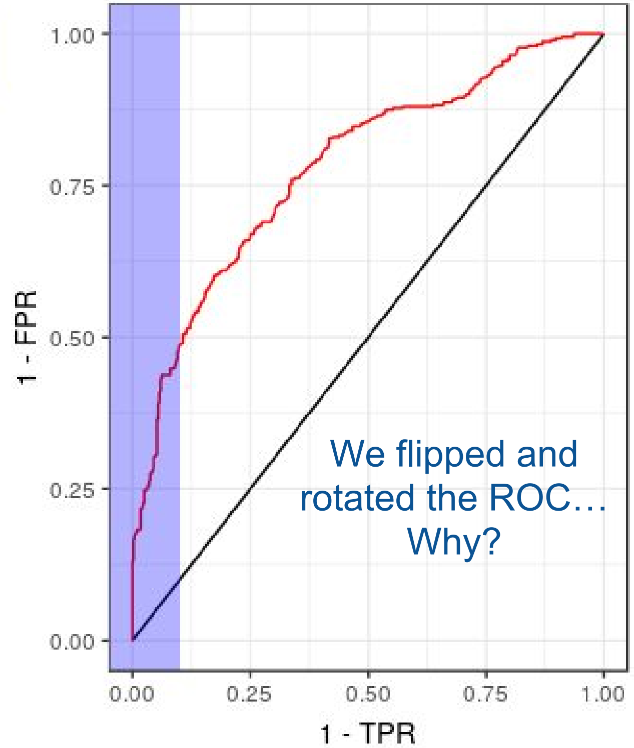
Low False Negative Rates on an ROC Curve

The other end of the ROC is also interesting, so we want to zoom there too…



…so we can invert the rates and swap the axes…

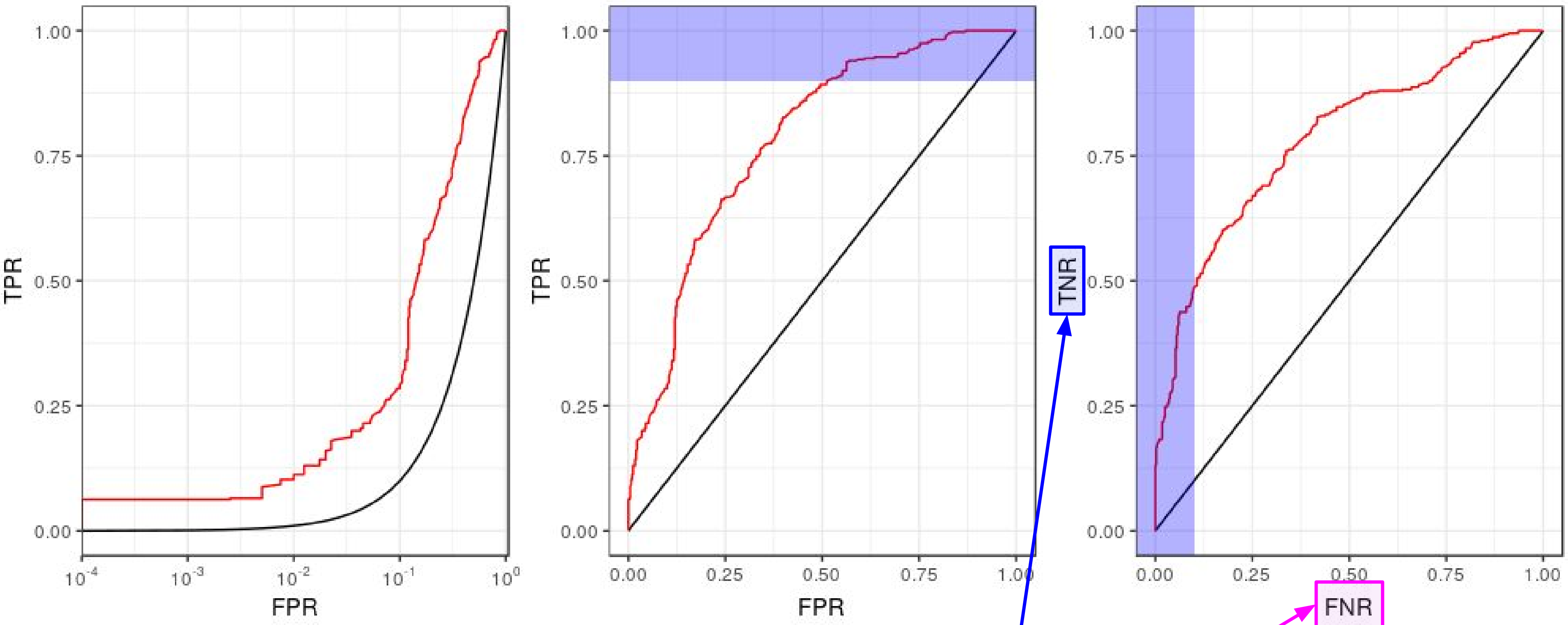




Low False Negative Rates on an ROC Curve

The other end of the ROC is also interesting, so we want to zoom there too…

…so we can invert the rates and swap their axes…



TNR: How much time do we spend on patients who need it by avoiding too much focus on those who don't?
FNR: How often do we miss something potentially really bad?

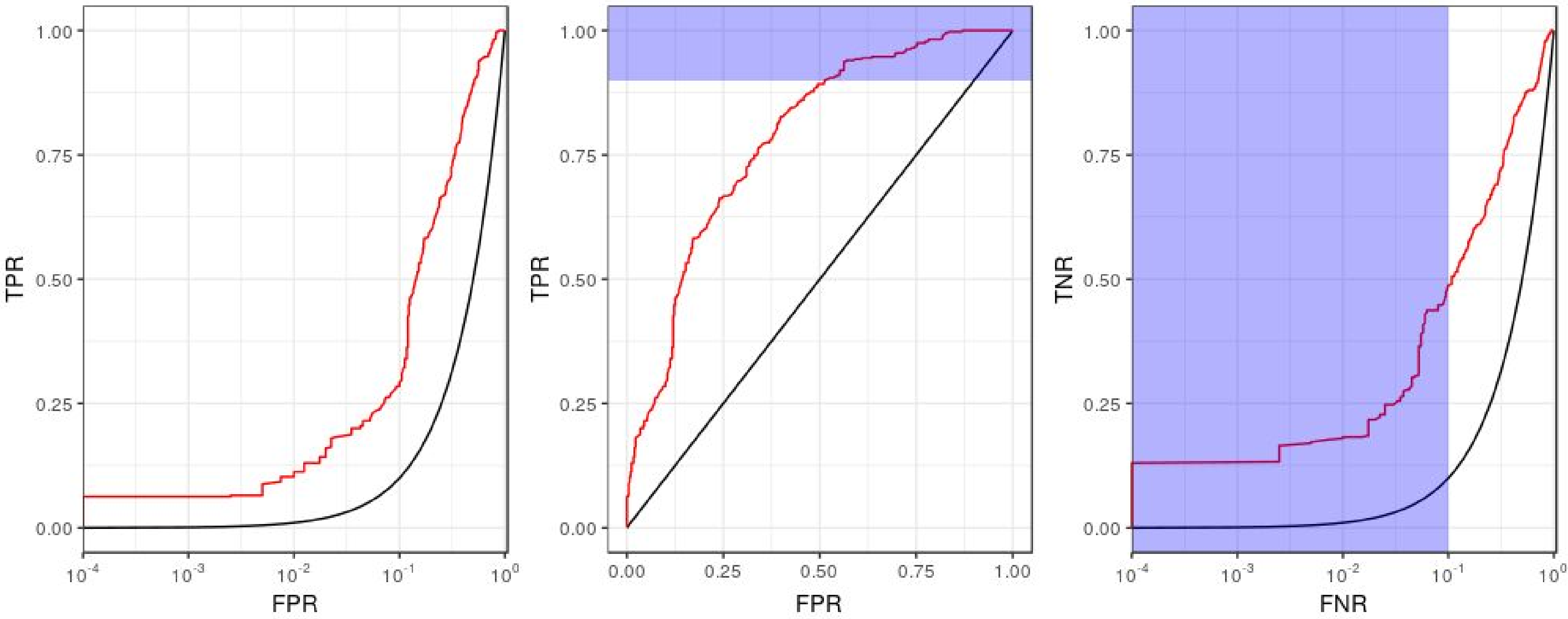
Note that:
- 1-FPR=**TNR**
- 1-TPR=**FNR**

Low False Negative Rates on an ROC Curve

The other end of the ROC is also interesting, so we want to zoom there too…

…so we can invert the rates and swap their axes…

…then plot the false negative rate (FNR) on the log scale.

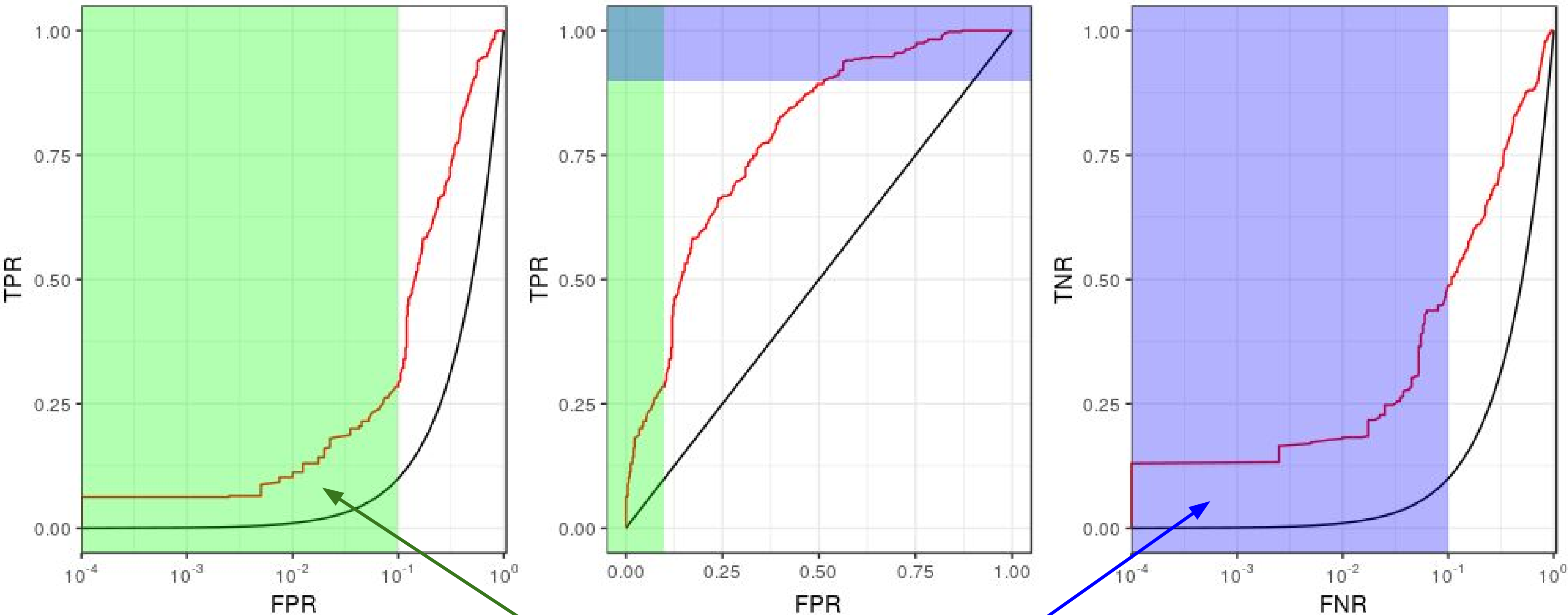


ROC Curve

The other end of the ROC is also interesting, so we want to zoom there too…

…so we can invert the rates and swap their axes…

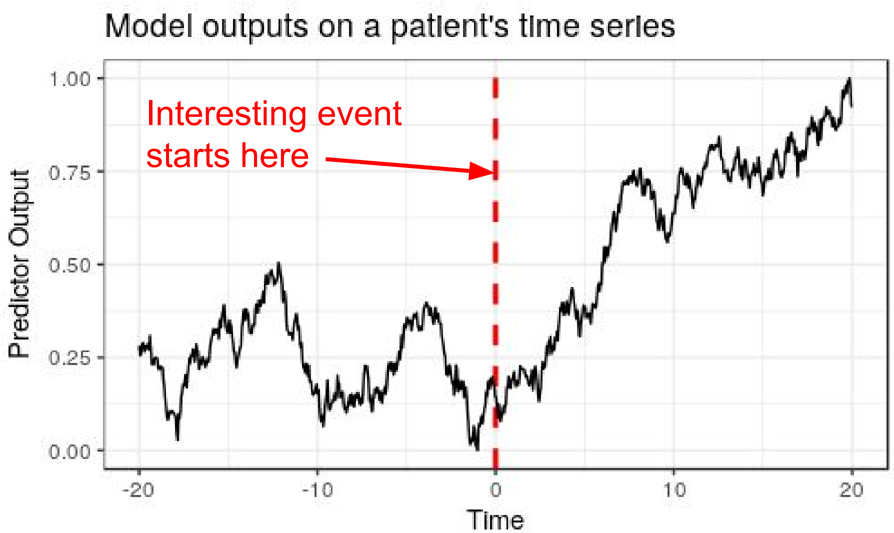
…then plot the false negative rate (FNR) on the log scale.



Now we see both interesting regions clearly!

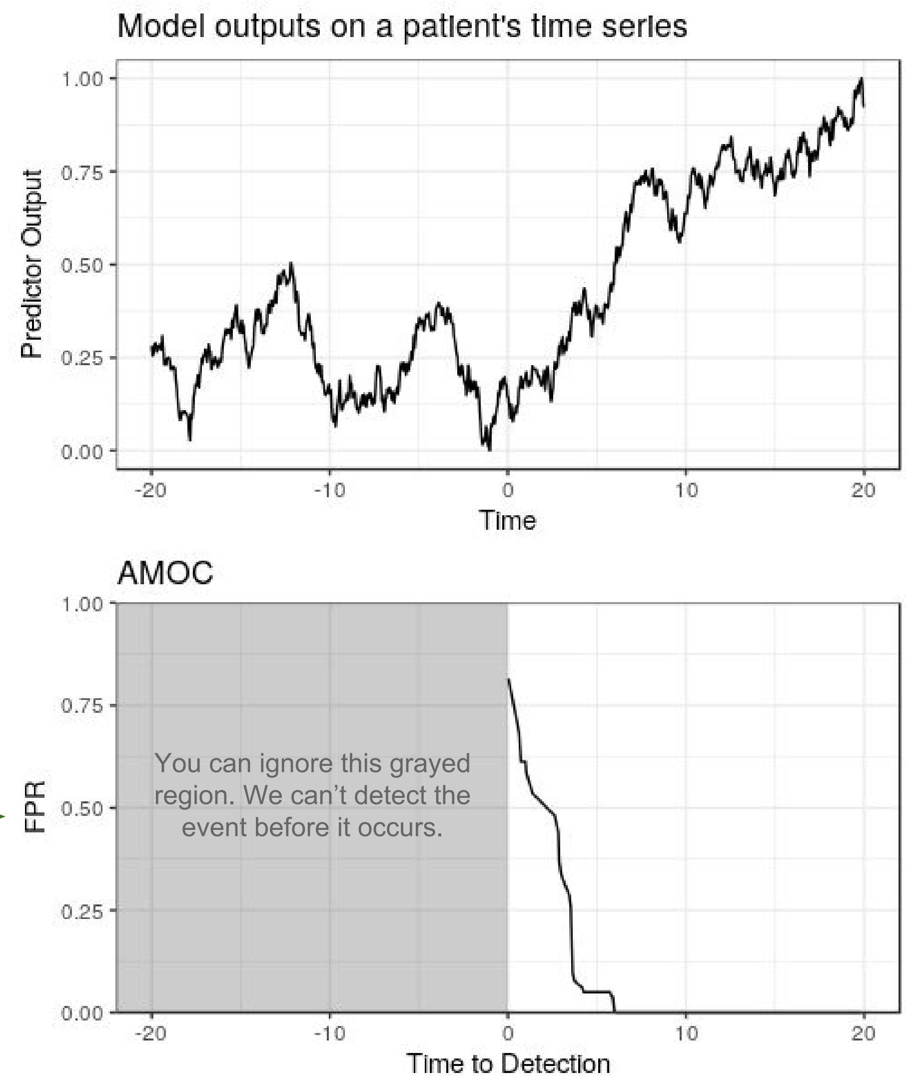
Purpose of the Activity Monitoring Operating Characteristic (AMOC) Curve

- Given a time series of predictor outputs generated by our model…



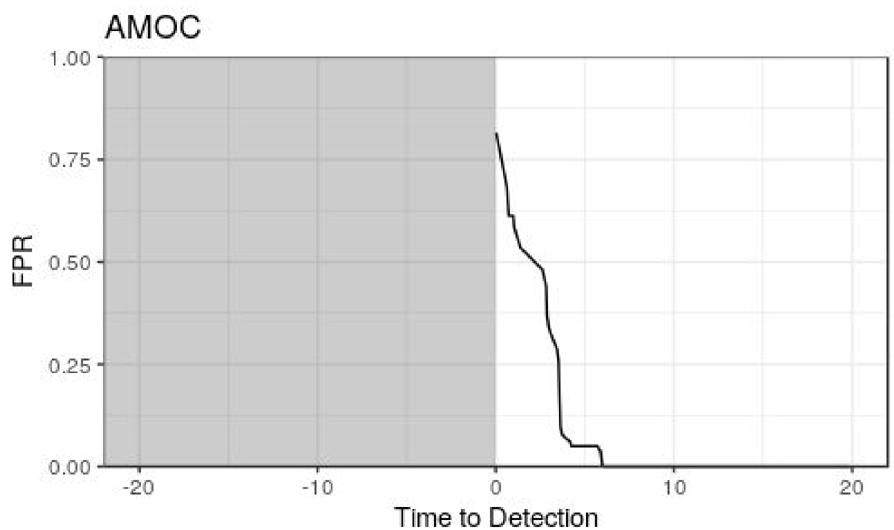
Purpose of the AMOC Curve

- Given a time series of predictor outputs generated by our model…
- …we want to characterize the **tradeoff** between detection **latency** (**time to detection**) and **false alarms** (**FPR**).



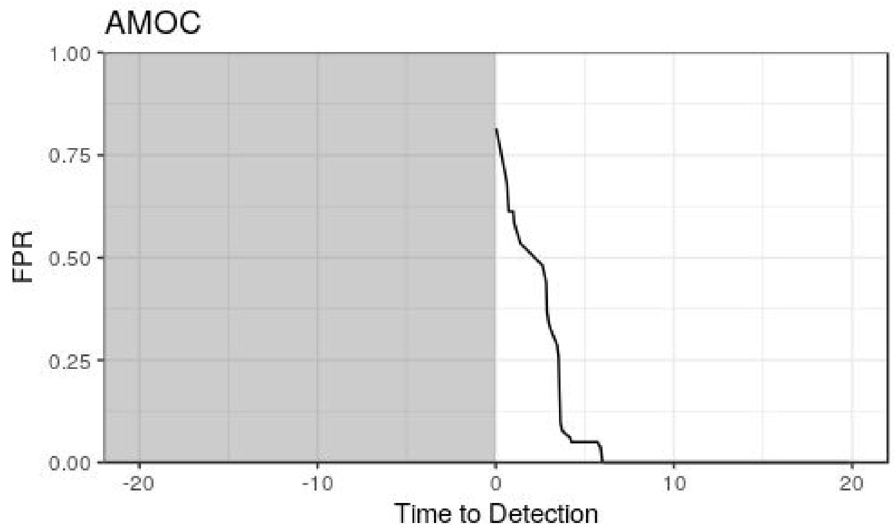
Computing an AMOC

- Given a time series of predictor outputs generated by our model…
- …we want to characterize the tradeoff between detection latency (time to detection) and false alarms (FPR).
- How do we compute this?



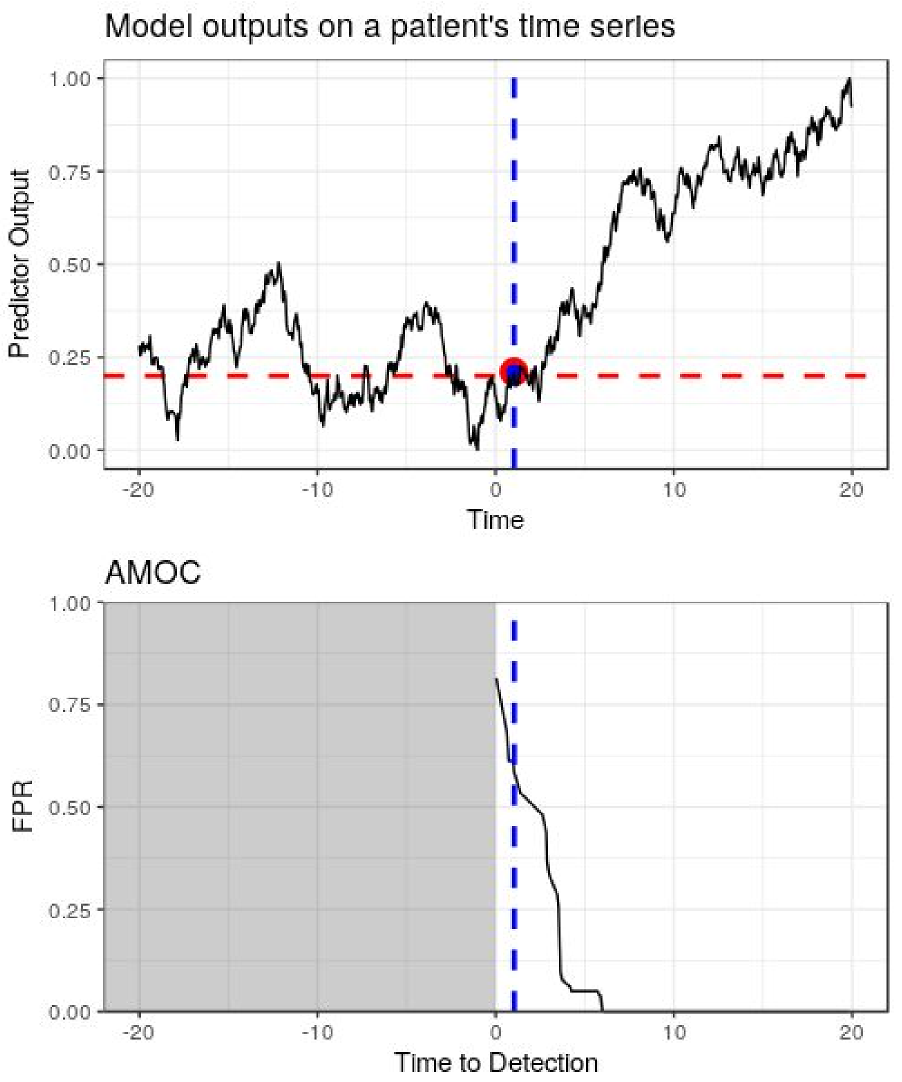
Computing an AMOC

- Given a time series of predictor outputs generated by our model…
- ...we want to characterize the tradeoff between detection latency (time to detection) and false alarms (FPR).
- How do we compute this?
  - Call a “detection” an output greater or equal to 0.2.



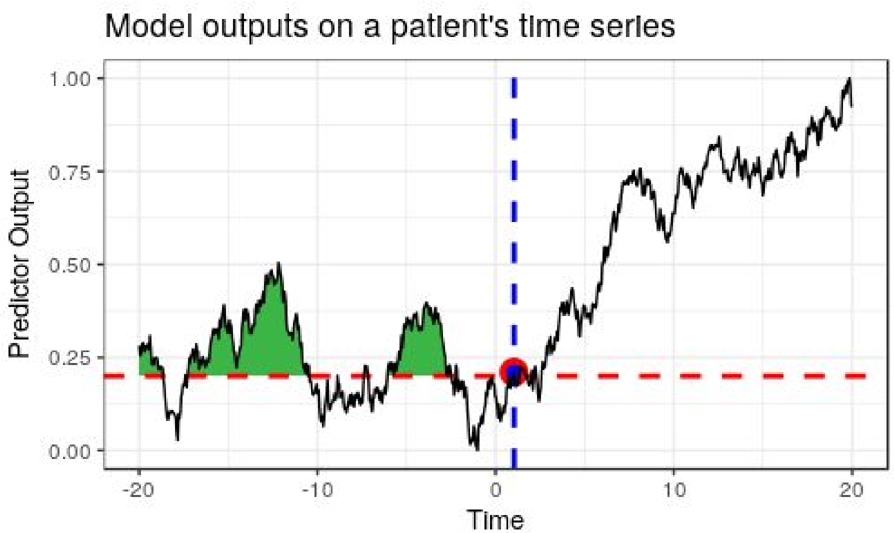
Computing an AMOC

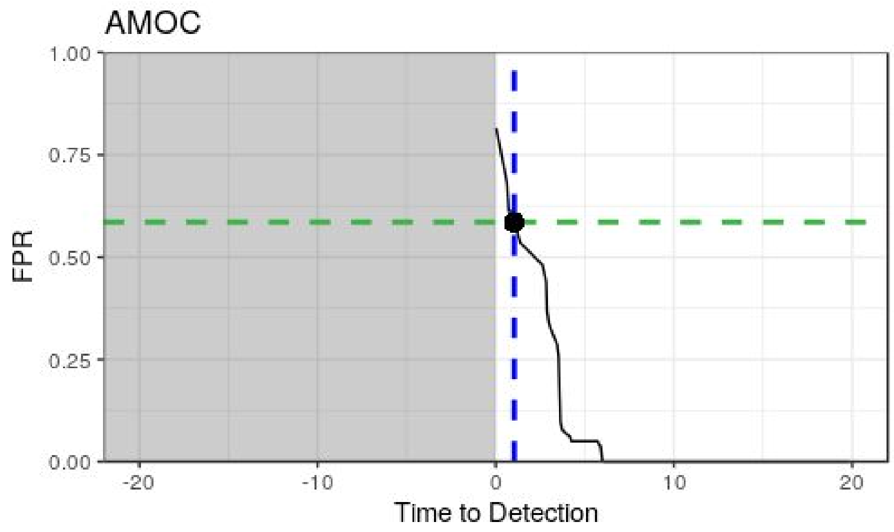
- Given a time series of predictor outputs generated by our model…
- ...we want to characterize the tradeoff between detection latency (time to detection) and false alarms (FPR).
- How do we compute this?
    - Call a “detection” an output greater or equal to 0.2. Assigning this threshold gives us
        - A time to detection (the first true positive).



Computing an AMOC

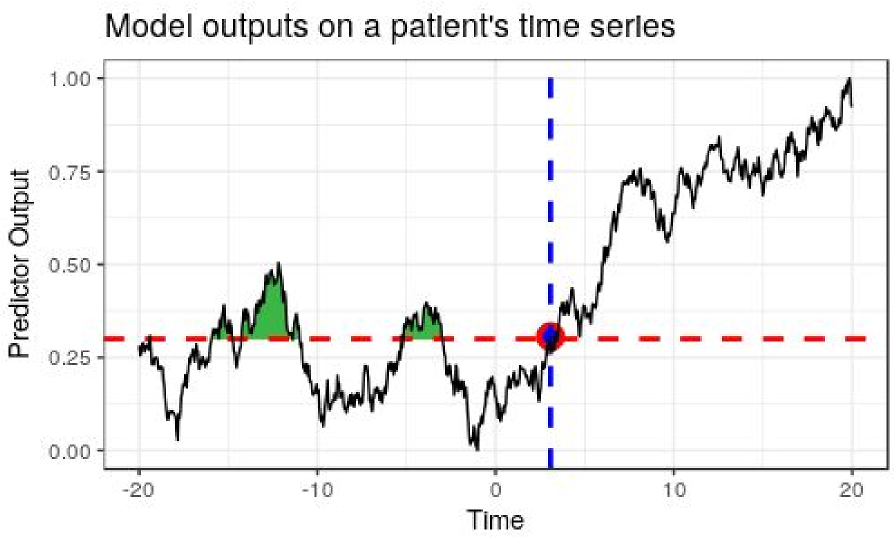
- Given a time series of predictor outputs generated by our model…
- ...we want to characterize the tradeoff between detection latency (time to detection) and false alarms (FPR).
- How do we compute this?
  - Call a “detection” an output greater or equal to 0.2. Assigning this threshold gives us
    - A time to detection (the first true positive).
    - A number of false positives (thus, FPR).

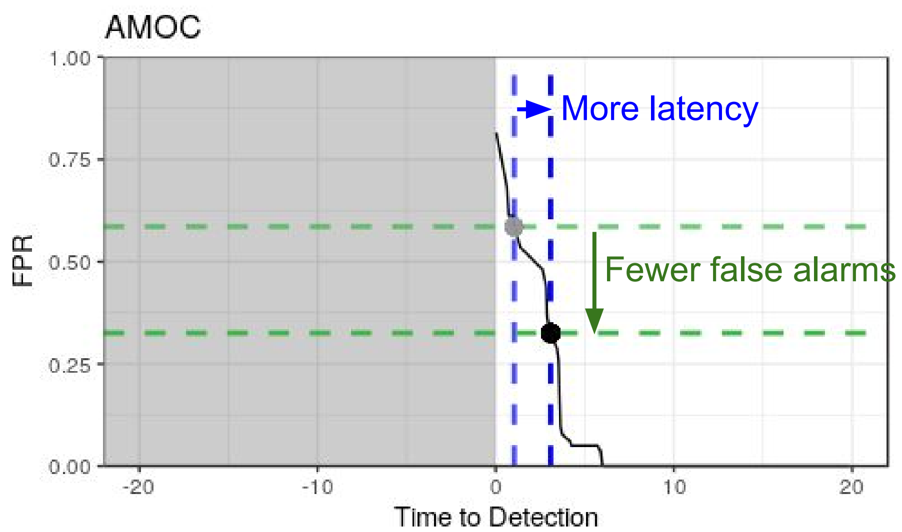




Computing an AMOC

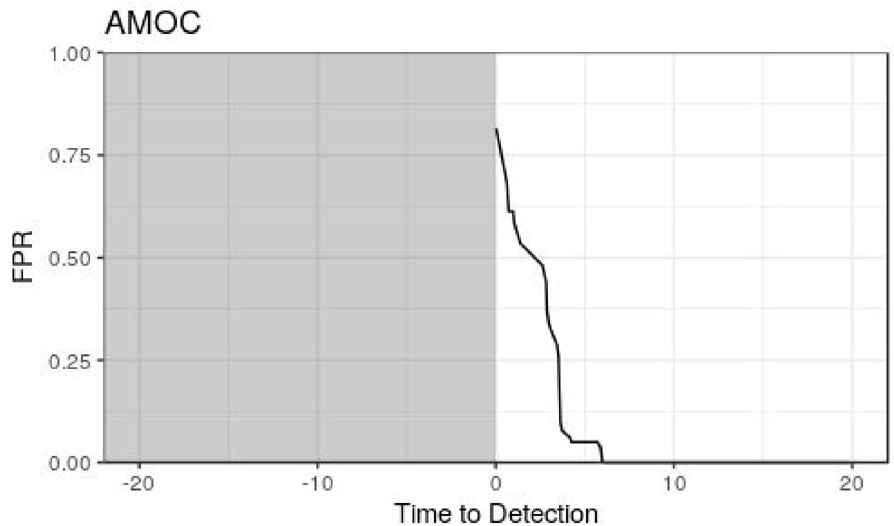
- Given a time series of predictor outputs generated by our model…
- ...we want to characterize the tradeoff between detection latency (time to detection) and false alarms (FPR).
- How do we compute this?
  - Call a “detection” an output greater or equal to 0.2. Assigning this threshold gives us
    - A time to detection (the first true positive).
    - A number of false positives (thus, FPR).
  - Do this again for another threshold, 0.3, and now there are two points on the AMOC.





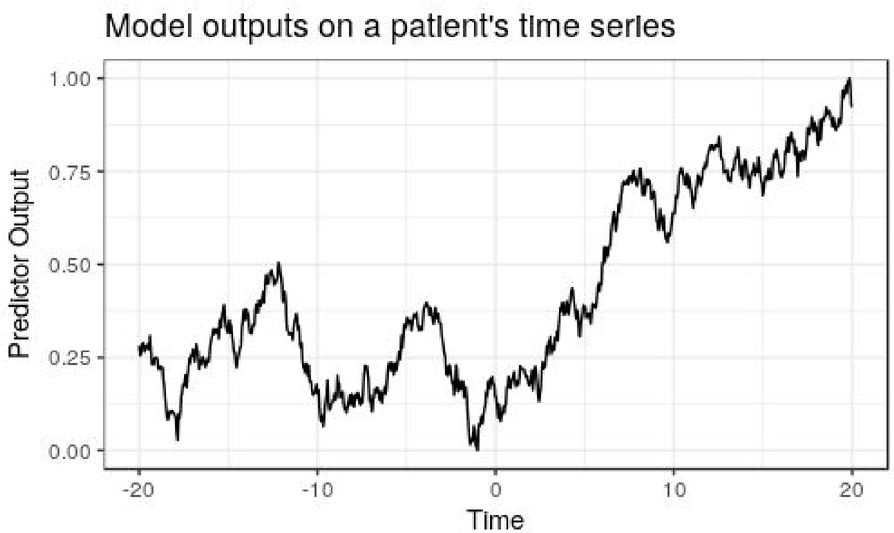
Computing an AMOC

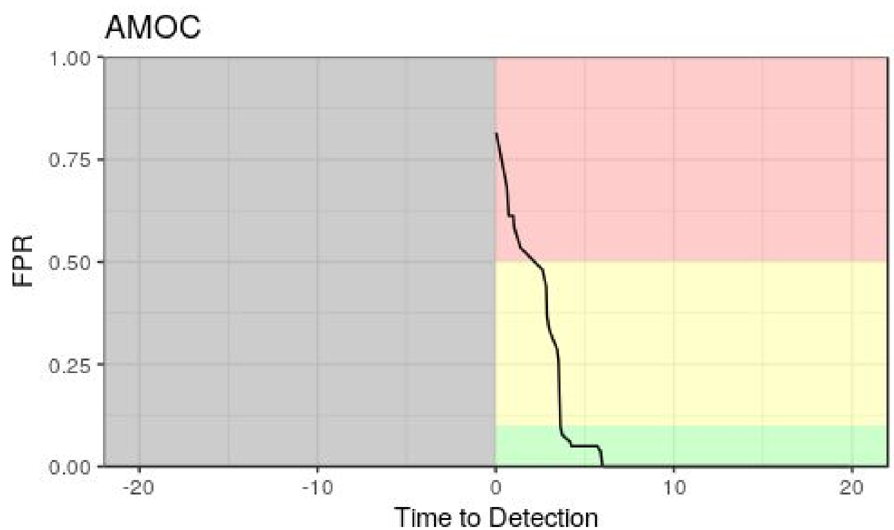
- Given a time series of predictor outputs generated by our model…
- ...we want to characterize the tradeoff between detection latency (time to detection) and false alarms (FPR).
- How do we compute this?
  - Call a “detection” an output greater or equal to 0.2. Assigning this threshold gives us
    - A time to detection (the first true positive).
    - A number of false positives (thus, FPR).
  - Do this again for another threshold, 0.3, and now there are two points on the AMOC.
  - Keep doing this for all thresholds for the complete curve.



Low False Positive Rates on an AMOC

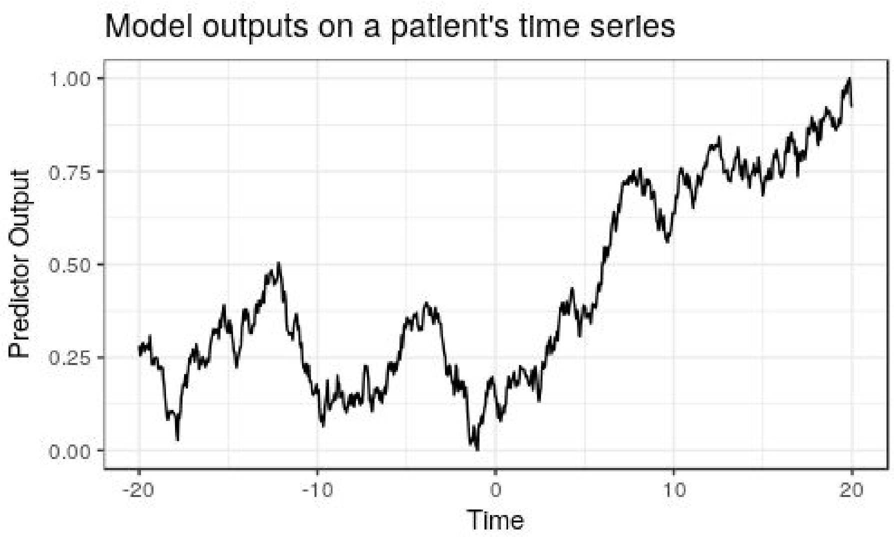
- Given a time series of predictor outputs generated by our model…
- ...we want to characterize the tradeoff between detection latency (time to detection) and false alarms (FPR).
- How do we compute this?
  - Call a “detection” an output greater or equal to 0.2. Assigning this threshold gives us
    - A time to detection (the first true positive).
    - A number of false positives (thus, FPR).
  - Do this again for another threshold, 0.3, and now there are two points on the AMOC.
  - Keep doing this for all thresholds for the complete curve.
- Lower FPR values are generally more operationally useful…

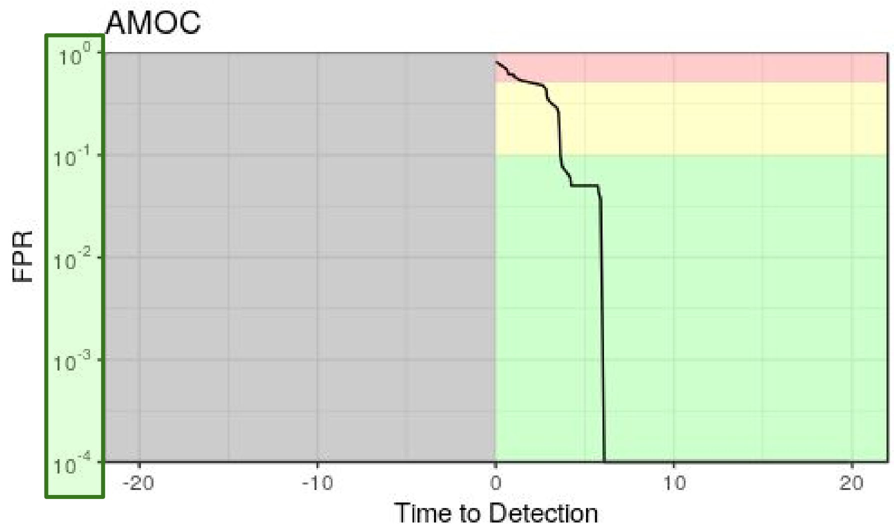




Low False Positive Rates on an AMOC

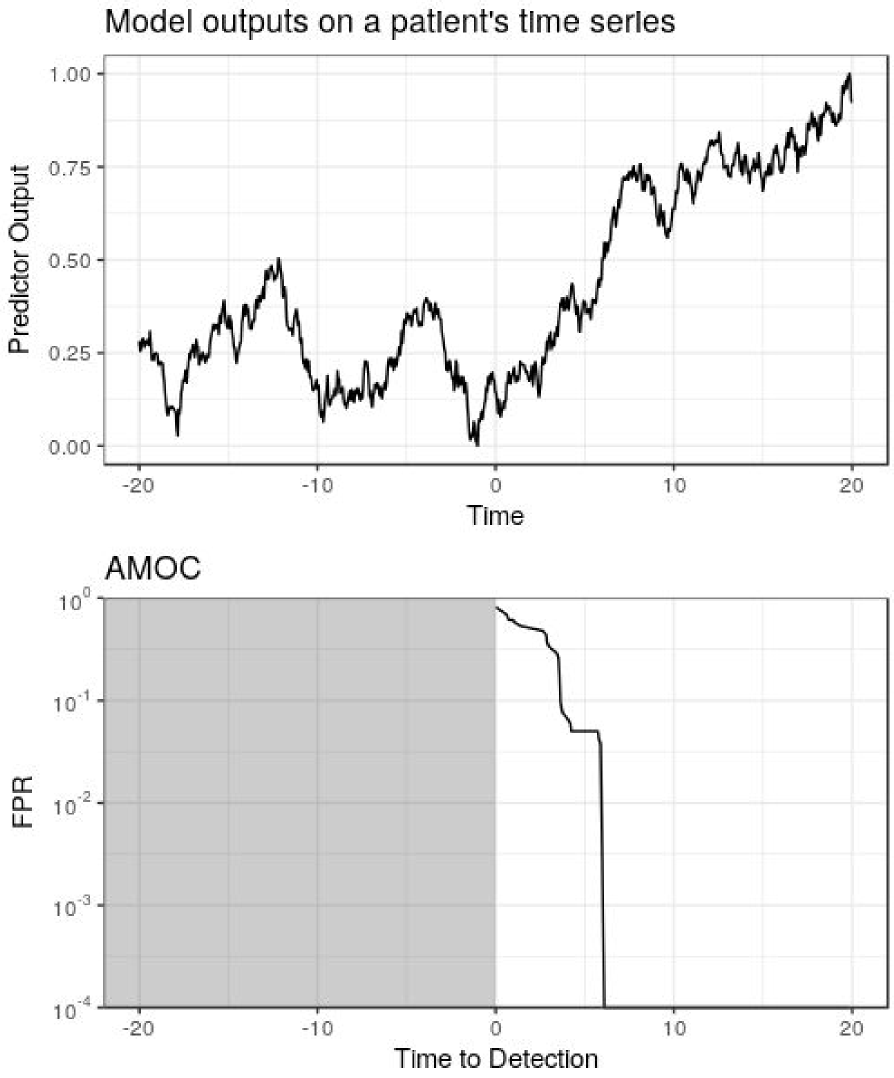
- Given a time series of predictor outputs generated by our model…
- ...we want to characterize the tradeoff between detection latency (time to detection) and false alarms (FPR).
- How do we compute this?
  - Call a “detection” an output greater or equal to 0.2. Assigning this threshold gives us
    - A time to detection (the first true positive).
    - A number of false positives (thus, FPR).
  - Do this again for another threshold, 0.3, and now there are two points on the AMOC.
  - Keep doing this for all thresholds for the complete curve.
- Lower FPR values are generally more operationally useful… so we put FPR on the log scale to zoom in to this region.





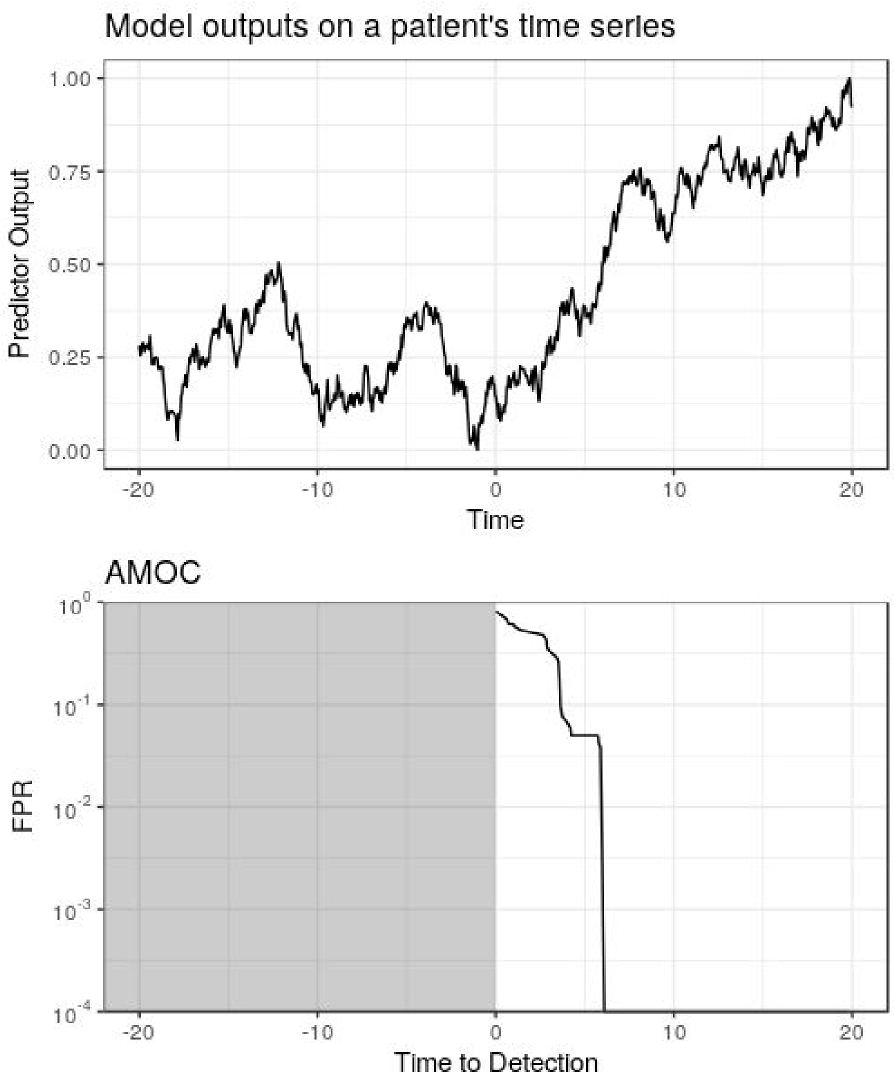
Missing Detections on AMOCs

- When computing AMOCs using multiple time series, not all series will have a detection for every output threshold.
  - If an event in a time series is not detected (i.e. zero true positives) we call that a “miss”.
  - As will be shown on subsequent slides, relaxing the minimum fraction of time series that must be detected can greatly reduce detection latency at the cost of missing some detections.

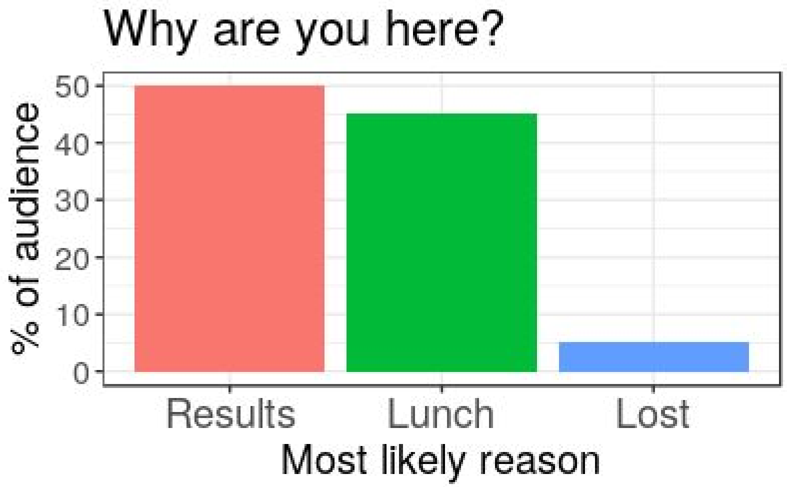


Aggregating AMOC Curves

- When computing AMOCs using multiple time series, not all series will have a detection for every output threshold.
  - If an event in a time series is not detected (i.e. zero true positives) we call that a “miss”.
  - As will be shown on subsequent slides, relaxing the minimum fraction of time series that must be detected can greatly reduce detection latency at the cost of missing some detections.
- The AMOCs on the next slides show the “maximum” time to detection for a given threshold.



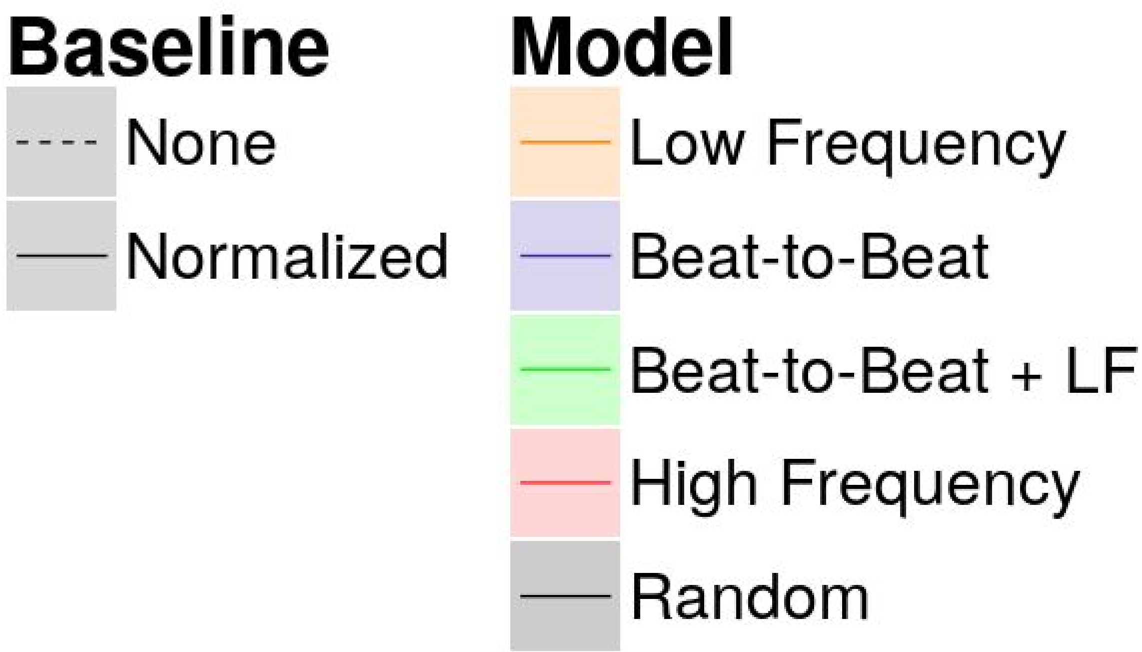
Results



Why are you here?
% of audience
50
40
30
20
10
0
Results
Lunch
Lost
Most likely reason

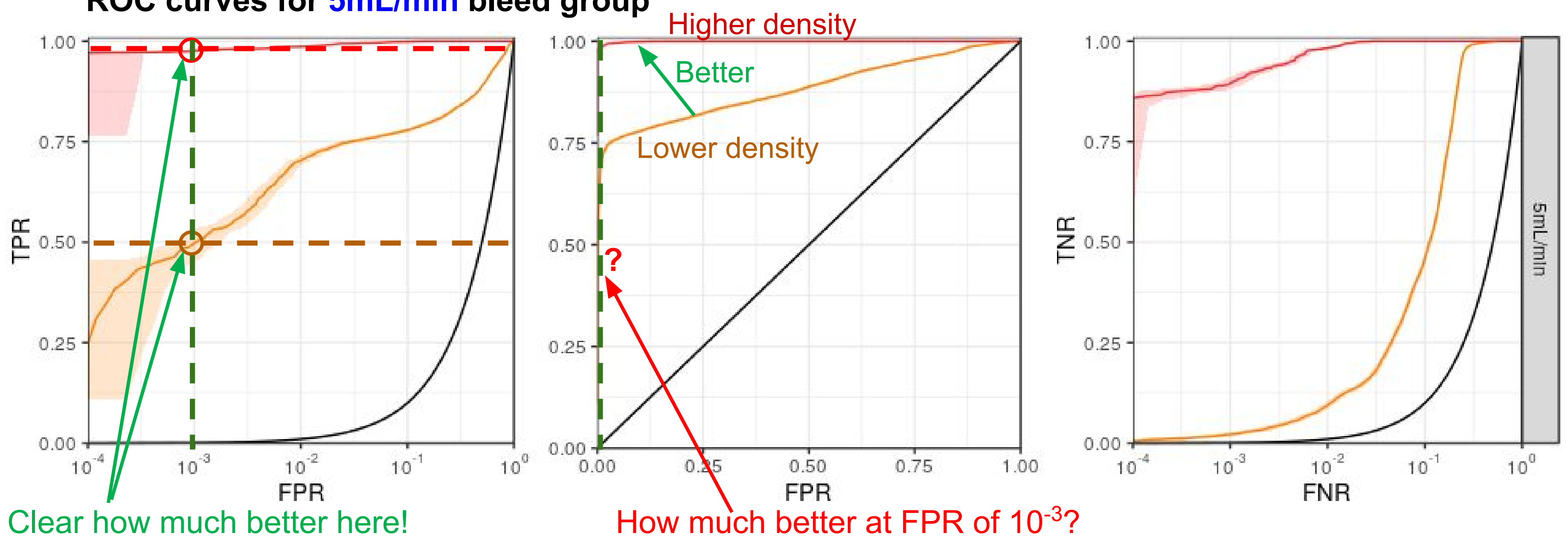
Understand the Legend

- We'll start with classifications results for the 5mL/min cohort. But first, the legend:



Performance Improves with Increasing Granularity

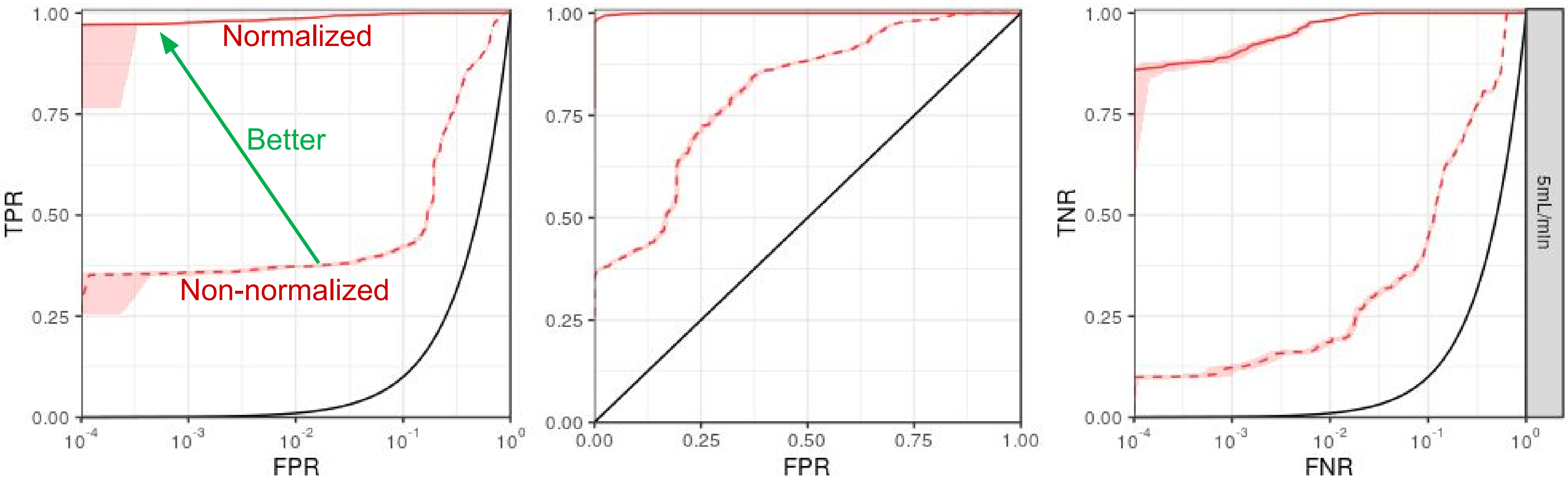
ROC curves for 5mL/min bleed group



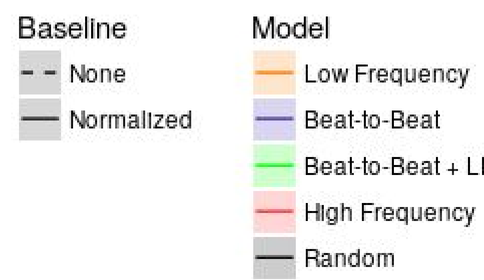
- These results are for the 5mL/min cohort.
- In general we see that greater data density yields better classification performance.

Performance Improves with Normalization

ROC curves for 5mL/min bleed group

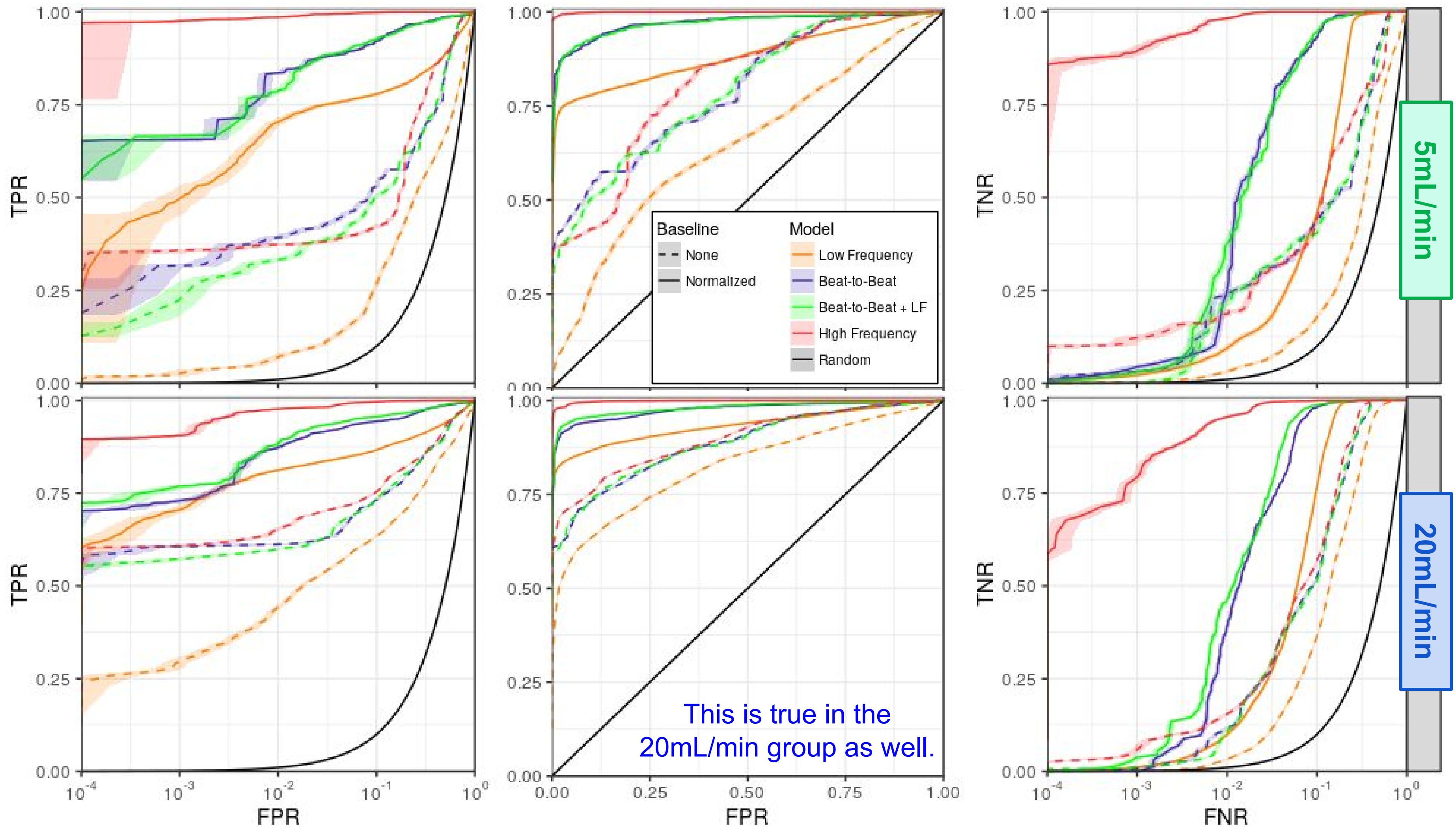


- These results are for the 5mL/min cohort.
- In general we see that greater data density yields better classification performance.
- Knowledge of individual baselines vastly improves performance.



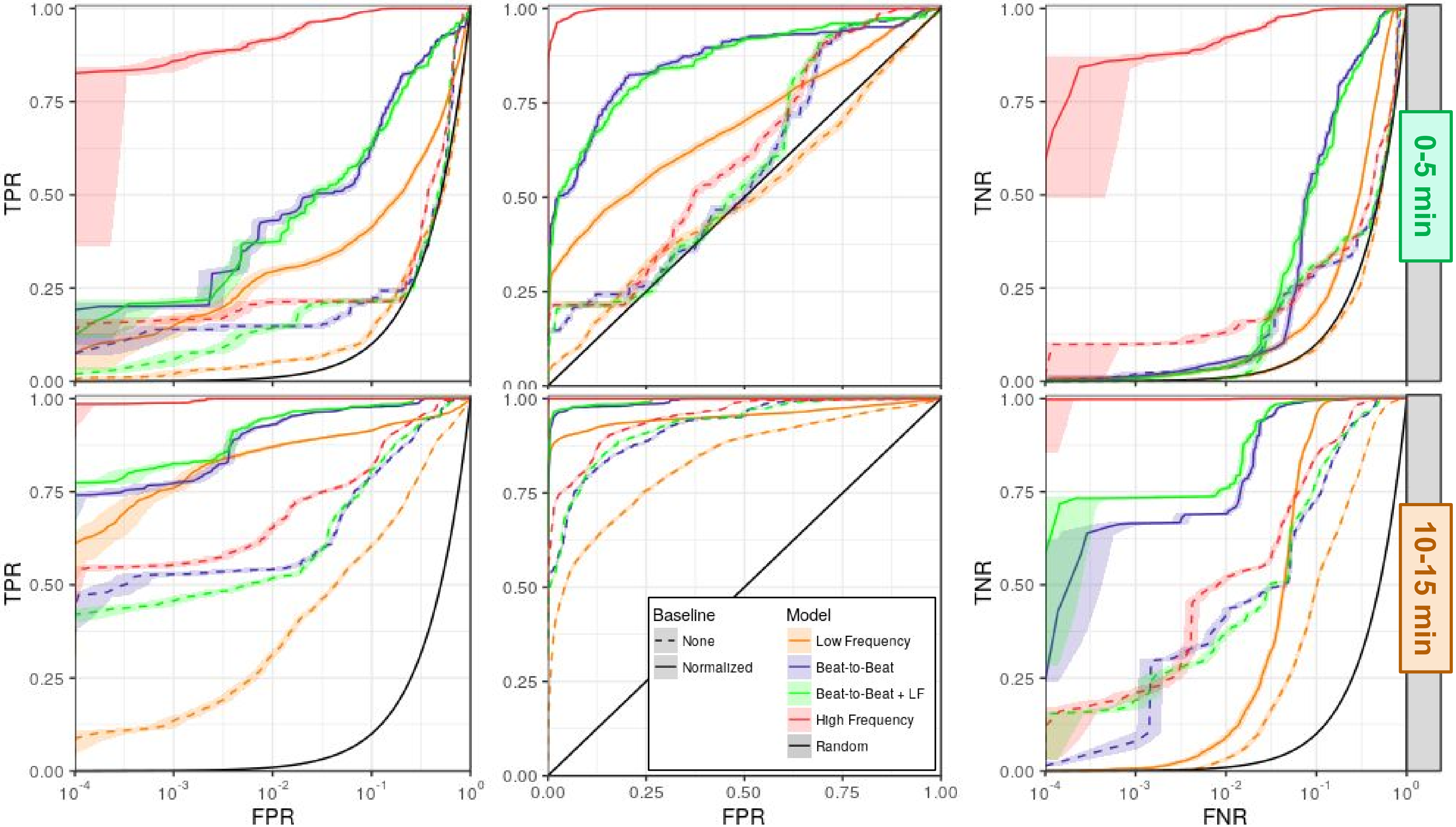
Performance is Similar Between Bleed Groups

ROC curves for 5mL/min and 20mL/min bleed groups



Early Performance is Much Better with Higher Granularity

ROC curves for 5mL/min bleed group 0-5 minutes (top) vs 10-15 minutes (bottom) into the bleed



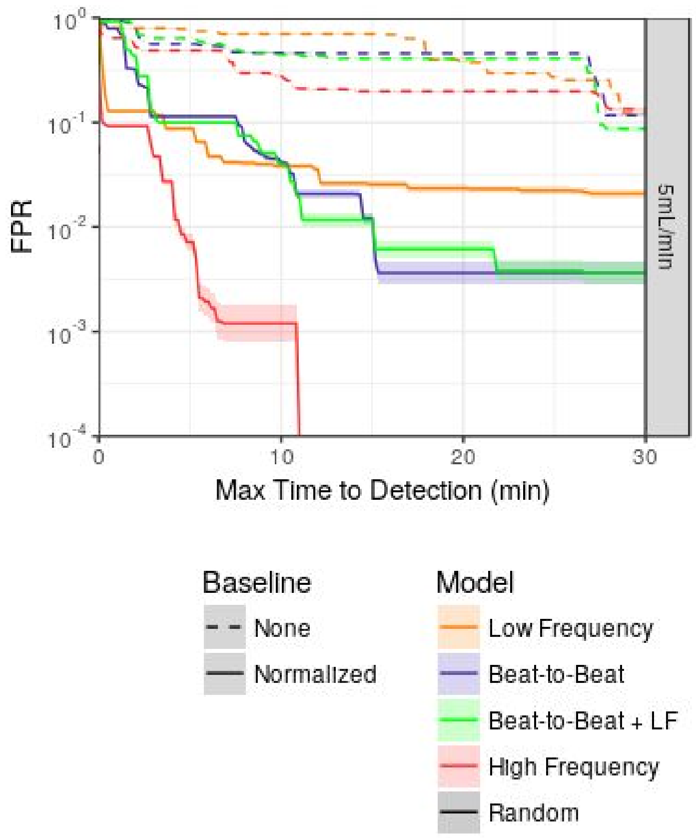
Moving from the ROC to the AMOC

- The ROC shows of the trade-off between correct and incorrect classifications.
- What about the timeliness of a detection?

Evaluating Time to Detection (Latency)

- These results are for the 5mL/min cohort.

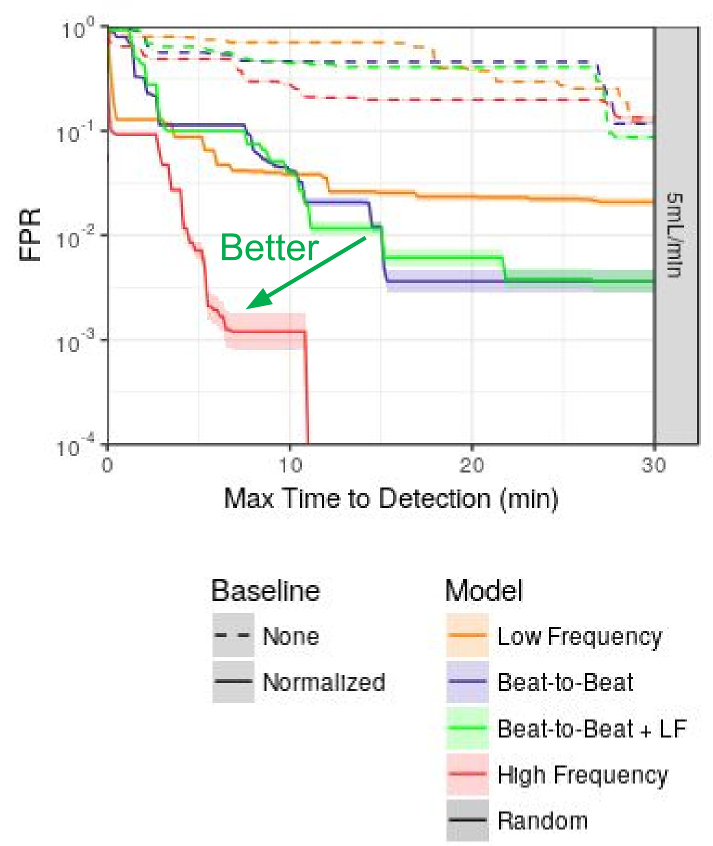
AMOC curve for 5mL/min bleed group



Evaluating Time to Detection (Latency)

- These results are for the 5mL/min cohort.
- The performance is better when we move to the bottom left of the plot (lower FPR, lower latency).

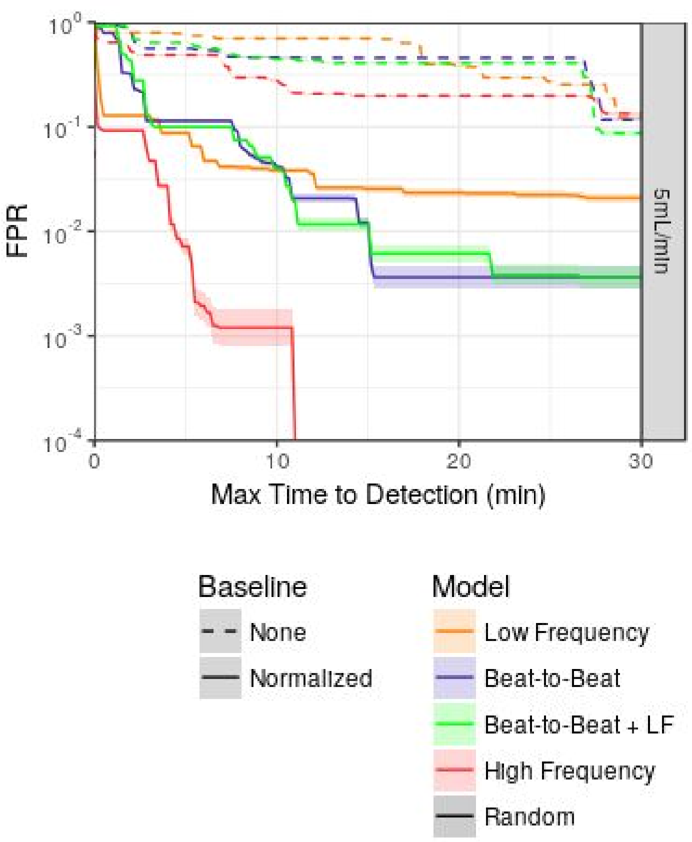
AMOC curve for 5mL/min bleed group



Evaluating Time to Detection (Latency)

- These results are for the 5mL/min cohort.
- The performance is better when we move to the bottom left of the plot (lower FPR, lower latency).
- This AMOC enforces the constraint that a detection is made on *all* pigs…

AMOC curve for 5mL/min bleed group

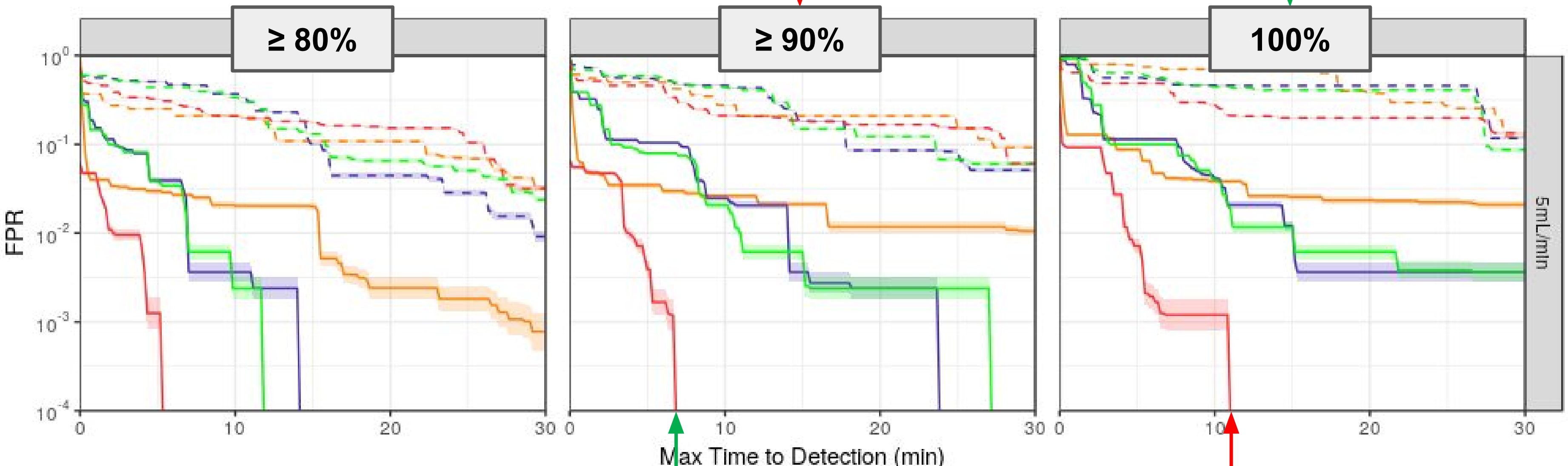


Faster Detections when Minimum Detected Fraction is Lower

AMOC curves for 5mL/min bleed group

At least 90% of bleeds detected

All bleeds detected



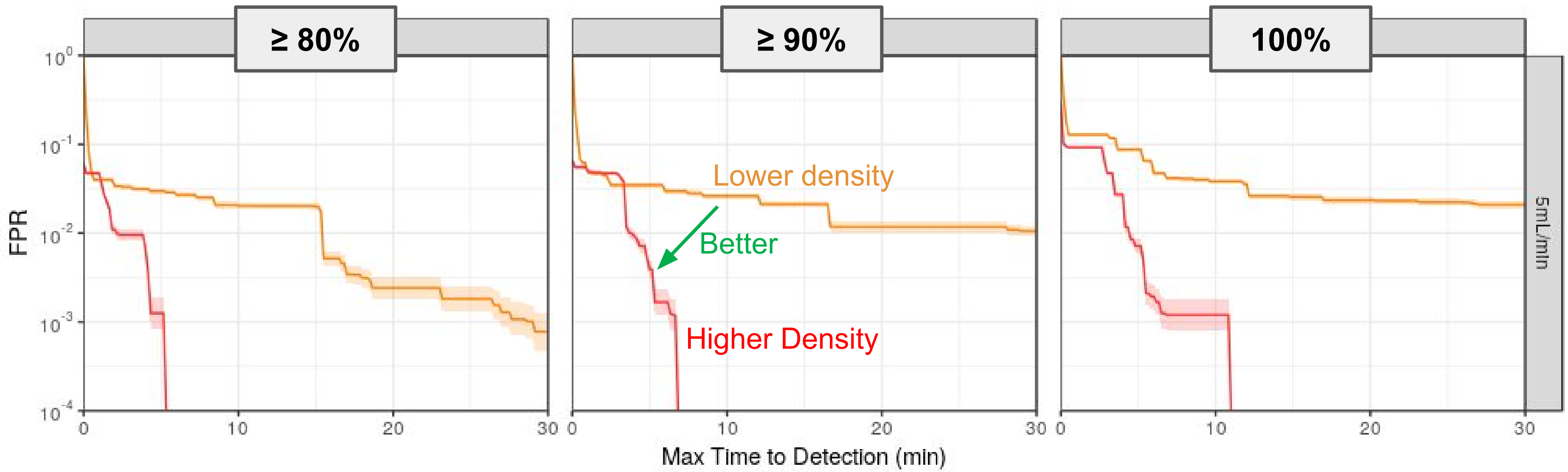
Faster detection

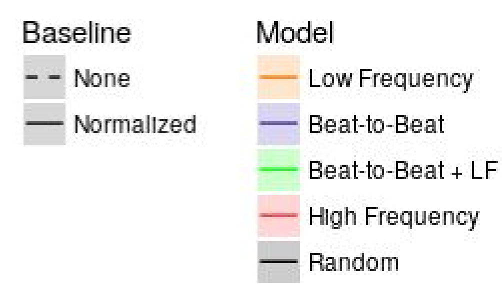
Slower detection

- These results are for the 5mL/min cohort.
- The performance is better when we move to the bottom left of the plot (lower FPR, lower latency).
- This AMOC enforces the constraint that a detection is made on *all* pigs… but we can loosen that constraint for faster detections at the expense of some missed detections.

Faster Detections and Fewer False Alarms with Higher Granularity

AMOC curves for 5mL/min bleed group

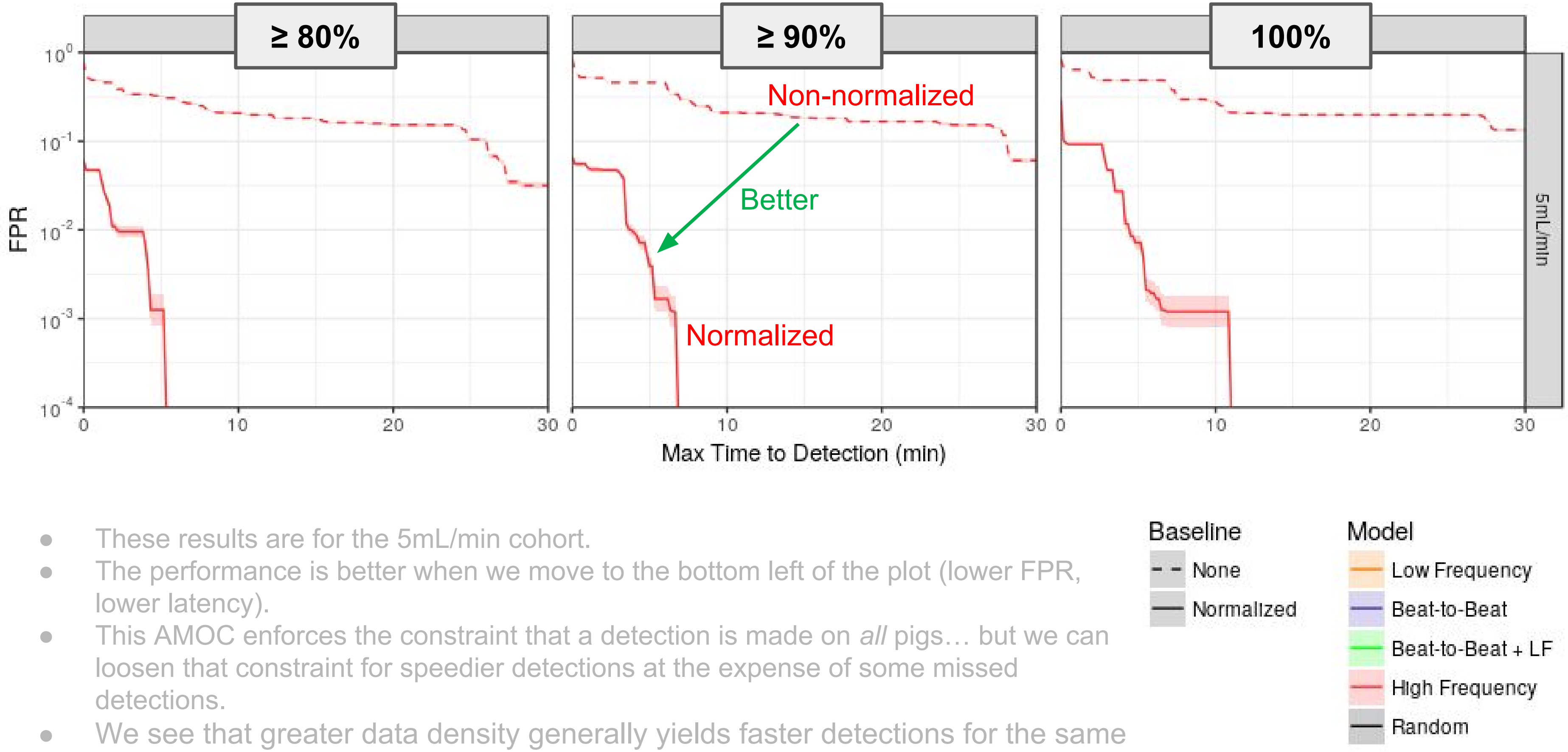




- These results are for the 5mL/min cohort.
- The performance is better when we move to the bottom left of the plot (lower FPR, lower latency).
- This AMOC enforces the constraint that a detection is made on *all* pigs… but we can loosen that constraint for speedier detections at the expense of some missed detections.
- We see that greater data density generally yields faster detections for the same FPR on *normalized* models.

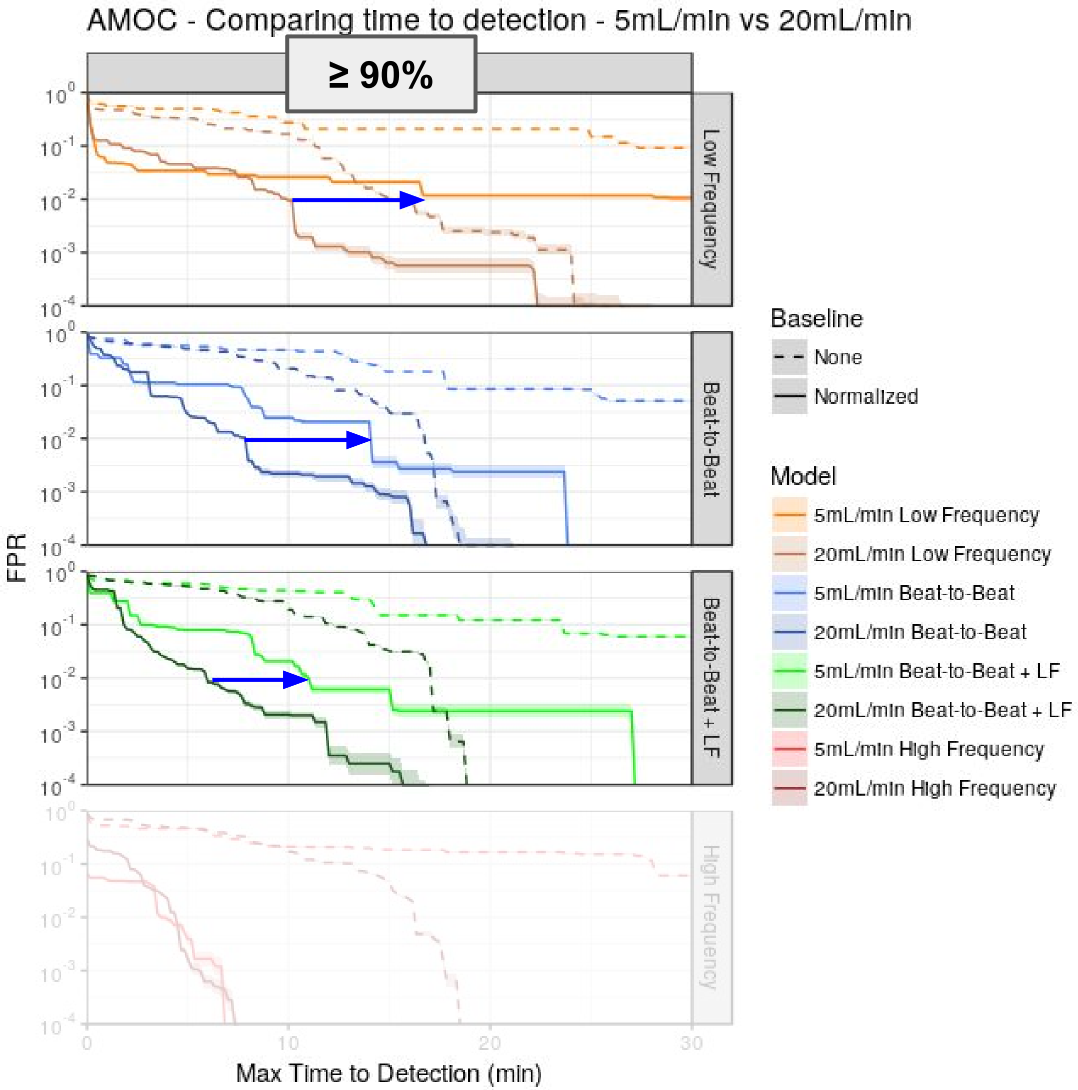
Faster Detections and Fewer False Alarms with Normalization

AMOC curves for 5mL/min bleed group



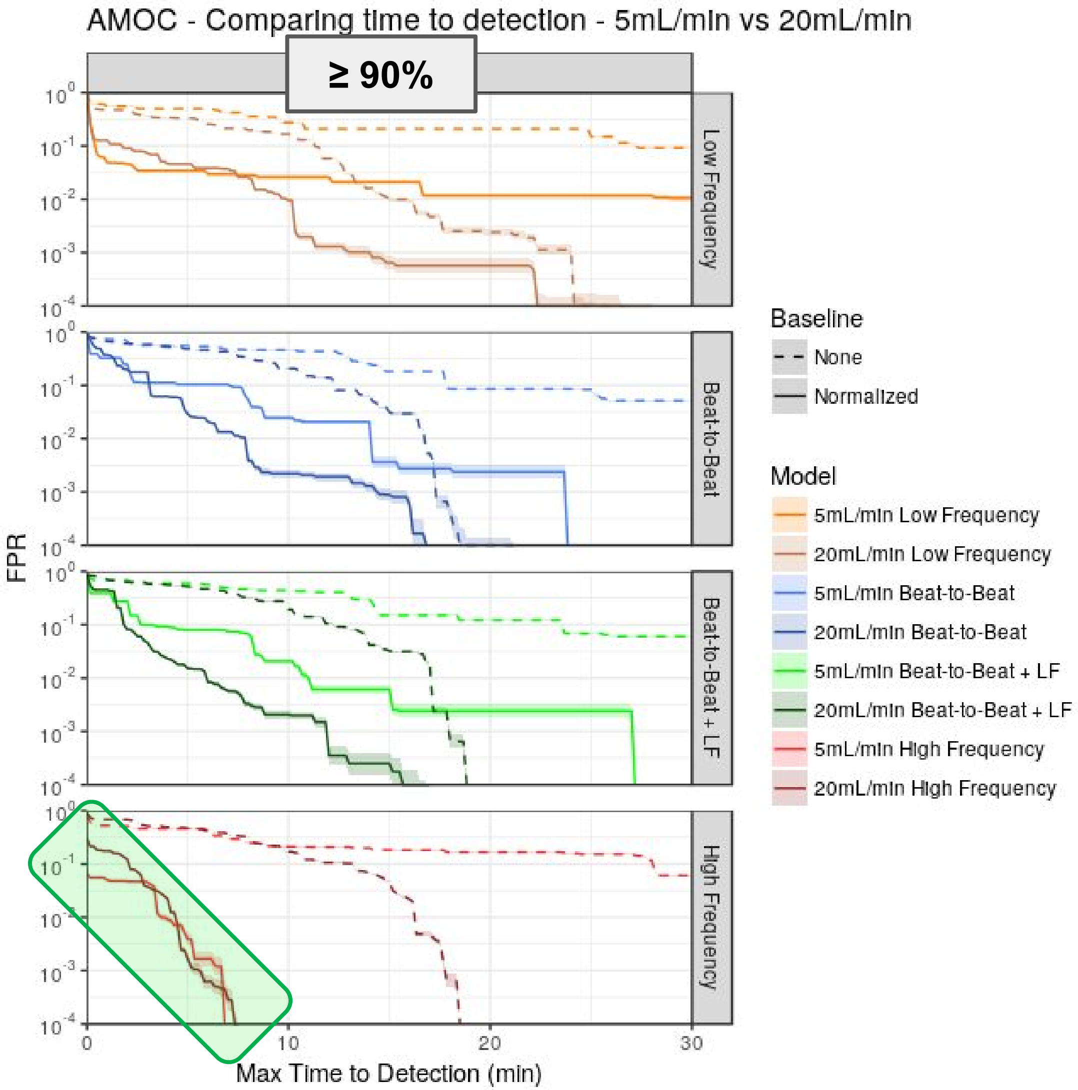
- These results are for the 5mL/min cohort.
- The performance is better when we move to the bottom left of the plot (lower FPR, lower latency).
- This AMOC enforces the constraint that a detection is made on *all* pigs… but we can loosen that constraint for speedier detections at the expense of some missed detections.
- We see that greater data density generally yields faster detections for the same FPR on *normalized* models.
- Knowledge of individual baselines allows faster detections for the same FPR.

Detection Performance by Time Latency



- Lower granularity models detect more slowly for the slower bleeding pigs.

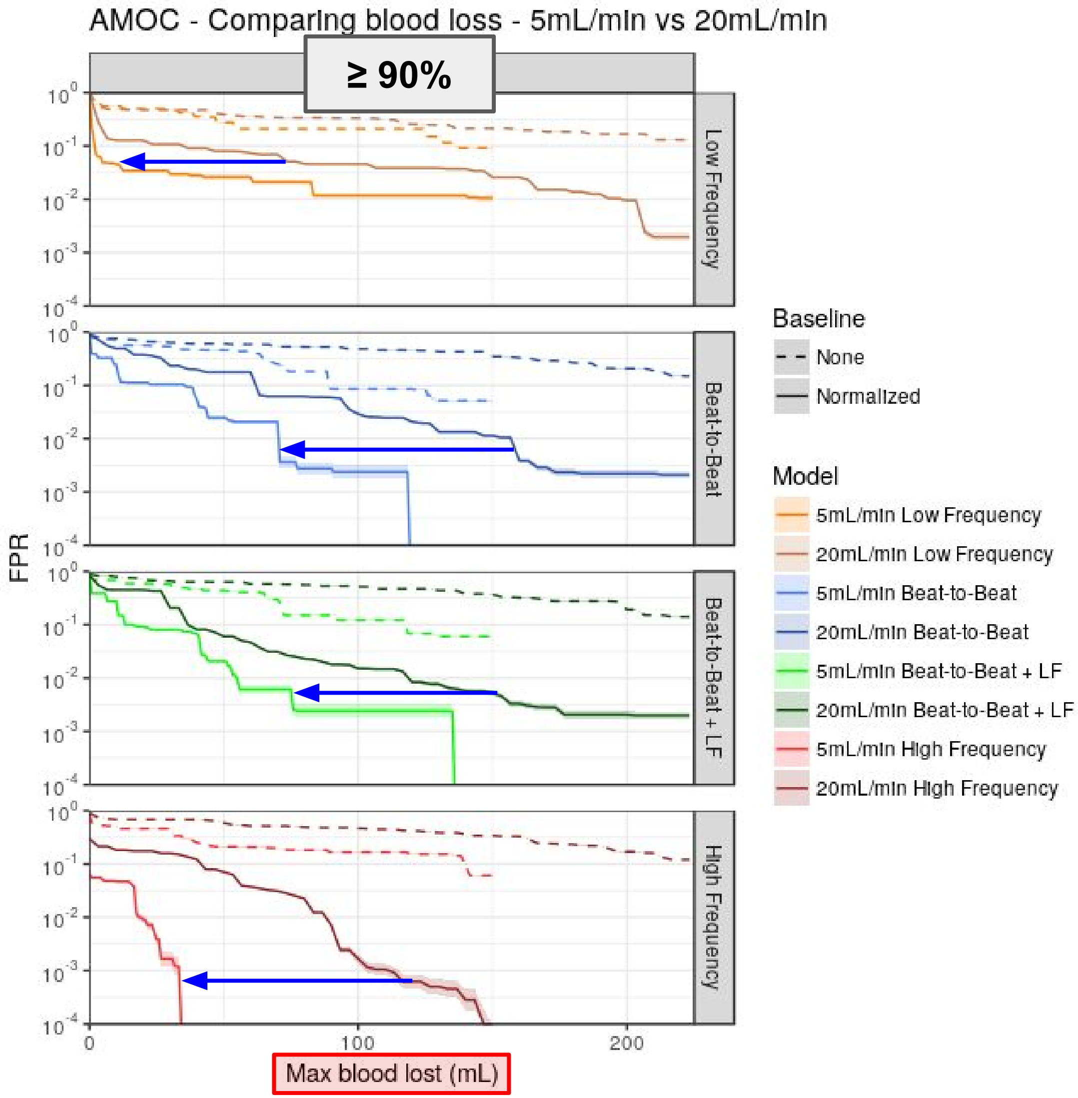
Detection Performance by Time Latency



- Lower granularity models detect more slowly for the slower bleeding pigs.
- But the highest granularity model detects them with the **same latency**.

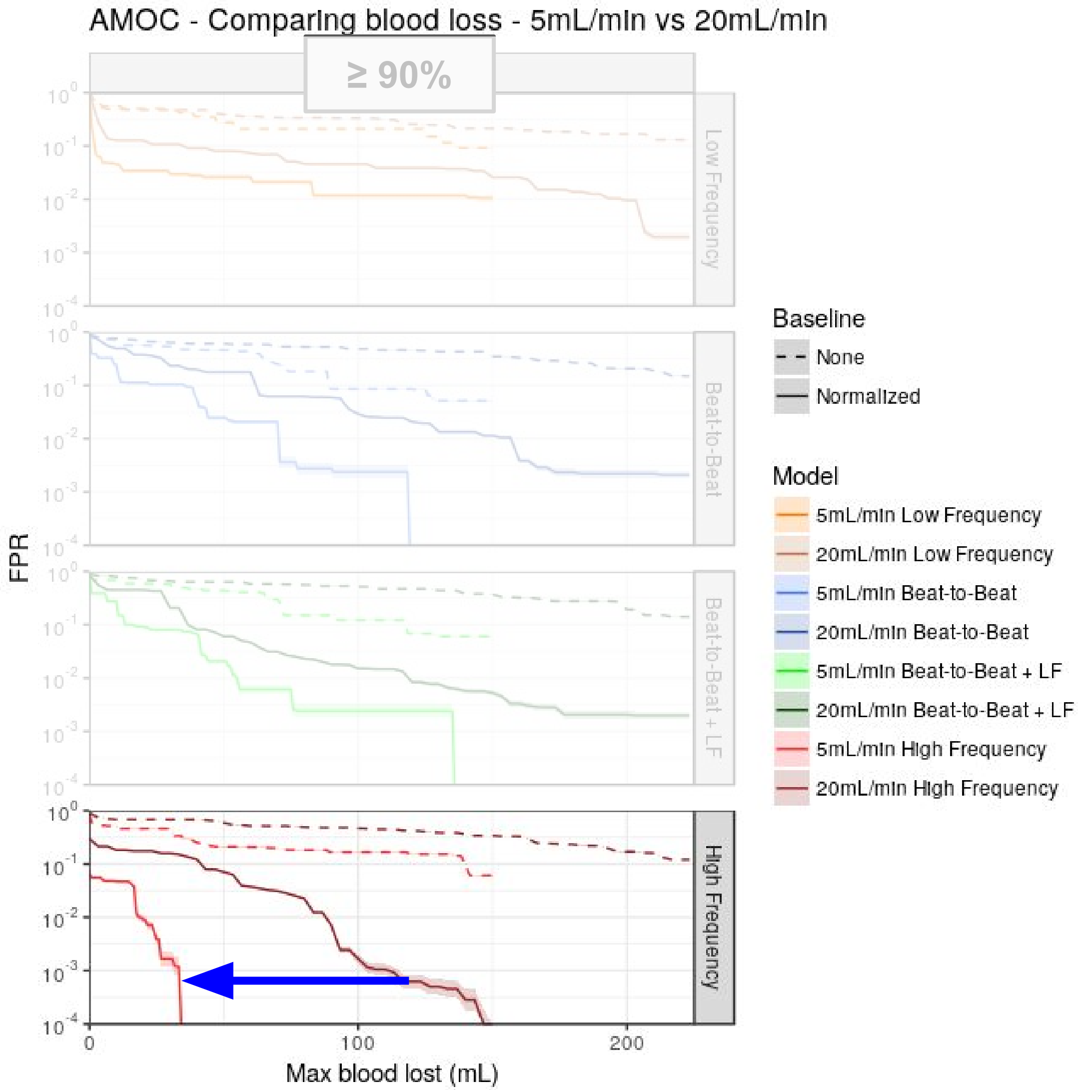
Detection Performance by Volume Lost

- Lower granularity models detect more slowly for the slower bleeding pigs.
- But the highest granularity model detects them with the **same latency**.
- Comparing by volume of blood loss reveals earlier detections in terms of volume of blood lost for the slower bleeding cohort.

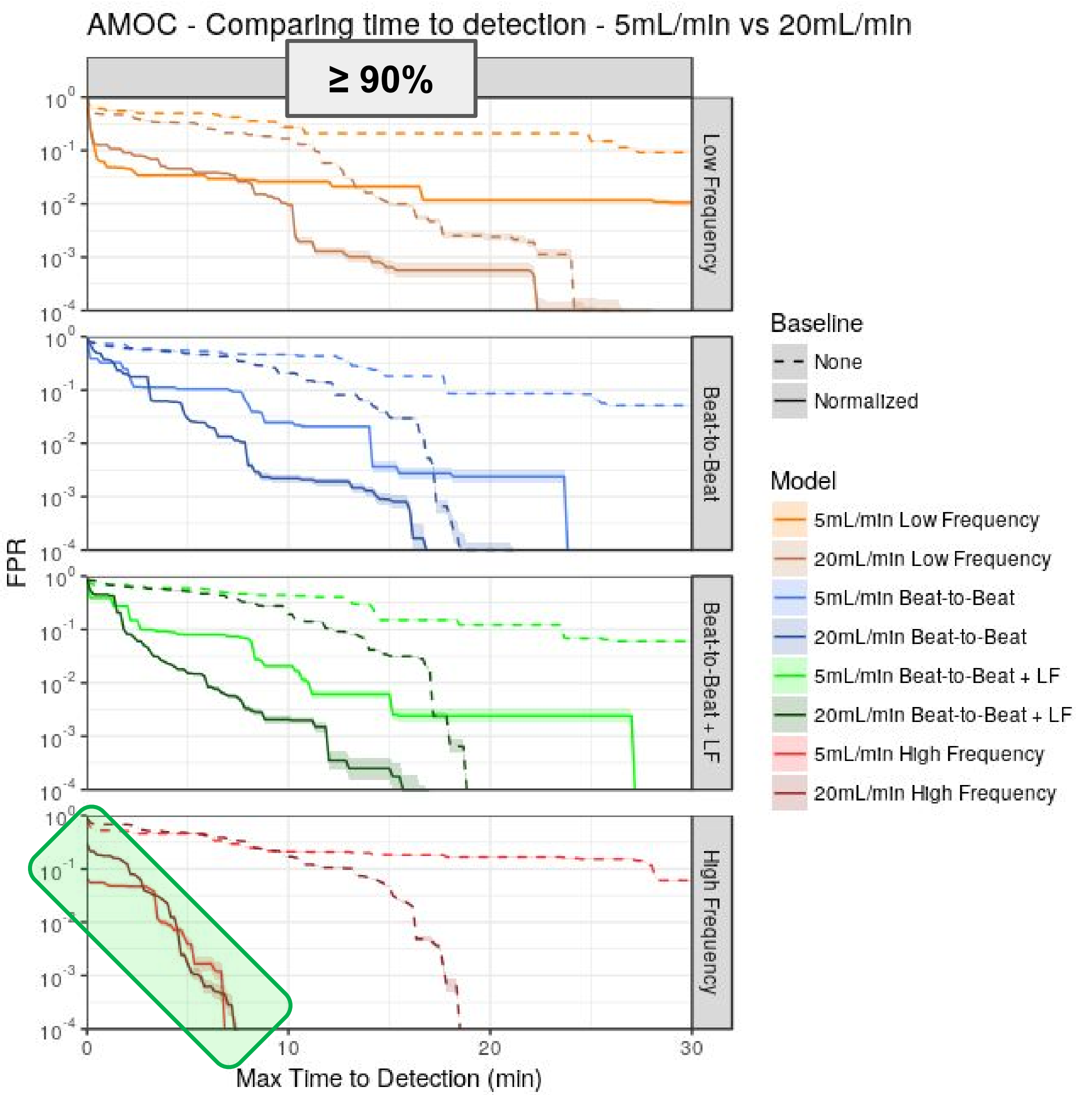


Detection Performance by Volume Lost

- Lower granularity models detect more slowly for the slower bleeding pigs.
- But the highest granularity model detects them with the **same latency**.
- Comparing by volume of blood loss reveals earlier detections in terms of volume of blood lost for the slower bleeding cohort.
- This is especially true in the case of the high frequency models.

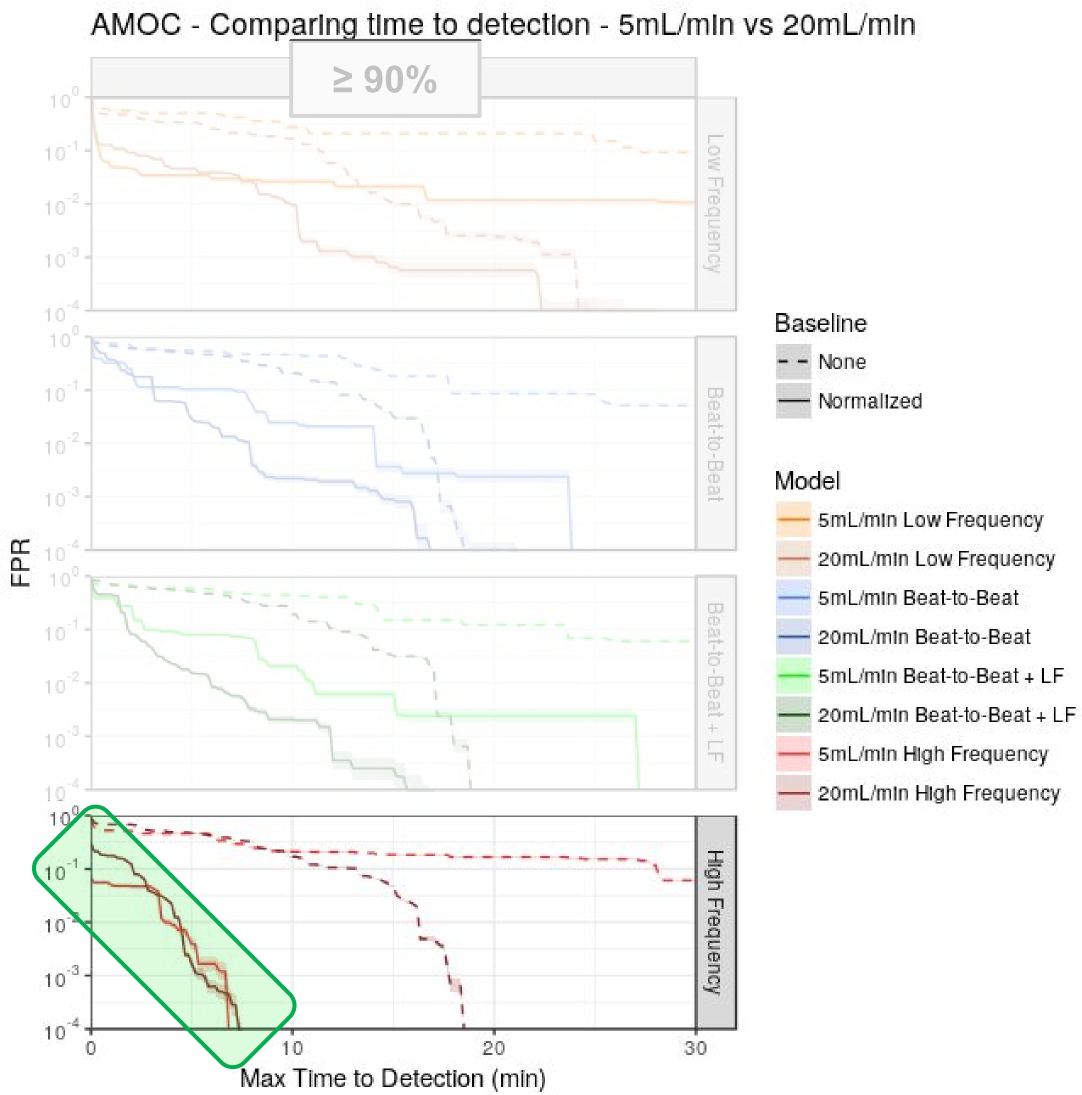


Detection Performance by Time and Volume Lost



- Lower granularity models detect more slowly for the slower bleeding pigs.
- But the highest granularity model detects them with the **same latency**.
- Comparing by volume of blood loss reveals earlier detections in terms of volume of blood lost for the slower bleeding cohort.
- This is especially true in the case of the high frequency models.
- Presence of a detectable response appears more dependent on some ***time delay*** from the onset of bleed rather than
  - ***volume*** of blood lost or
  - ***severity*** *(5 vs 20 mL/min)* of the bleeding.

Detection Performance by Time and Volume Lost



- Lower granularity models detect more slowly for the slower bleeding pigs.
- But the highest granularity model detects them with the **same latency**.
- Comparing by volume of blood loss reveals earlier detections in terms of volume of blood lost for the slower bleeding cohort.
- This is especially true in the case of the high frequency models.
- Presence of a detectable response appears more dependent on some *time delay* from the onset of bleed rather than
  - *volume* of blood lost or
  - *severity (5 vs 20 mL/min)* of the bleeding.
  - But seeing this requires ***denser data***.

Discussion

Clinical Implications:

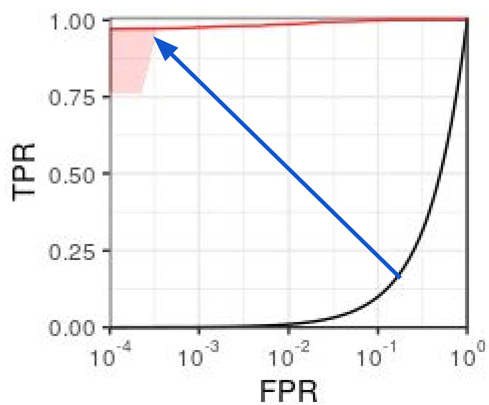
- Our results show that we can detect bleeding quickly at a low rate of false alarms, in particular when baseline data is available. (e.g. for patients prior to surgery, soldiers, astronauts…)
- Performance is improved when more granular data can be utilized, suggesting bedside monitoring equipment capable of capturing and processing higher density data can be beneficial.

What's next?

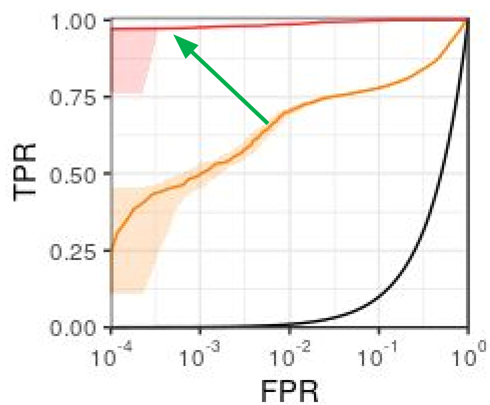
- Can models quantify the amount of blood lost?
- Can we eliminate the necessity of individualized baselines?
- How well can we do with non-invasive monitoring?
- Can we differentiate from other disease states? *(e.g. anaphylactic shock, septic shock…)*

The Main Take-Aways

Machine learning enables building powerful multi-variate models for bleeding detection.



Higher granularity data improves detection performance.



Knowledge of a personal baseline improves detection performance.