AMIA 2024
Annual Symposium

November 9-13 | San Francisco, CA

#AMIA2024

Personalized Uncertainty Quantification in Operating Room: Optimizing Surgical Time Estimation with Conformal Prediction

Session Title

Session Code: S10

Speaker

Sylvia Cheng, Ph.D. Candidate
University of California, Berkeley
Kaiser Permanente Division of Research
Connect with me: **sylvia-cheng@berkeley.edu**

SPEAKER DISCLOSURE

I have not had any relationships with ACCME-defined ineligible companies within the past 24 months.

In the past 24 months I have had the following relationship(s) with ACCME-defined ineligible companies:

NAME OF COMPANY	NATURE OF RELATIONSHIP

Operation Room Management in Perioperative Care

- In California, each minute spent in an operation room of an acute care hospital contributes to approximately $37 cost in 2014

Childers, C. P., & Maggard-Gibbons, M. (2018). Understanding costs of care in the operating room. JAMA surgery, 153(4), e176233-e176233.
Balch, C. M., & Shanafelt, T. (2010). Combating stress and burnout in surgical practice: a review. *Advances in surgery*, *44*(1), 29-47.
West, C. P., Huschka, M. M., Novotny, P. J., Sloan, J. A., Kolars, J. C., Habermann, T. M., & Shanafelt, T. D. (2006). Association of perceived medical errors with resident distress and empathy: a prospective longitudinal study. *Jama*, *296*(9), 1071-1078.

Operation Room Management in Perioperative Care

- In California, each minute spent in an operation room of an acute care hospital contributes to approximately $37 cost in 2014

- Distress and burnout are often reported by surgeons and medical staff working in suboptimal operating room environment, potentially leading to
  - worsened health status of care providers
  - lowered healthcare quality for patients
  - increased medical errors

Childers, C. P., & Maggard-Gibbons, M. (2018). Understanding costs of care in the operating room. JAMA surgery, 153(4), e176233-e176233.
Balch, C. M., & Shanafelt, T. (2010). Combating stress and burnout in surgical practice: a review. *Advances in surgery*, *44*(1), 29-47.
West, C. P., Huschka, M. M., Novotny, P. J., Sloan, J. A., Kolars, J. C., Habermann, T. M., & Shanafelt, T. D. (2006). Association of perceived medical errors with resident distress and empathy: a prospective longitudinal study. *Jama*, *296*(9), 1071-1078.

Challenges in Improving Surgical Time Estimation

- Rare procedures

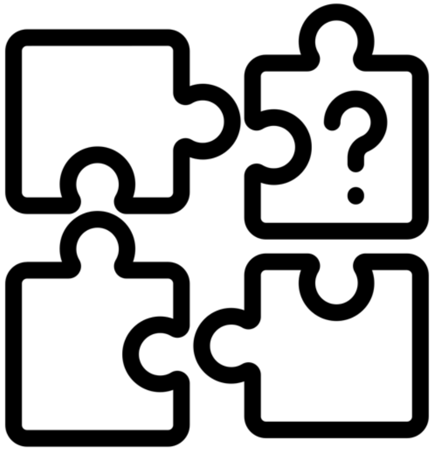
Challenges in Improving Surgical Time Estimation

- Rare procedures
- Uniqueness in patient features and health histories

Challenges in Improving Surgical Time Estimation

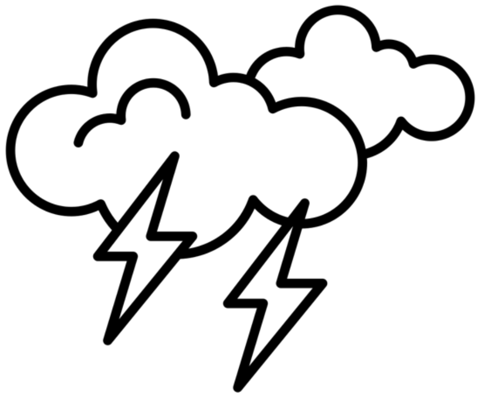
- Rare procedures
- Uniqueness in patient features and health histories
- Randomness not captured in the data






Challenges in Improving Surgical Time Estimation

- Rare procedures
- Uniqueness in patient features and health histories
- Randomness not captured in the data
  - missingness of information
  - Systematic reasons, e.g. design of the database
  - Unexpected changes when carrying out the case
  - Individual-level uncertainty, e.g. inpatient vs. outpatient, pre-existing health conditions…



Conformal prediction: capturing & quantifying the underlying uncertainty in an estimate

- Uncertainty quantification: delivers exact prediction intervals on the **individual** level for future observations

Conformal prediction: capturing & quantifying the underlying uncertainty in an estimate

- Uncertainty quantification: delivers exact prediction intervals on the **individual** level for future observations

- Reliability of the prediction: each predictive interval varies according to the **difficulty** of the individual task

Conformal prediction: capturing & quantifying the underlying uncertainty in an estimate

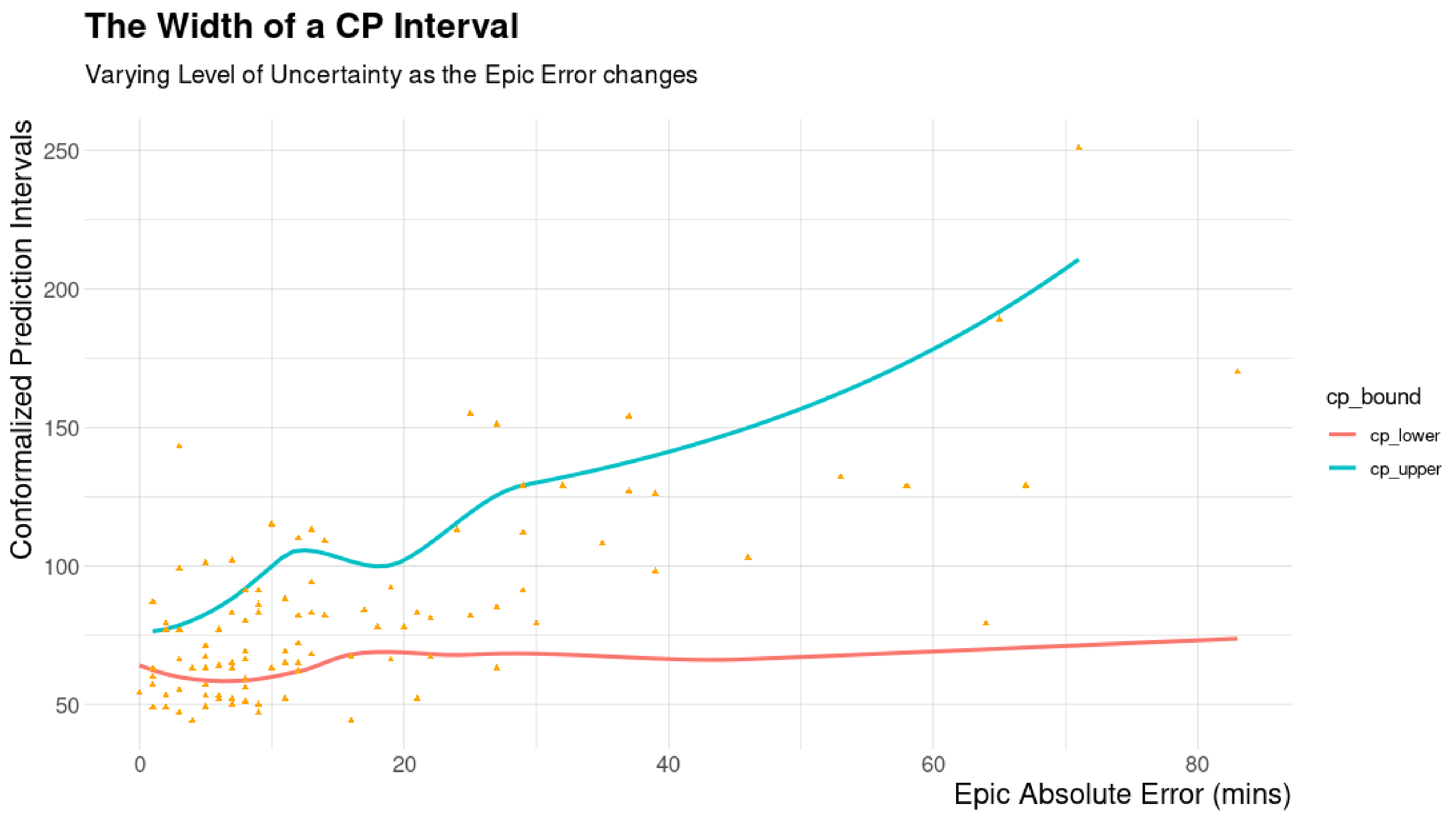
- Uncertainty quantification: delivers exact prediction intervals on the **individual** level for future observations

- Reliability of the prediction: each predictive interval varies according to the **difficulty** of the individual task

- Validity in finite, random samples
    - Guaranteed coverage of Y_i: pre-specified probability $1 - \alpha$ (error rate)

$$\mathscr{P}\{Y \in C(X) \,|\, X = x\} \geq 1 - \alpha$$

Guaranteed coverage on the individual level while reflecting the level of uncertainty

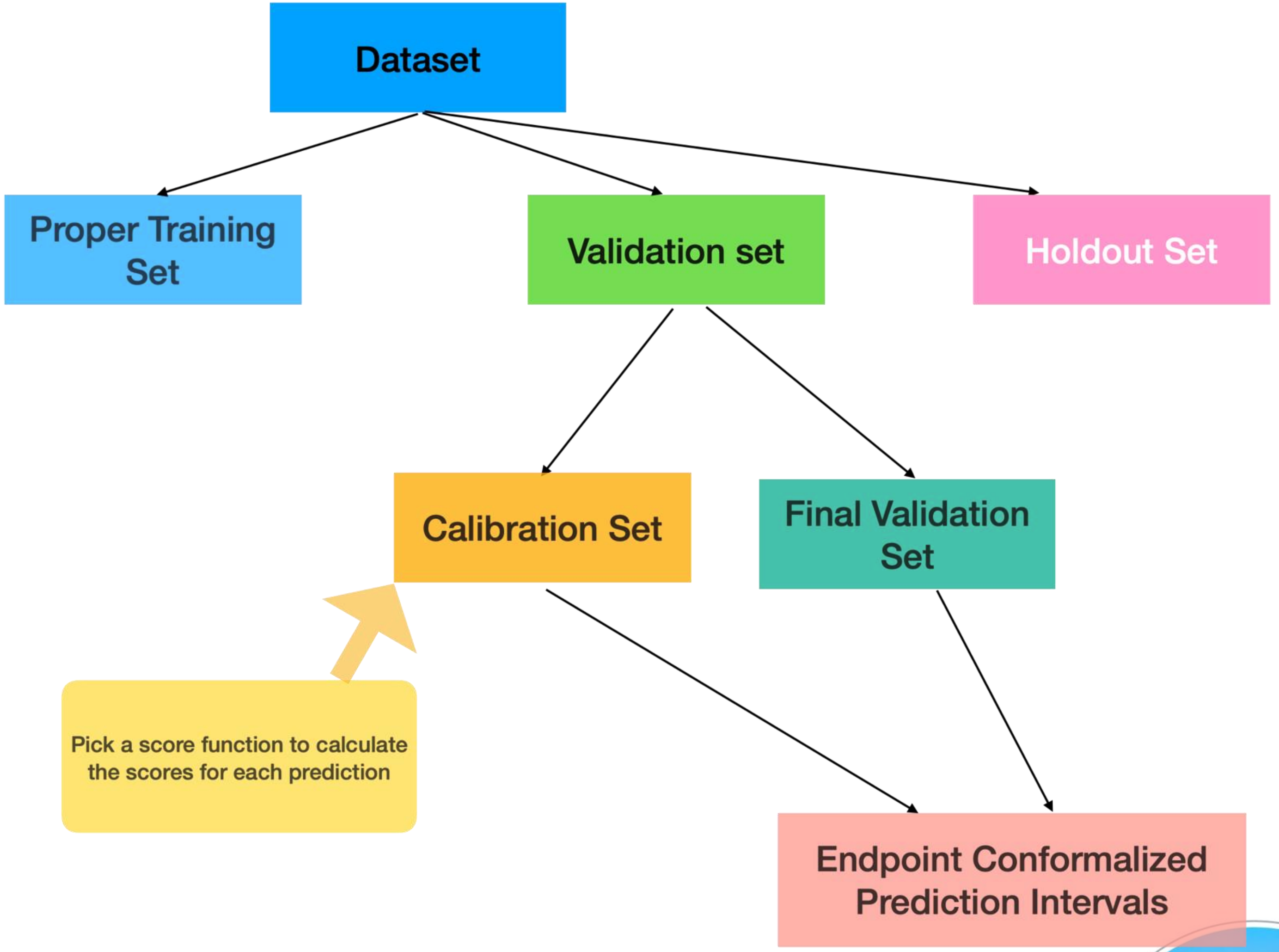


Model training

- 912,644 surgeries with dates from Jan 2018 to December 2022

- Data-splitting
    - training set 60% (n=547,942)
    - internal validation set 12% (n=109,306)
    - calibration set 3% (n=27,326)
    - holdout set 25% (n=228,070)

- A cross-validated gradient boosting model using Conformal Quantile Prediction with 435 variables (3,628 features after one-hot encoding)

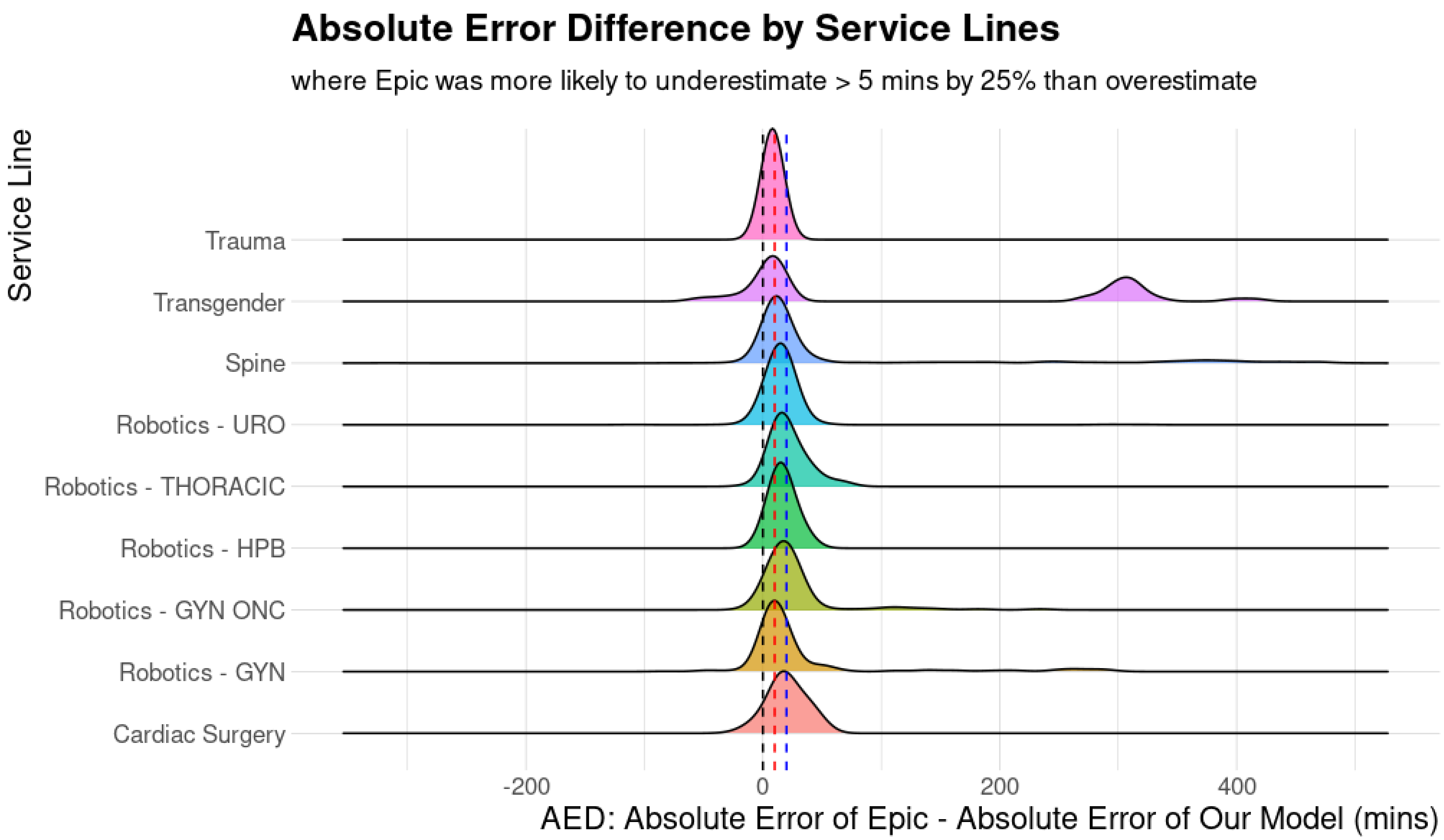


Model performance: across all surgical cases in the validation set

Models	RMSE (mins)	MAE (mins)
Our model with 1k features	37.49	22.38
Our model with 100 features	37.58	22.39
Surgeon Score	36.35	21.93
Epic Score	44.50	23.48

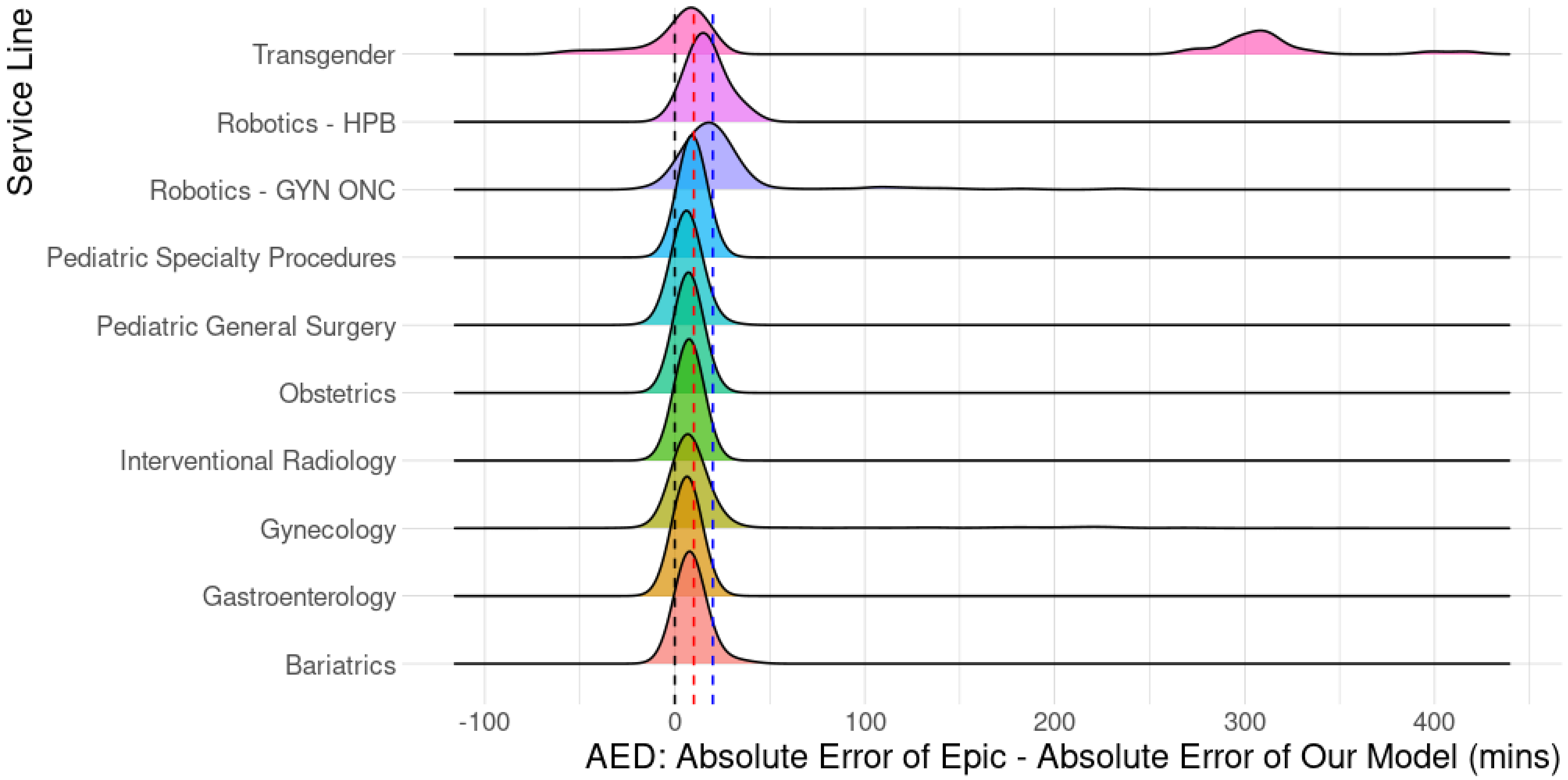
Model performance: in the holdout set

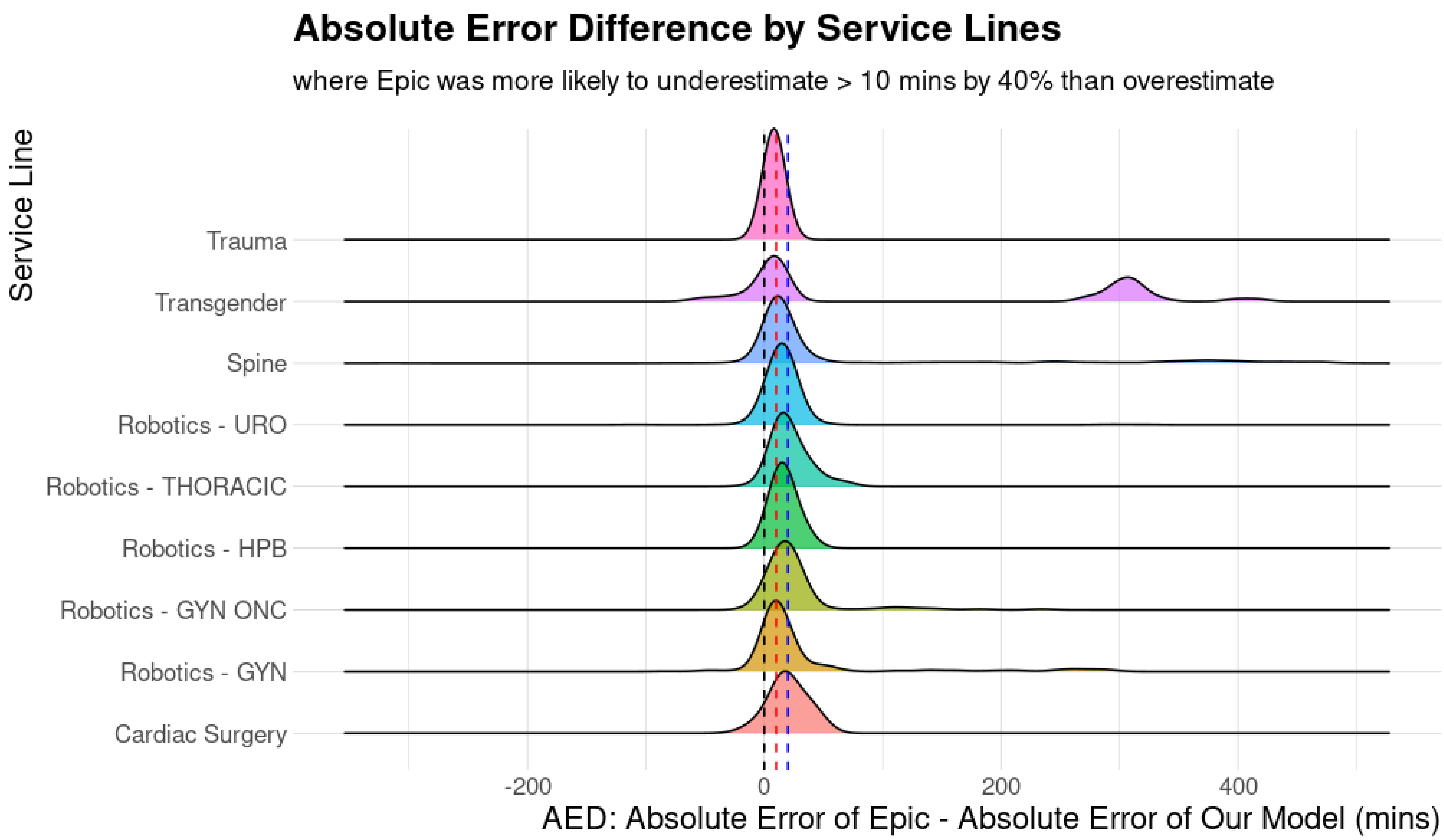
Models	RMSE (mins)	MAE (mins)
Our model with 100 features	37.49	22.39
Surgeon Score	36.34	21.96
Epic Score	44.18	23.41


Absolute Error Difference by Service Lines
where Epic was more likely to underestimate > 5 mins by 25% than overestimate
Service Line
Trauma
Transgender
Spine
Robotics - URO
Robotics - THORACIC
Robotics - HPB
Robotics - GYN ONC
Robotics - GYN
Cardiac Surgery
-200
0
200
400
AED: Absolute Error of Epic - Absolute Error of Our Model (mins)

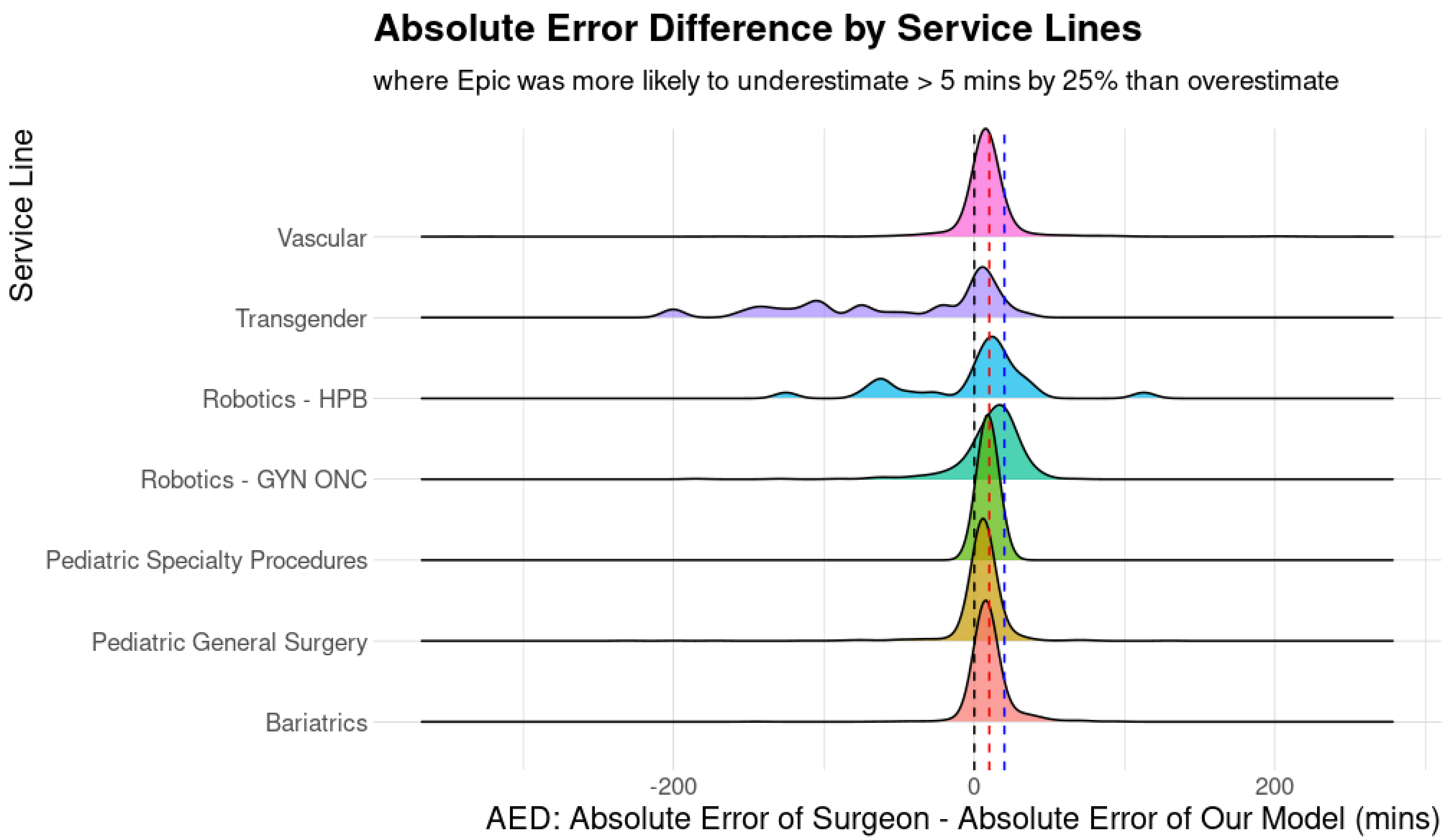
Absolute Error Difference by Service Lines

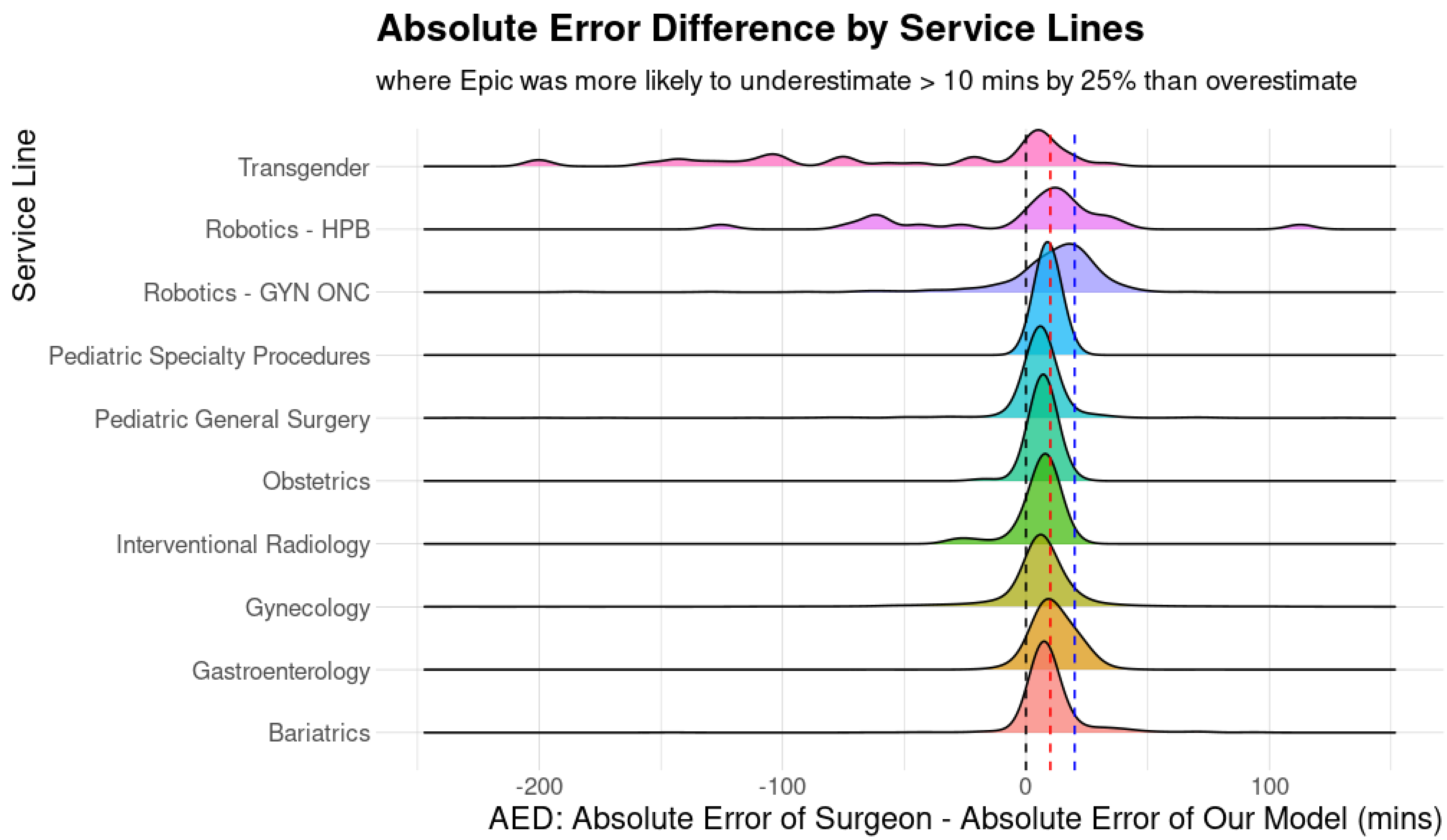
where Epic was more likely to underestimate > 10 mins by 25% than overestimate




Absolute Error Difference by Service Lines
where Epic was more likely to underestimate > 10 mins by 40% than overestimate
Service Line
Trauma
Transgender
Spine
Robotics - URO
Robotics - THORACIC
Robotics - HPB
Robotics - GYN ONC
Robotics - GYN
Cardiac Surgery
-200
0
200
400
AED: Absolute Error of Epic - Absolute Error of Our Model (mins)

How does our model compare to surgeons when Epic underestimates in certain service lines?

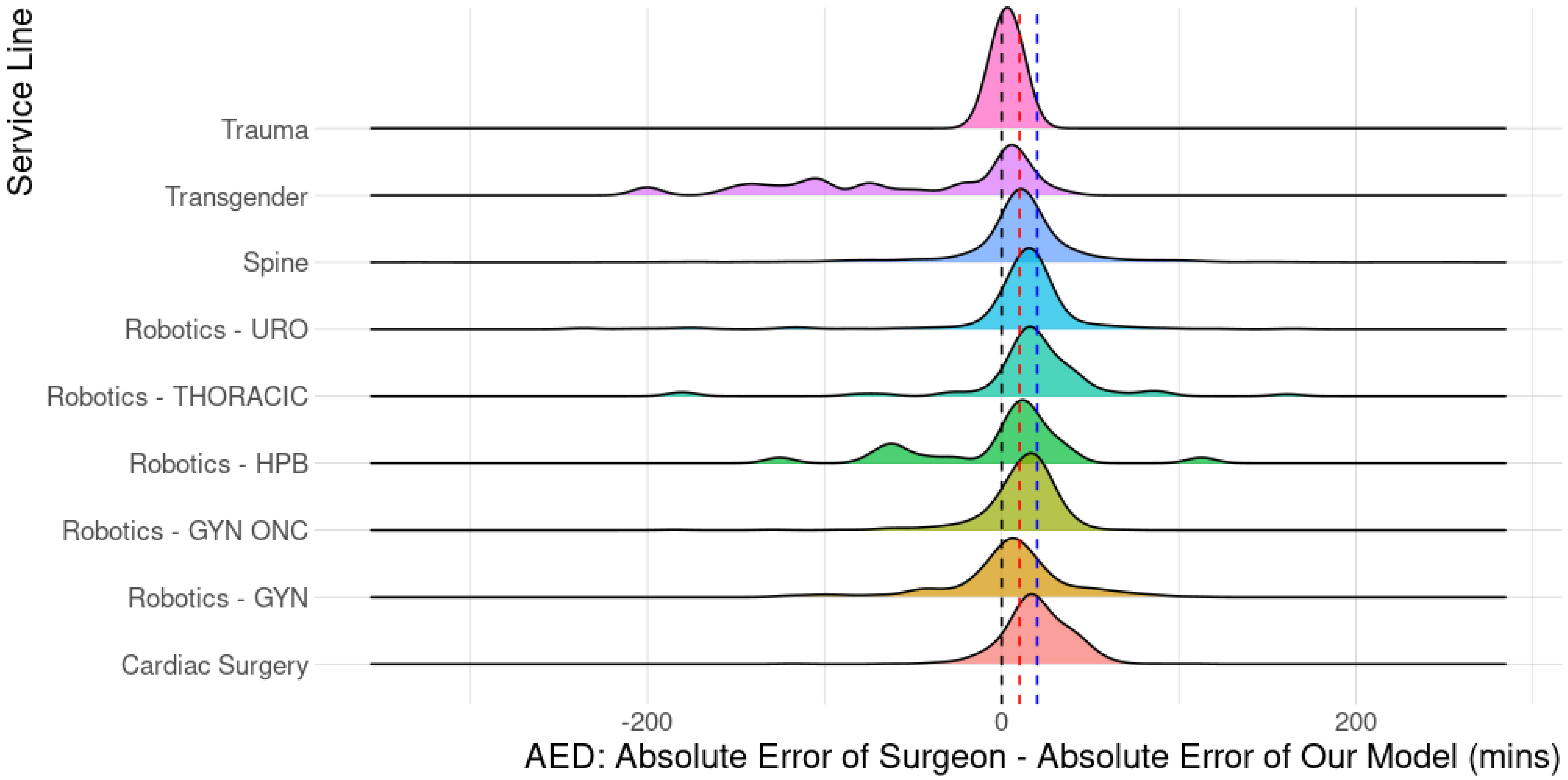

Absolute Error Difference by Service Lines
where Epic was more likely to underestimate > 5 mins by 25% than overestimate
Service Line
Vascular
Transgender
Robotics - HPB
Robotics - GYN ONC
Pediatric Specialty Procedures
Pediatric General Surgery
Bariatrics
-200
0
200
AED: Absolute Error of Surgeon - Absolute Error of Our Model (mins)



Absolute Error Difference by Service Lines
where Epic was more likely to underestimate > 10 mins by 25% than overestimate
Service Line
Transgender
Robotics - HPB
Robotics - GYN ONC
Pediatric Specialty Procedures
Pediatric General Surgery
Obstetrics
Interventional Radiology
Gynecology
Gastroenterology
Bariatrics
-200
-100
0
100
AED: Absolute Error of Surgeon - Absolute Error of Our Model (mins)

Absolute Error Difference by Service Lines

where Epic was more likely to underestimate > 10 mins by 40% than overestimate



Service Lines in the holdout set

Table. Service Lines in the Holdout Set

ServiceLine	count	mean_Epic_AbsError	mean_cp_width
Psychiatry	966	3.76	8.99
Interventional Pain Management	3,123	6.32	11.99
Dermatology	19	8.95	15.03
Ophthalmology	22,493	9.04	16.46
Interventional Radiology	313	14.70	29.57
Gastroenterology	5,555	15.63	29.62
Obstetrics	168	17.74	31.11
Pediatric Specialty Procedures	25	20.76	31.29
Bariatrics	2,395	15.44	31.96
Trauma	4	68.75	32.20
Podiatry	12,426	21.61	38.87
Orthopedics	49,700	21.56	39.18
Urology	19,007	20.57	39.26
Pediatric General Surgery	2,050	23.19	39.54
Retinal Surgery	301	23.34	40.00
General Surgery	49,698	23.49	42.54
Cardiology	1,955	17.80	43.03
Pulmonary	558	27.44	44.01
Head and Neck	16,047	27.79	44.40
Gynecology	13,409	29.64	44.45
Dentistry	620	23.57	46.39
Maxillofacial	1,568	27.97	50.70
Plastics	5,891	31.48	54.78
Vascular	5,378	30.01	55.36
Robotics - TORS	36	26.92	56.00
Robotics - GEN SUR	798	42.30	63.28
Robotics - HPB	64	59.00	65.50
Robotics - GYN ONC	772	43.30	66.91
Thoracic	1,228	42.91	66.97
Pediatric Cardiac	1	53.00	71.87
Robotics - URO	1,378	38.60	74.82
Spine	3,322	62.10	81.81
Neurosurgery	4,373	45.24	83.27
Robotics - GYN	880	59.81	85.14
Robotics - THORACIC	194	56.87	89.08
Cardiac Surgery	941	52.55	106.95
Robotics - COLO	339	62.76	109.89
Transgender	75	163.27	138.94

MANY THANKS TO MY COLLABORATORS

In alphabetical order:

Kaiser Permanente Division of Research
Bradley Cohn, M.D.
Brian Lawson, Ph.D.
Patricia Kipnis, Ph.D.
Vincent Liu, M.D., M.S.

University of California, Berkeley, Division of Biostatistics
Alejandro Schuler, Ph.D.

LEARNING OUTCOMES

At the conclusion of this session attendees will understand how the level of uncertainty in an estimate for surgical duration can be captured by the method we used in the model to assist decision making.

LEARNING OUTCOMES EVALUATION QUESTION

A conformalized prediction interval

a. Is equal to a confidence interval
b. Doesn’t have guaranteed coverage of truth
c. Can’t have its coverage customized based on the user’s need
d. Can be reflective of the level of uncertainty in the model for a certain data point