(Q1-Q4) The table below shows the results of heart failure screening from 100 adult patients undergoing major surgery. The HF expert and the ML algorithm determined whether heart failure was present or absent prior to surgery. Please respond to the following questions.

		HF Expert Review		
		HF	No HF	Total
ML Detection	HF	39	40	79
	No HF	2	19	21
	Total	41	59	100

Note: True Positive Rate = Sensitivity; False Positive Rate = 1 – Specificity

Question 1. What is the **True Positive Rate (TPR)** from the table above?

A. $39/79$
B. $2/21$
C. $39/41$
D. $40/59$
E. $39/100$

Question 2. What is the **False Positive Rate (FPR)** from the table above?

A. $1 - 39/79$
B. $2/21$
C. $39/79$
D. $1 - 19/59$
E. $1 - 39/100$

Question 3. The ML algorithm detected 79 positive cases, 39 true positive and 40 false positive cases. Because of the high number of false positives, this ML algorithm triggers many false alarms, causing preoperative clinicians to be overburdened. In this situation, how would you like to modify its **threshold** (cut-off probability) to reduce the number of false positives?

A. Increase the threshold
B. Decrease the threshold

Question 4. How would TPR (sensitivity) and FPR (1 – specificity) respond to the threshold change based on your response in Question 3?

A. TPR increases, FPR increases
B. TPR decreases, FPR decreases
C. TPR increases, FPR decreases
D. TPR decreases, FPR increases
E. No changes in TPR and FPR

(Q5) In this study, Sensitivity (TPR), Specificity (1-FPR), and the AUC plot are available to assist your clinical decision. The TPR and FPR follow the line of the curve in the AUC plot based on various thresholds.

Model Performance:

AUROC = 90%, **Threshold** = 38%, **Sensitivity** = 82%, **Specificity** = 82%

In ML Reference, *ML recommendation* (HF vs No HF) and *HF probability* are available to help your clinical judgment in individual cases. In addition, the overall model performance shown above indicates that the model correctly detects HF and No HF by more than 80%.

We want you to effectively utilize ML-generated information when reviewing cases. Especially, when your clinical judgment and ML recommendation conflict, we encourage you to review the absolute difference (nonnegative number) between *"HF probability"* and the *"threshold."* This absolute difference shows the likelihood that the ML algorithm correctly detects HF or No HF.

Question 5. Using the *ML recommendation* and *HF probability* in the table below and the model performance above, calculate the absolute difference between *HF probability* and the *threshold*, then choose the **most** appropriate likelihood that ML correctly detects HF or No HF.

The likelihood of correctly detecting HF or No HF:
Choice 1. Very likely (>= 90%)
Choice 2. Likely (60-89%)
Choice 3. Neither likely nor unlikely (40-59%)

	HF Recommendation	HF Probability	Difference from the threshold	Likelihood of correct detection
Case #1.	HF	95%		
Case #2.	HF	45%		
Case #3.	No HF	10%		

Please keep in mind that the likelihood in this question is for your conceptual understanding to support your clinical reasoning. The primary source of the final HF clinical decision should be your clinical reasoning.

Hint:

- The further away from the threshold, the greater the likelihood of correctly detecting HF or No HF. For example, 97% or 5% of HF probability is far from 38% of the threshold.