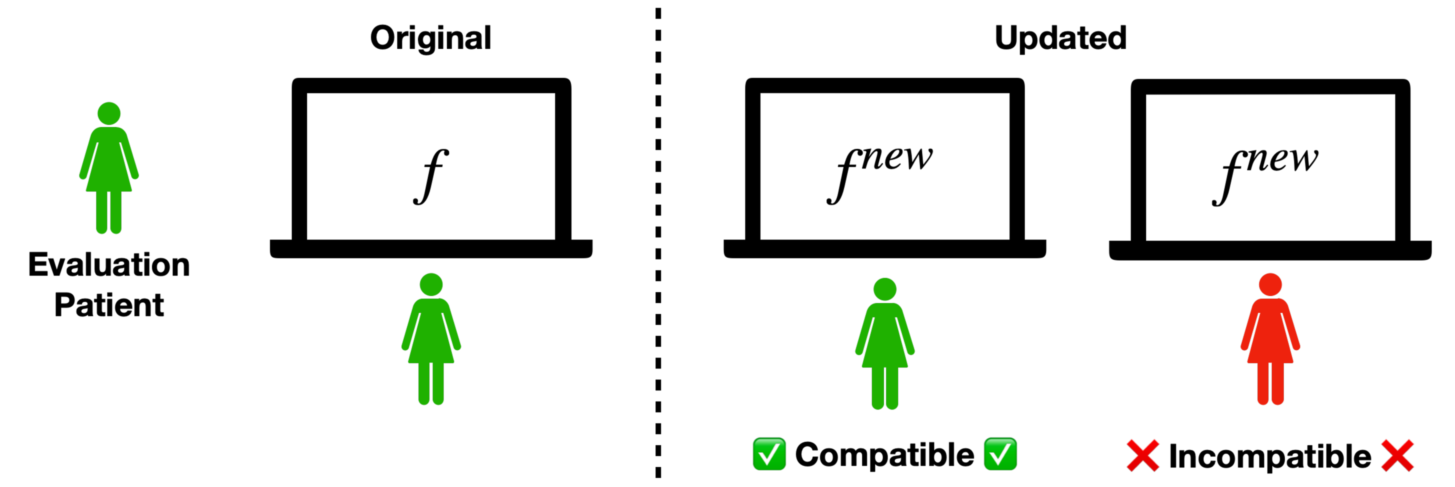
Updating Clinical Risk Stratification Models Using Rank-Based Compatibility

Erkin Ötleş, Brian T. Denton, Jenna Wiens
Medical Scientist Training Program, Industrial & Operations Engineering, Computer Science & Engineering
eotles@umich.edu

UNIVERSITY OF MICHIGAN

Updating clinical risk models pose challenges when updated models don't meet user expectations.

Compatibility quantifies how an updated model **continues the correct behavior** exhibited by an original model



Existing measures operate by comparing labels

$$C^{BT}(f^o, f^u) = \frac{\#\,patients\ both\ models\ label\ correctly}{\#\,patients\ original\ model\ labels\ correctly}$$

Gap: Limited use in updating healthcare risk stratification models; incongruous with multiple thresholds, resource-based usage, and AUROC

How can we measure and optimize compatibility in a way that is better suited for updating risk stratification models?

We propose a **rank-based** compatibility measure based on the concordance of **risk estimate pairs**

Problem Setup

Risk stratification models, $f(\cdot)$, map features, $\boldsymbol{x}_i$, to risk estimates, $\hat{p}_i$, for each patient, i

Goal: we seek to assess the compatibility between an original model $f^o(\cdot)$ and an updated model $f^u(\cdot)$

The set of patients, I, can be split based on their labels: 0-labeled, I^0, and 1-labeled, I^1

A *patient-pair*, are two patients i and j, that do not share the same label $i \in I^0$ and $j \in I^1$

Intuition: we develop a compatibility measure inheriting AUROC's notion of correct risk estimate ordering.

Rank-based Compatibility

We propose

$$C^R(f^o, f^u) = \frac{\sum_{i \in I^0} \sum_{j \in I^1} \mathbf{1}(\hat{p}_i^o < \hat{p}_j^o) \cdot \mathbf{1}(\hat{p}_i^u < \hat{p}_j^u)}{\sum_{i \in I^0} \sum_{j \in I^1} \mathbf{1}(\hat{p}_i^o < \hat{p}_j^o)}$$

the proportion of patient-pairs correctly ranked by both models normalized by the original model's AUROC

Optimizing for rank-based compatibility

$$\widetilde{L^R}(f^o, f^u) = 1 - \frac{\sum_{i \in I^0} \sum_{j \in I^1} \sigma(\hat{p}_j^o - \hat{p}_i^o) \cdot \sigma(\hat{p}_j^u - \hat{p}_i^u)}{\sum_{i \in I^0} \sum_{j \in I^1} \sigma(\hat{p}_j^o - \hat{p}_i^o)}$$

Approximate rank-based incompatibility loss, $\widetilde{L^R}$, is a loss function based on approximation of C^R and uses the *ranking sigmoid function:*

$$\sigma(\hat{d}_{ji}) = \frac{1}{1 + \exp(-s \cdot \hat{d}_{ji})}$$

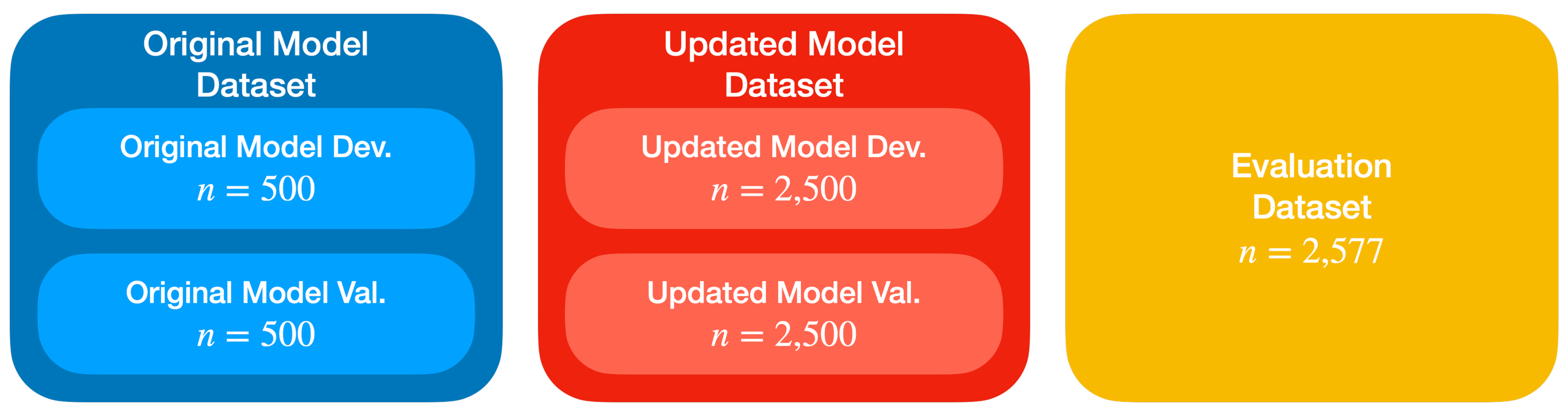
This can be weighted against binary cross entropy loss, L^{BCE}

$$\alpha L^{BCE}(f^u) + (1 - \alpha)\widetilde{L^R}(f^o, f^u)$$

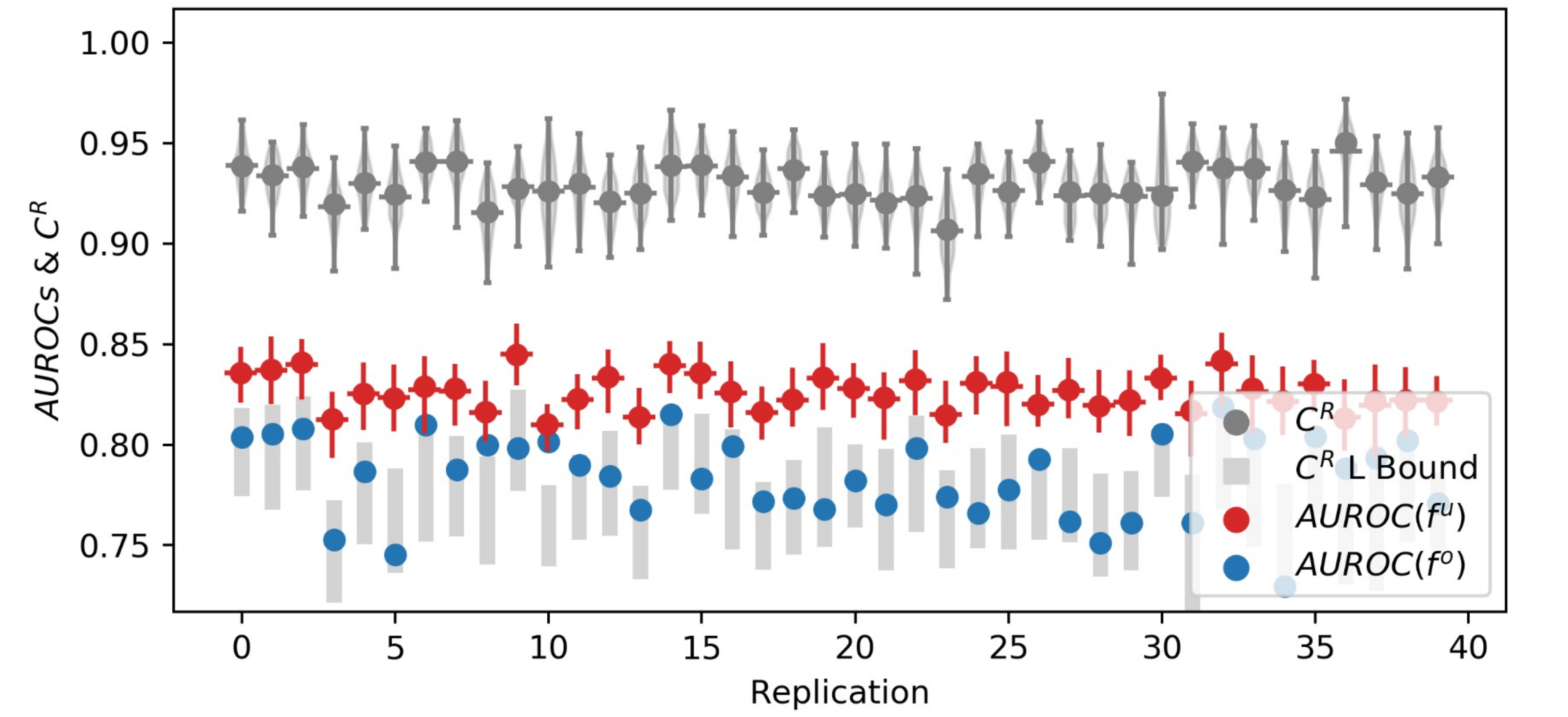
Does C^R come for free?

Q1: *What is the empirical distribution of C^R achieved using standard model updates?*
Q2: *Compared to standard model update generation and selection approaches, can we use L^R to generate updates with better C^R?*

Data: MIMIC-III
Task: predict in-hospital mortality based on the first 48 hours of ICU stay



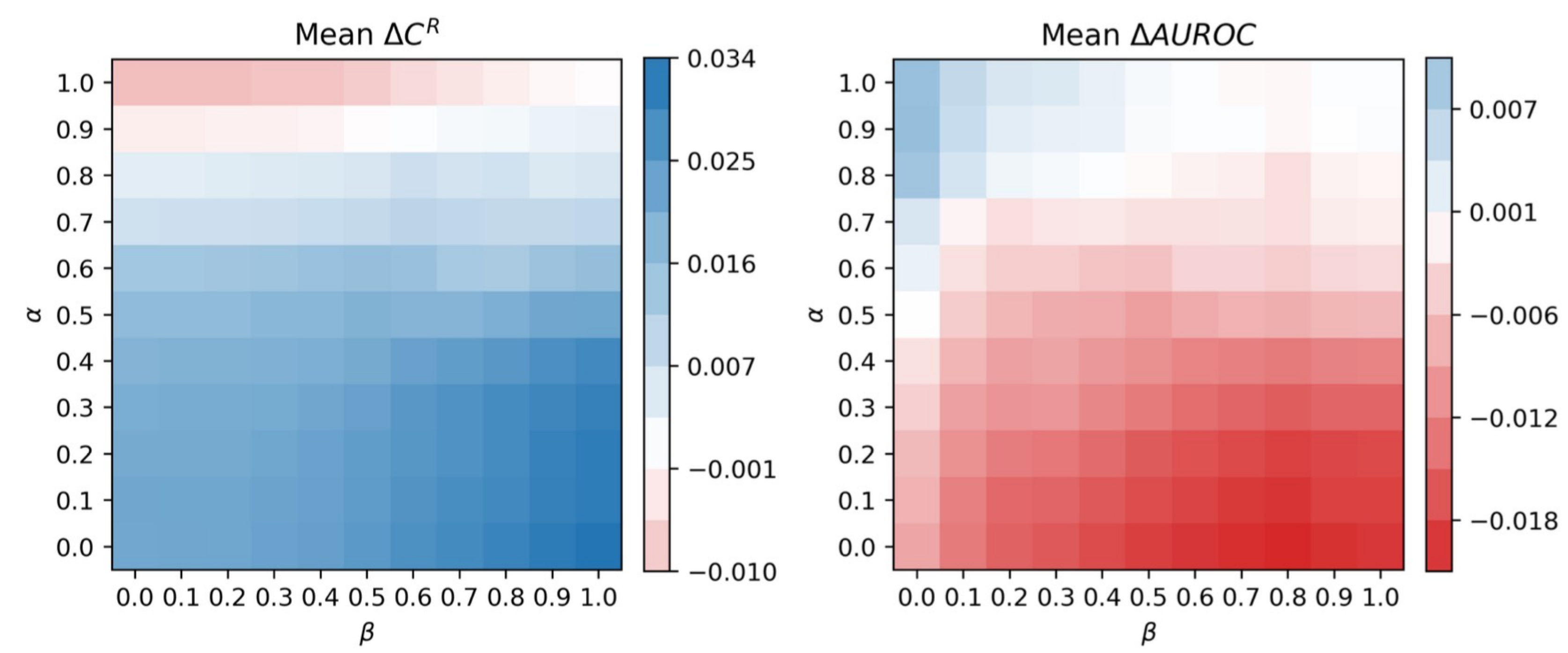
Results



Q1: *Model developers may be limited when selecting updated models that maximize C^R when using standard update generation procedures.*

For **Q2** we assessed the difference in performance and compatibility between models trained using only L^{BCE} and those trained with a weighted combination of L^{BCE} and L^R. This selection was done using:

$$\beta AUROC(f^u) + (1 - \beta)C^R(f^o, f^u)$$



Q2: *Incorporating C^R into the objective function generates model updates with larger C^R than obtained through standard procedures*

High rank-based compatibility is not guaranteed but can be **achieved through optimization**, which can yield updated models that better meet user expectations, **promoting clinician-model team performance.**

Funding: T32GM007863, Sloan Foundation