Evaluating Forecasts in the Context of Public Health Decision-Making

Aaron Gerding, **Evan L. Ray**

Royal Society Satellite Meeting on
Forecasting Infectious Disease Incidence
15 March 2023

UMassAmherst
School of Public Health
& Health Sciences
Biostatistics and Epidemiology

Connecting forecast uses and targets

Use	Targets
Planning expansions to hospital bed or ICU capacity	Peak (all-cause) hospitalizations in a given location
Allocation of limited medical supplies (e.g. ventilators, oxygen)	Demand for resources in multiple locations (e.g. [severe] hospitalizations)
Site selection for vaccine trials	Case counts across multiple locations
Situational awareness, public communications	Quantities of relevance to the public; cases, hospitalizations, deaths, …

Connecting forecast uses and targets

Use	Targets
Planning expansions to hospital bed or ICU capacity	Peak (all-cause) hospitalizations in a given location
Allocation of limited medical supplies (e.g. ventilators, oxygen)	Demand for resources in multiple locations (e.g. [severe] hospitalizations)
Site selection for vaccine trials	Case counts across multiple locations
Situational awareness, public communications	Quantities of relevance to the public; cases, hospitalizations, deaths, …

- **Question:** How should we quantify the value of forecasts in the context of each of these uses of the forecasts?

Our scope: actions with quantifiable loss

Use	
Planning expansions to hospital bed or ICU capacity	Potentially quantifiable
Allocation of limited medical supplies (e.g. ventilators, oxygen)	
Site selection for vaccine trials	
Situational awareness, public communications	Not easily quantifiable

- We focus on settings where we could plausibly quantify the loss or utility associated with a particular decision

3-step recipe (example: resource allocation)

1. Specify a way to measure the loss (or utility) associated with taking an action $\mathbf{x}$ when the outcome $\mathbf{y}$ is realized

3-step recipe (example: resource allocation)

1. Specify a way to measure the loss (or utility) associated with taking an action $\mathbf{x}$ when the outcome $\mathbf{y}$ is realized
   - $\mathbf{y} = (y_1, \ldots, y_n)$: number of hospitalizations in each location
   - $\mathbf{x} = (x_1, \ldots, x_n)$: resources allocated to each location, with $\sum x_i \leq K$
   - Loss is the amount of unmet need
   - Example: two locations, $K = 30$ units of resources
     - allocate $\mathbf{x} = (10, 20)$ units of resources to locations 1 and 2
     - eventually the value $\mathbf{y} = (15, 18)$ is observed
     - unmet need is $s(\mathbf{x}, \mathbf{y}) = (15 - 10) + 0 = 5$

3-step recipe (example: resource allocation)

1. Specify a way to measure the loss (or utility) associated with taking an action $\mathbf{x}$ when the outcome $\mathbf{y}$ is realized
   - $\mathbf{y} = (y_1, \ldots, y_n)$: number of hospitalizations in each location
   - $\mathbf{x} = (x_1, \ldots, x_n)$: resources allocated to each location, with $\sum x_i \le K$
   - Loss is the amount of unmet need
   - Example: two locations, $K = 30$ units of resources
     - allocate $\mathbf{x} = (10, 20)$ units of resources to locations 1 and 2
     - eventually the value $\mathbf{y} = (15, 18)$ is observed
     - unmet need is $s(\mathbf{x}, \mathbf{y}) = (15 - 10) + 0 = 5$

2. Given a probabilistic forecast F for Y, determine the action $\mathbf{x}^F$ that minimizes expected loss under the distribution F

3-step recipe (example: resource allocation)

1. Specify a way to measure the loss (or utility) associated with taking an action $\mathbf{x}$ when the outcome $\mathbf{y}$ is realized
   - $\mathbf{y} = (y_1, \ldots, y_n)$: number of hospitalizations in each location
   - $\mathbf{x} = (x_1, \ldots, x_n)$: resources allocated to each location, with $\sum x_i \leq K$
   - Loss is the amount of unmet need
   - Example: two locations, $K = 30$ units of resources
     - allocate $\mathbf{x} = (10, 20)$ units of resources to locations 1 and 2
     - eventually the value $\mathbf{y} = (15, 18)$ is observed
     - unmet need is $s(\mathbf{x}, \mathbf{y}) = (15 - 10) + 0 = 5$

2. Given a probabilistic forecast F for Y, determine the action $\mathbf{x}^F$ that minimizes expected loss under the distribution F
   - divide available resources across locations so that according to F, the expected benefit of 1 additional unit is the same everywhere

3-step recipe (example: resource allocation)

1. Specify a way to measure the loss (or utility) associated with taking an action $\mathbf{x}$ when the outcome $\mathbf{y}$ is realized
   - $\mathbf{y} = (y_1, \ldots, y_n)$: number of hospitalizations in each location
   - $\mathbf{x} = (x_1, \ldots, x_n)$: resources allocated to each location, with $\sum x_i \leq K$
   - Loss is the amount of unmet need
   - Example: two locations, $K = 30$ units of resources
     - allocate $\mathbf{x} = (10, 20)$ units of resources to locations 1 and 2
     - eventually the value $\mathbf{y} = (15, 18)$ is observed
     - unmet need is $s(\mathbf{x}, \mathbf{y}) = (15 - 10) + 0 = 5$

2. Given a probabilistic forecast F for Y, determine the action $\mathbf{x}^F$ that minimizes expected loss under the distribution F
   - divide available resources across locations so that according to F, the expected benefit of 1 additional unit is the same everywhere

3. Once the outcome $\mathbf{y}$ is observed, score the forecast based on the loss incurred by the action $\mathbf{x}^F$ in relation to the outcome $\mathbf{y}$

3-step recipe (example: resource allocation)

1. Specify a way to measure the loss (or utility) associated with taking an action $\mathbf{x}$ when the outcome $\mathbf{y}$ is realized
   - $\mathbf{y} = (y_1, \ldots, y_n)$: number of hospitalizations in each location
   - $\mathbf{x} = (x_1, \ldots, x_n)$: resources allocated to each location, with $\sum x_i \leq K$
   - Loss is the amount of unmet need
   - Example: two locations, $K = 30$ units of resources
     - allocate $\mathbf{x} = (10, 20)$ units of resources to locations 1 and 2
     - eventually the value $\mathbf{y} = (15, 18)$ is observed
     - unmet need is $s(\mathbf{x}, \mathbf{y}) = (15 - 10) + 0 = 5$

2. Given a probabilistic forecast F for Y, determine the action $\mathbf{x}^F$ that minimizes expected loss under the distribution F
   - divide available resources across locations so that according to F, the expected benefit of 1 additional unit is the same everywhere

3. Once the outcome $\mathbf{y}$ is observed, score the forecast based on the loss incurred by the action $\mathbf{x}^F$ in relation to the outcome $\mathbf{y}$
   - How much unmet need would have resulted with that allocation?

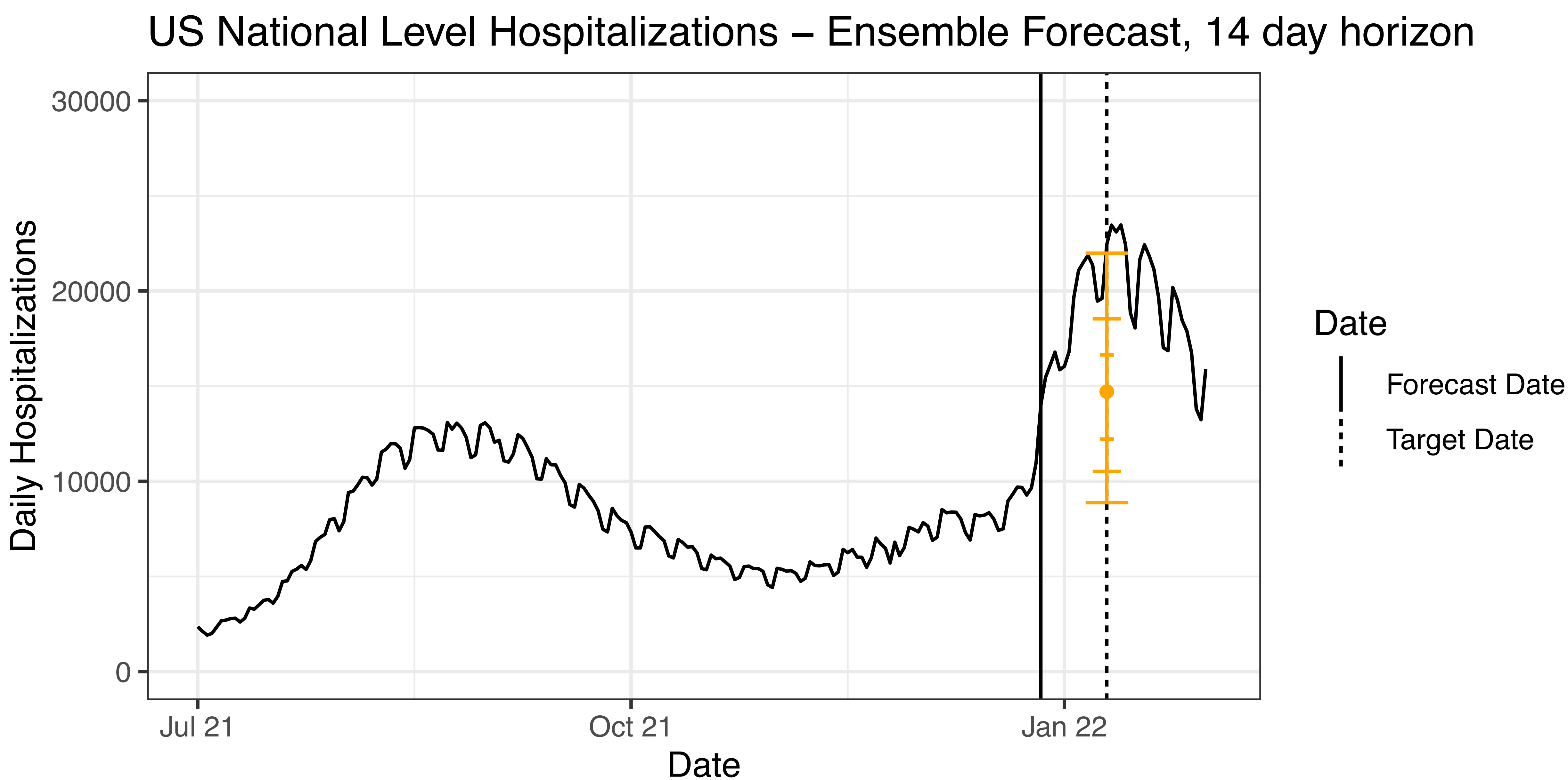
Note: the recipe is general

1. Define a loss function $s(\mathbf{x}, \mathbf{y})$
2. Get the Bayes act $\mathbf{x}^F$ for a probabilistic forecast F
3. Score the forecast via $s(\mathbf{x}^F, \mathbf{y})$

- This process is standard procedure in decision theory
- Under some technical conditions, forecast scoring rules obtained from this recipe are proper
  - Gneiting, T. and Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, *102*(477), pp.359-378.

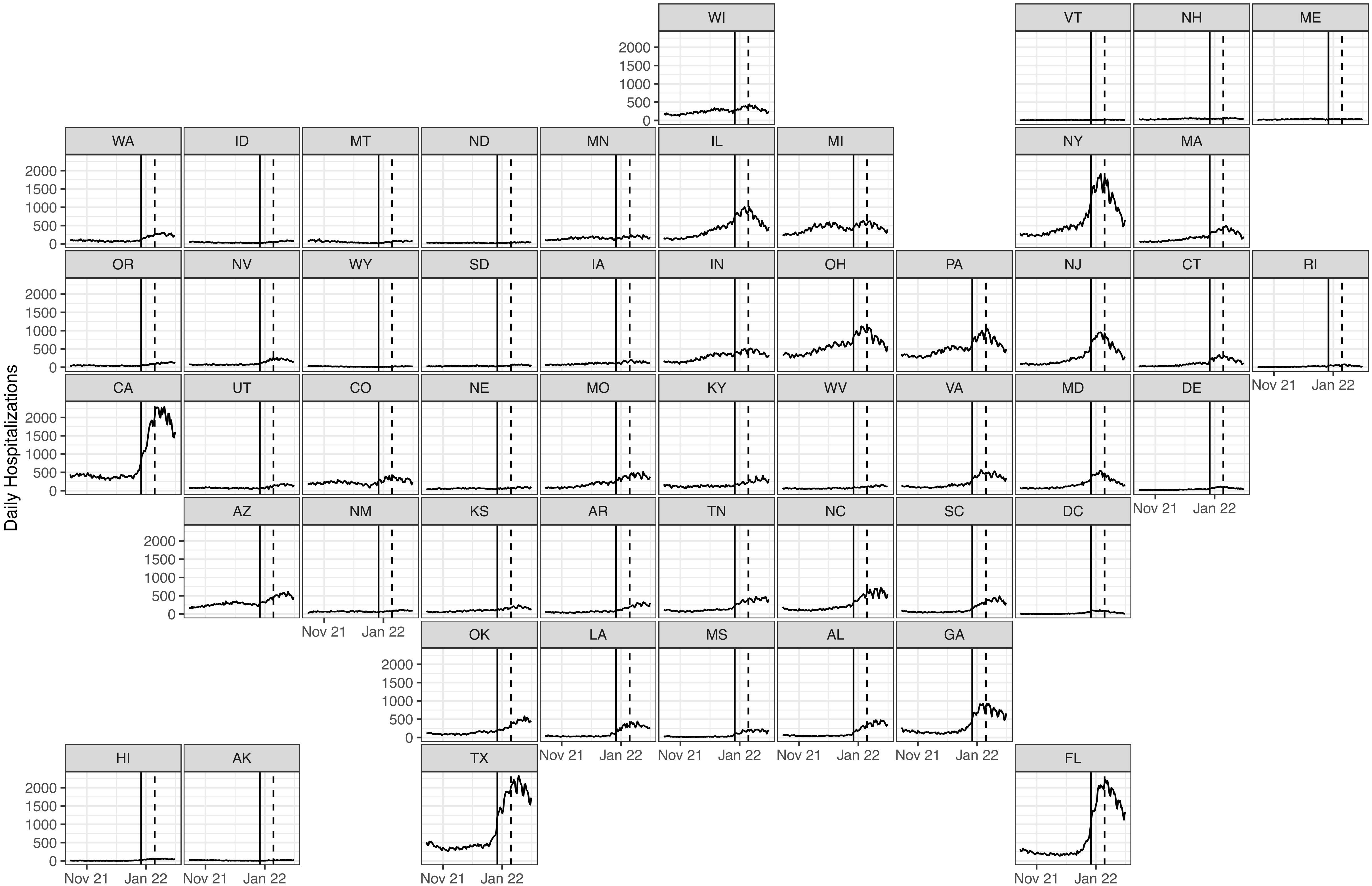
A “simple” example

- Allocation of federal resources to the US states heading into the Omicron hospitalizations wave



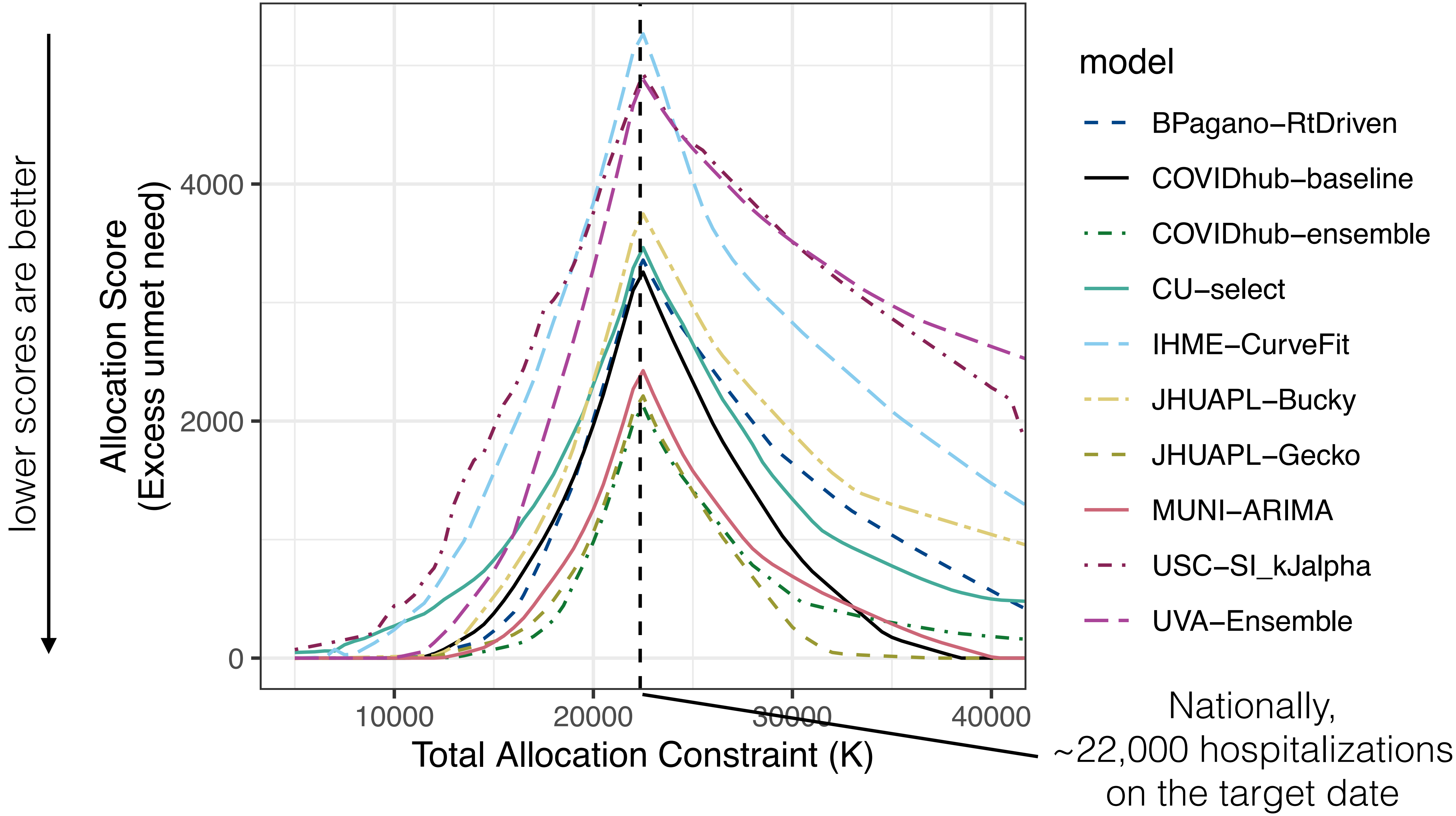
A “simple” example

- Allocation of federal resources to the US states heading into the Omicron hospitalizations wave



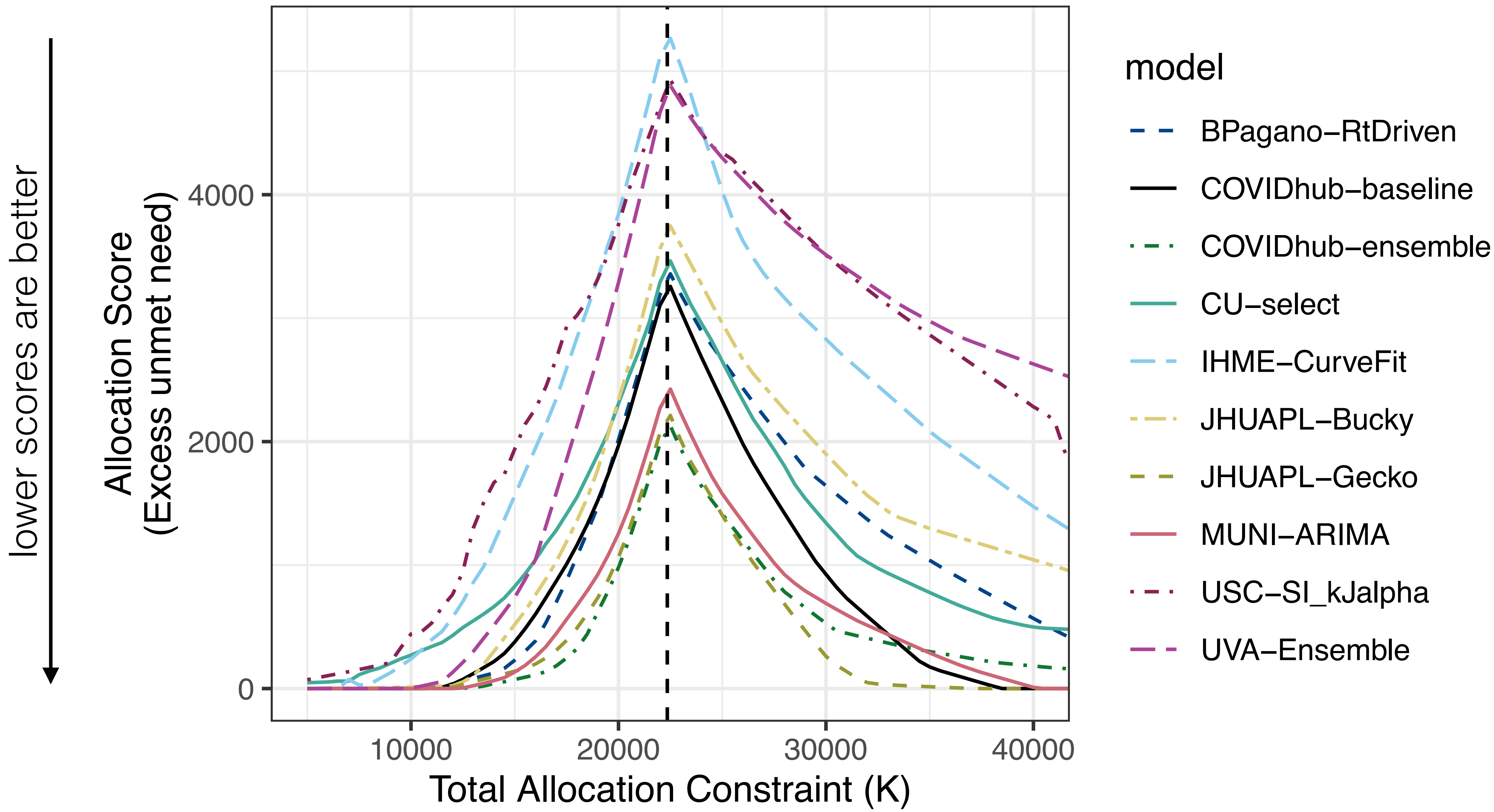
Results

- Excess loss relative to oracle forecaster: how much unmet need beyond what was unavoidable given resource constraints?
- Peaks when constraint K is equal to total observed hospitalizations
- For very large or small K, all forecasts lead to similar levels of unmet need



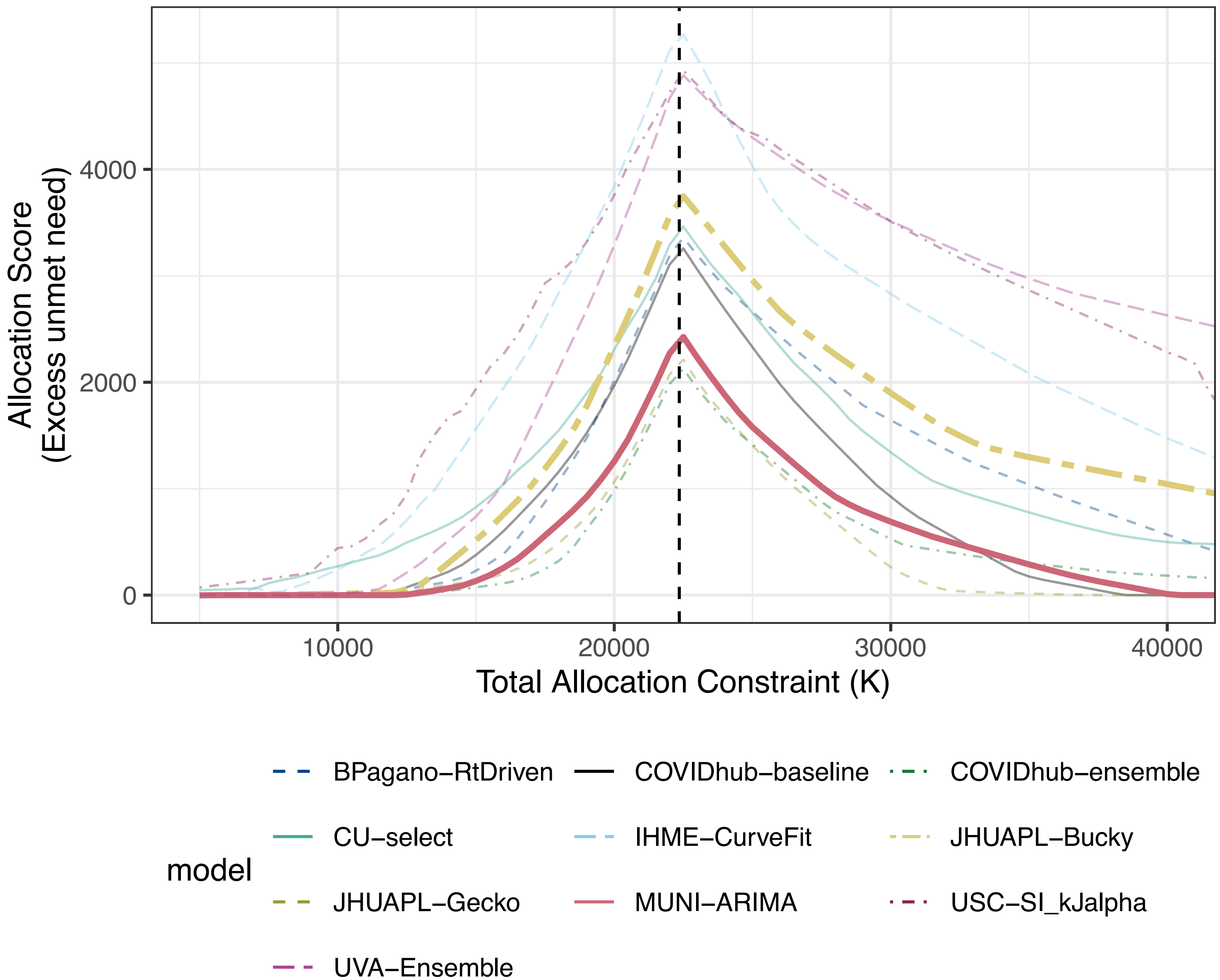
Results

- Allocation scores are fairly stable as the constraint K varies



Best allocation score ⇎ best WIS

- Consider the model pair:
  - JHUAPL-Bucky: best WIS, middling allocation score
  - MUNI-ARIMA: middling WIS, good allocation score



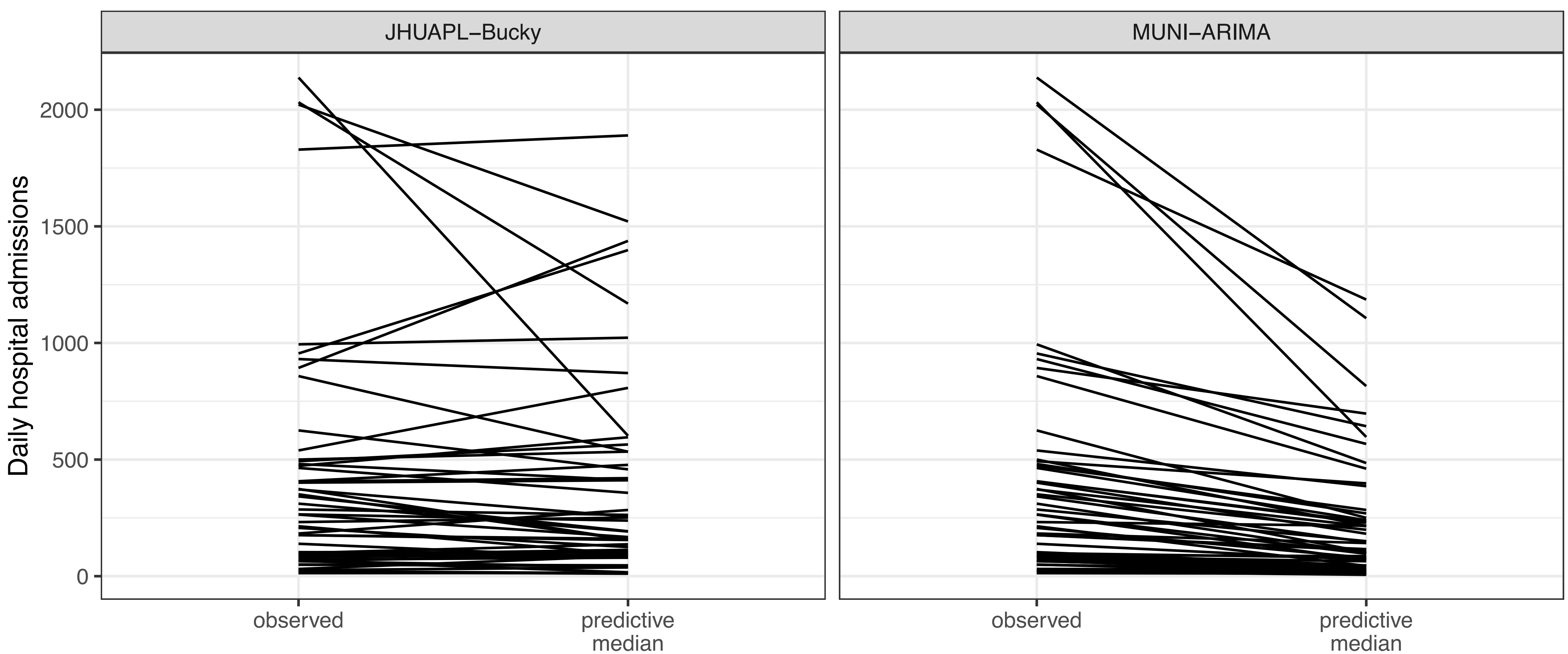
Model	WIS
JHUAPL-Bucky	**4831**
CU-select	5297
COVIDhub-ensemble	8139
JHUAPL-Gecko	8160
IHME-CurveFit	8209
USC-SI_kJalpha	8402
MUNI-ARIMA	**8668**
BPagano-RtDriven	8729
COVIDhub-baseline	9789
UVA-Ensemble	11711

Results

- JHUAPL-Bucky: best WIS, middling allocation score
- MUNI-ARIMA: middling WIS, good allocation score

- Figure compares observed values vs predictive median, one line per state

- JHUAPL-Bucky lines are closer to level on average, but more “crossings” indicate inaccurate relative rankings across states

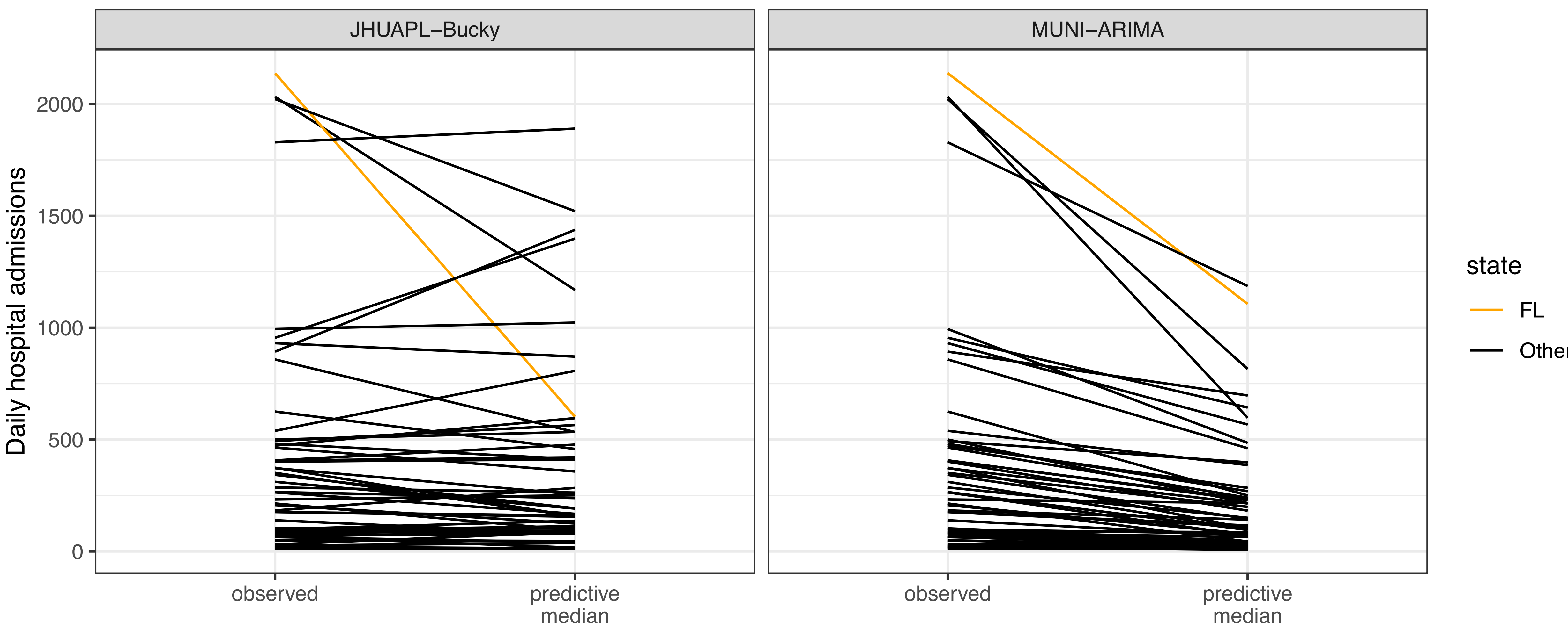


Results

- JHUAPL-Bucky: best WIS, middling allocation score
- MUNI-ARIMA: middling WIS, good allocation score

- Figure compares observed values vs predictive median, one line per state

- JHUAPL-Bucky lines are closer to level on average, but more "crossings" indicate inaccurate relative rankings across states



Limitations, future work

- In practice, decision makers use many inputs alongside model-based predictions to inform decisions
- In many (most?) instances, it's challenging to quantify the loss associated with a decision
- We do not account for important considerations such as equity/fairness of allocations
- We do not account for other broader elements of the decision-making context, such as the balance of multiple mitigation measures, increasing the resource constraint K, etc.
- It would be valuable to consider other decision-making contexts

Summary

- What is required of a forecast for it to be useful for specific decision-making purposes?
  - A forecast ***target*** that is relevant to the decision
  - A record of ***accuracy*** in relation to the decision

- In some settings, it is possible to evaluate forecast skill in ways that are responsive to the decision-making context

- Such evaluation methods can yield model rankings that are substantively different from generic measures like WIS

- We have illustrated in a simple example, but this could be taken much further

UMass Amherst | School of Public Health & Health Sciences
Biostatistics and Epidemiology

Thanks!

With acknowledgments to:
Aaron Gerding, who has done most of the work I presented today
Nicholas Reich, who has provided valuable input

The Reich Lab and the COVID-19 Forecast Hub have been supported by the National Institutes of General Medical Sciences (R35GM119582) and the US Centers for Disease Control and Prevention (1U01IP001122). The content is solely the responsibility of the authors and does not necessarily represent the official views of NIGMS, the National Institutes of Health, or US CDC.

References

- Gneiting, T. and Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. Journal of the American statistical Association, 102(477), pp.359-378.
  - 1-paragraph outline of the decision-theoretic set up
  - We have not identified a beginner-friendly reference for the general audience (suggestions?)
- Reference suggested by Johannes: Brehmer, J.R., Gneiting, T., 2020. Properization: constructing proper scoring rules via Bayes acts. *Ann Inst Stat Math* **72**, 659–673. https://doi.org/10.1007/s10463-019-00705-7
- Choi, T.M. ed., 2012. Handbook of Newsvendor problems: Models, extensions and applications (Vol. 176). Springer Science & Business Media.
  - Problems related to allocation of resources subject to constraints

Notation

- $Y = (Y_1, \ldots, Y_n)$: random variables representing the count of future hospitalizations in each of n locations
- $y = (y_1, \ldots, y_n)$: specific values of the outcome variables, not yet observed at the time the forecast is generated
- $F = (F_1, \ldots, F_n)$: forecast distributions for each location
- $x = (x_1, \ldots, x_n)$: the level of resources (e.g. hospital beds, oxygen, ventilators) allocated to each location
- K: a constraint on the total resource allocation.

$$\sum_i x_i \leq K$$

Step 1: decision loss function

- Recall notation:
  - $y = (y_1, \ldots, y_n)$: specific values of the outcome variable
  - $x = (x_1, \ldots, x_n)$: resources allocated to each location

- We could measure loss as the amount of unmet need:

$$s(\mathbf{x}, \mathbf{y}) = \sum_i (y_i - x_i)_+ = \sum_i \begin{cases} 0 \text{ if } x_i \geq y_i \\ (y_i - x_i) \text{ if } x_i < y_i \end{cases}$$

- Example:
  - We allocate $\mathbf{x} = (10, 20)$ units of resources to locations 1 and 2
  - Eventually the value $\mathbf{y} = (15, 18)$ is observed
  - Unmet need is $s(\mathbf{x}, \mathbf{y}) = (15 - 10) + 0 = 5$

- In work in progress, we generalize this in several ways
  - Most importantly, allow for varying penalties for over- or under-allocation

- Everything on this slide has been done previously

Step 2: optimal allocation given a forecast

- We could measure loss as the amount of unmet need:

$$s(\mathbf{x}, \mathbf{y}) = \sum_i (y_i - x_i)_+ = \sum_i \begin{cases} 0 \text{ if } x \geq y \\ (y - x) \text{ if } x < y \end{cases}$$

- With this set up, the optimal allocation is $x_i^F = F_i^{-1}(1 - \lambda)$, where λ is chosen so that $\sum_i x_i^F = K$

- In words: choose a quantile for all locations at a probability level such that the resource constraint is satisfied

- Everything on this slide has been done previously

Step 3: score forecasts with allocation loss

- Given a probabilistic forecast $F = (F_1, \ldots, F_n)$, find the optimal allocation as $x_i^F = F_i^{-1}(1 - \lambda)$

- Once the outcome $\mathbf{y}$ is observed, calculate the score $s(\mathbf{x}^F, \mathbf{y})$, measuring the unmet need resulting from the allocation suggested by the forecast.

- For interpretability, we subtract the score obtained from an Oracle that knows the true value of $\mathbf{y}$:

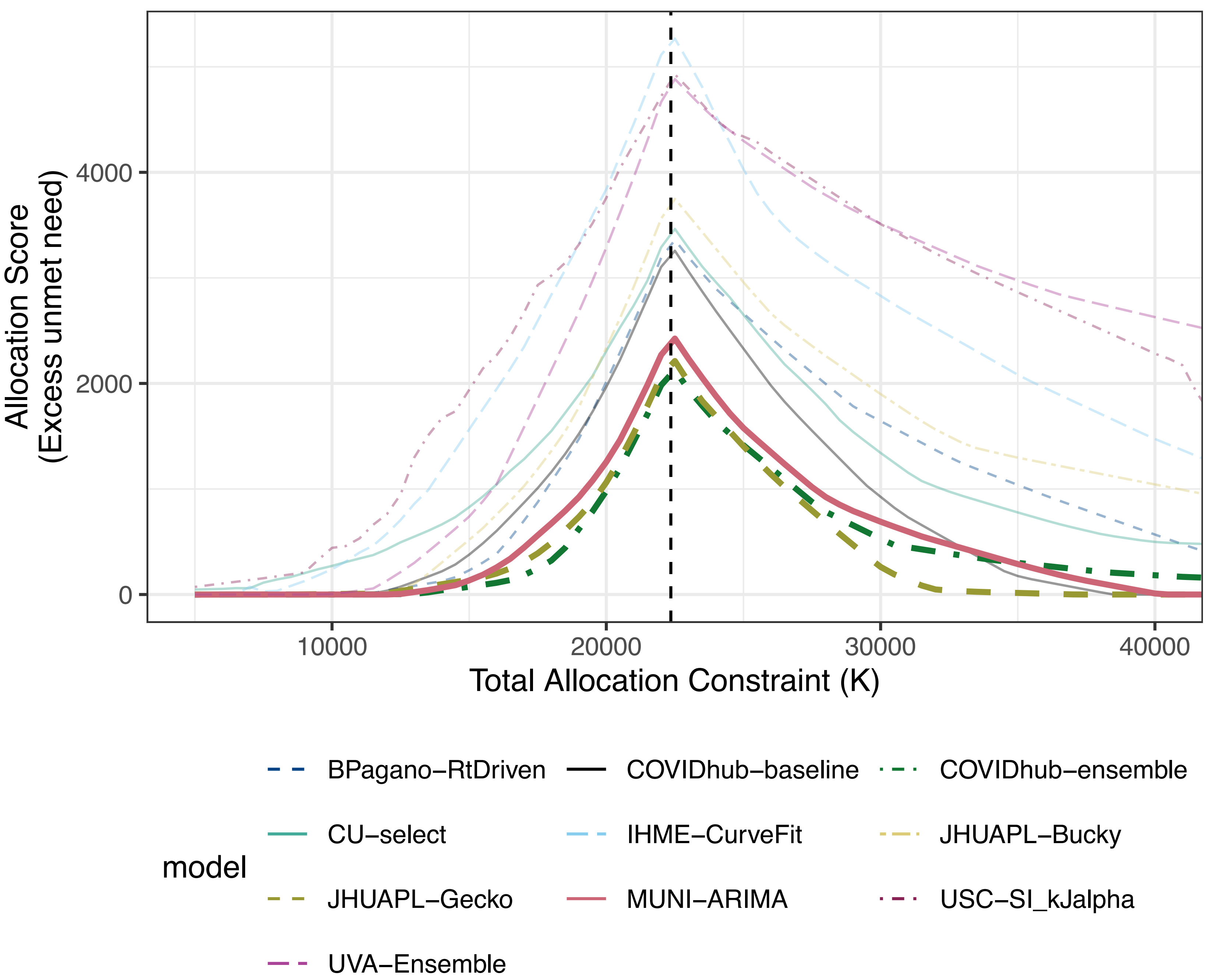
$$s(\mathbf{x}^F, \mathbf{y}) - s(\mathbf{x}^{Oracle}, \mathbf{y})$$

- Interpretation: "How much excess unmet need was there above what was necessary given the constraints and realized need?"

- To our knowledge, no one has done this previously with the allocation loss

Results

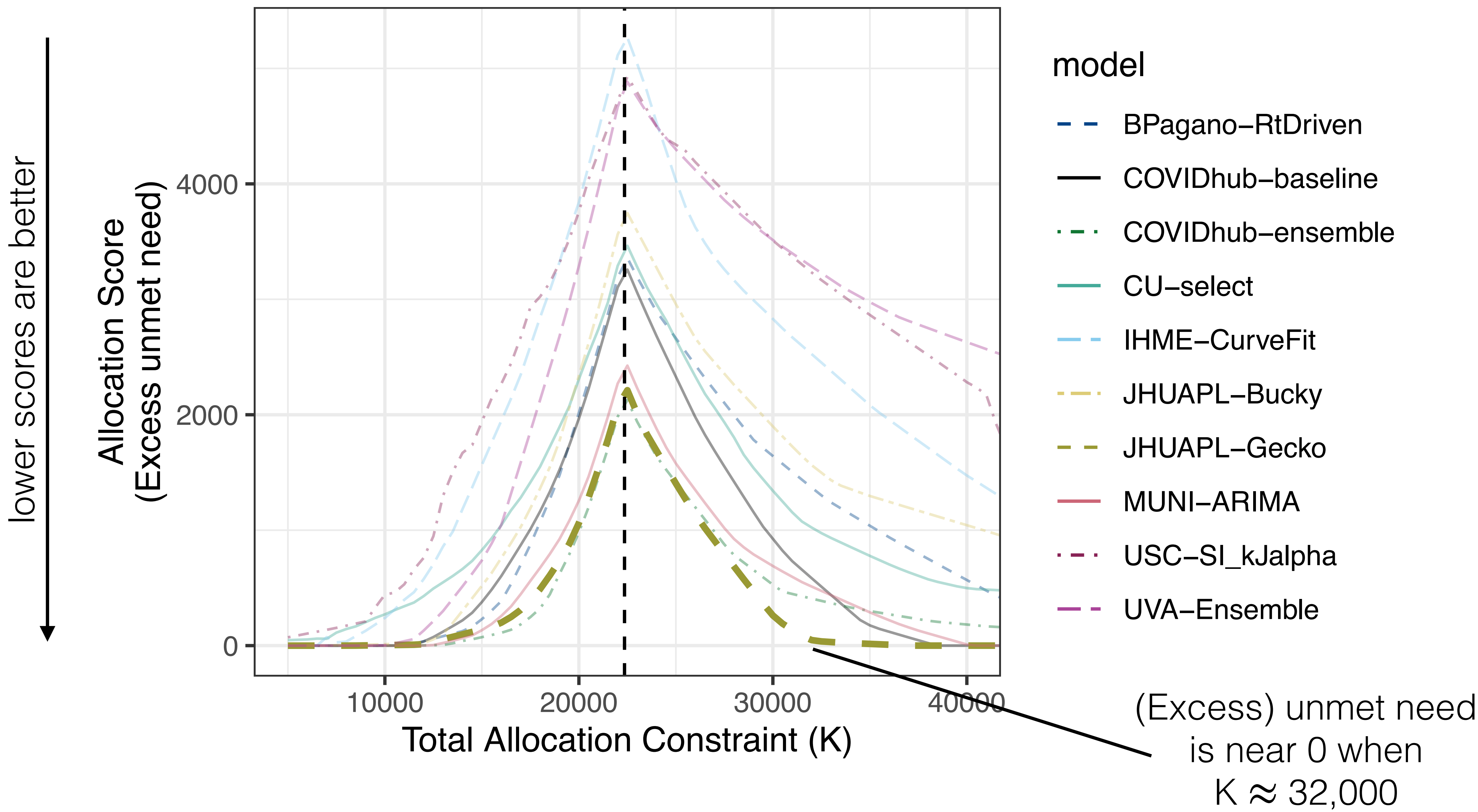
- Models with the best allocation scores do not necessarily have the best WIS



Model	WIS
UVA-Ensemble	11711
COVIDhub-baseline	9789
BPagano-RtDriven	8729
MUNI-ARIMA	**8668**
USC-SI_kJalpha	8402
IHME-CurveFit	8209
JHUAPL-Gecko	**8160**
COVIDhub-ensemble	**8139**
CU-select	5297
JHUAPL-Bucky	4831

Results

- Allocation according to JHUAPL-Gecko needed ~10,000 more beds than there were hospitalizations to achieve near 0 unmet need



Working with forecasts in quantile format

- In the application to Hub forecasts, we infer distributions from quantiles using
  - a monotonic spline to interpolate the quantiles
  - parametric assumptions about tail behavior
- Our current thinking (to be thought through carefully): The score is still proper, but we believe that the “quantiles” elicited by this process are not quantiles.
- More thought would be needed to either adjust the score or the forecast representation for use as an official scoring metric for a Hub.