Erkin Ötleş
hi@eotles.com

Board Preparation

A Primer on Artificial Intelligence / Machine Learning in Medicine

OR The things you should feel comfortable asking as an intern

NOTE: I use AI & ML synonymously

FDA certified 692nd AI models in 2023
- 5y ago ≤ 75
- Only the tip of the iceberg
- Most not certified
- Not necessary if informing MD

Growing # of models
- Onus on MDs to catch bad ones

Talk Goal:
- Not making you an expert
- But make you comfortable asking Qs

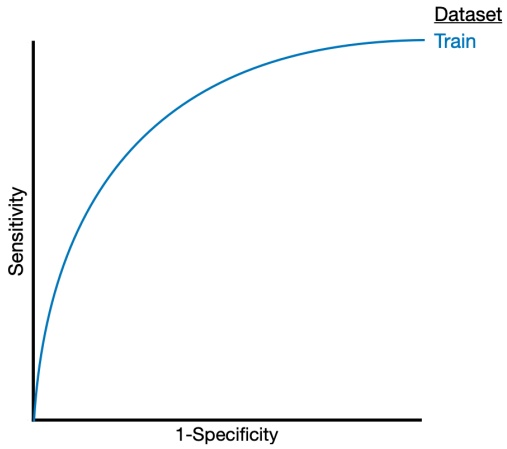
If interacting with an new AI model
- feel comfortable asking 3 Qs

1. How was it made?
2. Is it any good?
3. How is it being used?

How was it made?

Pt	Sex	Age	Vitals	...	Abx	Future Sepsis
1	M	65	130/...	...	Yes	Yes
2	F	70	...	...	Yes	Yes
3	M	72	...	...	No	No

Is it any good?

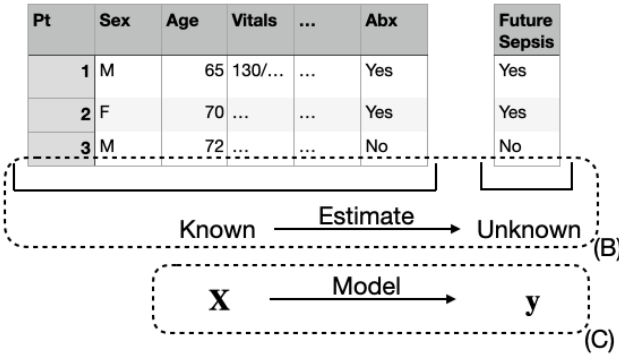


How is this being used?

$$\text{EHR} \rightarrow \mathbf{X} \rightarrow \text{Model} \rightarrow \mathbf{y}$$

Erkin Ötleş
hi@eotles.com

How was it made?
Development (A)



How was it made?

Engineers call this *development*
(A)

Goal: use AI to estimate unknown info from known info
(B)

Model: thing that does the estimation

To make a model:
- Collect lots of data
- Use math & programming

Note: data is retrospective bc we need to know the unknown info

Engineers refer to known data as $\mathbf{X}$
& unknown data/target outcome as $\mathbf{y}$

So a model uses $\mathbf{X}$ to estimate $\mathbf{y}$
(C)

It is important we use the right info for $\mathbf{X}$, $\mathbf{Y}$
- clinical perspective is essential
- accidental inclusion of info tied $\mathbf{y}$

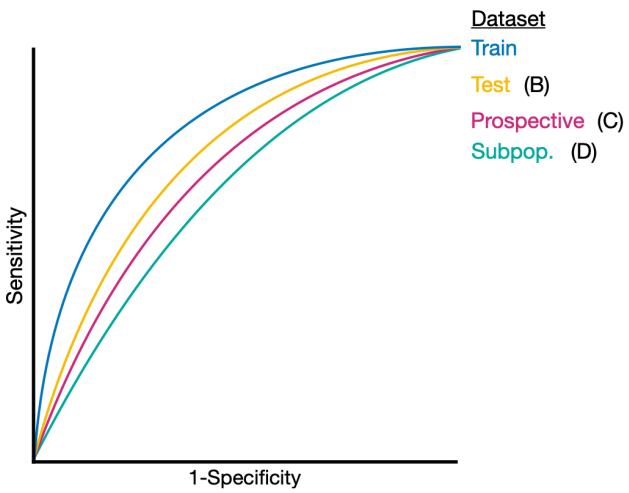
Toy example.
Let's say we have a sepsis prediction model that uses info about fluid & abx admin.
Why might this be problematic?

Markers of resuscitation
Model depends on them
The alerts arrive too late
↓ clinical utility

An extreme case.
But studies we have shown evidence of models inadvertently relying on clinical markers of the outcome. [1]

Erkin Ötleş
hi@eotles.com

Is it any good?
Validation (A)



Is it any good?
Called *validation*
(A)

Test the model's *performance* on data it hasn't seen before

Performance measurement should be familiar
- closely related to EBM & biostats
- Can you think of some measures?

- Accuracy,
- Sensitivity, specificity
- PPV, NPV
- NNT
- Area under the ROC curve

What are data the model hasn't seen before?
1. Data held aside for testing (B)
2. Data seen in prospective usage (C)

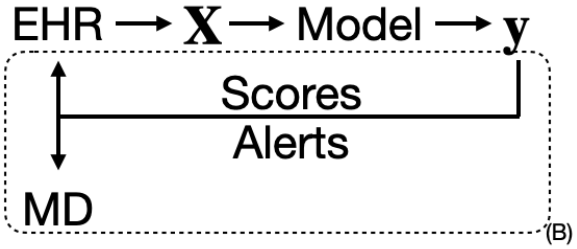
Why do we assess on held aside data?
- Models can memorize training data, *Overfitting*
- Won't work in well new situations

Related, you want to understand what populations this model works well on
- Models made for 1 pop may not work well for others. [2]
- Models often optimize for the majority
- Need to check vulnerable pops. (D)

Finally, validation should happen multiple times
- Across model development
- Through the transition to clinical use
- And then continually thereafter

Erkin Ötleş
hi@eotles.com

How is this being used?
Implementation (A)



(B)

How is this being used?
Implementation
(A)

Developers attempt to connect the model to clinical workflows

Can you think of ways to do this?
(B)

The ultimate goal is to deliver info that is :

timely
clinically relevant
to the right clinician

Developers don't know clinical workflows

If you receiving alerts that:

Don't add value
Don't improve your decision making

Feel empowered to ask for improvements [3]

You are the clinical expert you know what you need to know, the developers do not.

Erkin Ötleş
hi@eotles.com

Summary

If you're concerned about a model or even curious about it

You should ask:
1. How was it made?
2. Is it any good?
3. How is it being used?

Every developer should be able to answer these questions
Or share a paper that studies them

Note: if you're looking for info regarding Epic models you need to log into their documentation site.
Contact me if you want help

Erkin Ötleş
hi@eotles.com

Bibliography

1. Wong, A., et al., *External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients.* JAMA Internal Medicine, 2021.
2. Ötleş, E., et al., *Development and validation of models to predict pathological outcomes of radical prostatectomy in regional and national cohorts.* The Journal of Urology, 2022. **207**(2): p. 358-366.
3. Wong, A., et al., *Quantification of Sepsis Model Alerts in 24 US Hospitals Before and During the COVID-19 Pandemic.* JAMA Network Open, 2021. **4**(11): p. e2135286.