The main idea is to asset uncertainty in specific prediction – confidence interval for certain prediction.

Potential use cases:

1. One simple use case can be as complementary for explainability and reveal the uncertainty for specific prediction – it might improve the confidence and credibility of the model. I am not sure it is important enough to start with.
2. Use to detect groups that the model is uncertain about and lead to ideas to improve the model – feature engineer, etc. Not for now.
3. Use as part of the process and link to action items to reduce uncertainty. For example, if the uncertainty is due to missing lab tests – send patient to lab tests...

We will focus on potential use case #3 and try to evaluate it on 2 use cases: LGI and lung cancer.

We have two options:

~~3a. Trigger by lab test (as the situation is now)~~

3b. Independent periodically check - to point out patients without recent lab tests that would improve process performance after completing the lab tests.

For now we start with 3b.

Primary endpoint:

Under XXX resources for colonoscopies screening, we want to maximize the #true cases by completing additional YYY lab tests for patients before running their scores. how many additional cases were found using the additional preliminary lab tests to enrich the data?

Retrospective simulation

Since we can “add” lab tests to patients we will need to erase tests to mimic ability to add tests.

My suggestion for the simulation is that we will simulate the experiment in this manner for each patient:

* Select random Hgb test date – can happened multiple times
  + Add history\_limit rep processor of 1 day to “erase” the lab tests in that day or set prediction date on 1 day before the Hgb test.
  + Since in LGI model the expected input is to have Hgb test in the prediction date we will set prediction date to previous Hgb test but will “test” the outcome (cancer no cancer) from the current Hgb test

It will mimic what happens to the model when you don’t use most recent Hgb test but the one before it on whole population and whether we can find who to screen by lab tests (and then use most recent Hgb test) to improve performance. We can think about it differently – it mimics what will happen if the health system will stop sending patients to CBC’s can we help them with that?

Potential issue: we select Hgb test before cancer, so if we move prediction date to a test before it will cause bias that patient will not have cancer is short time windows. Isn’t that similar for testing performance in time window 180-365 (if next Hgb test is 6 months ahead?) – if patient had cancer within 6 months, he is excluded so we can be certain he does not cancer in the next 6 months.

Final measures

# of detected cancers (or PPV which is equivalent on same population) in 97% specificity threshold. And primary endpoint. We might evaluate it on different cutoffs as well

We will compare both arms A and B:

* Arm A – no additional Hgb tests (We will use previous most recent Hgb tests) or random sampling of test/ or other baseline logic
* Arm B – after completing additional YYY lab tests – what is out performance

Arm A is the baseline which mimics what happens if no more Hgb tests will be performed in the future. Arm B will be a potential solution for the problem.

Assumptions:

1. Lab tests are preformed unconditional with outcome
2. Full compliance to blood tests
3. Short time interval for performing lab test

Strategies:

1. Compare to train additional model for CRC with last time that is not always 0
2. Train model that predicts who will change state from flagged to unflagged and vice versa and use this model to perform lab test
3. Data explorations for