

Predicting COPD-Related ED/Inpatient Visits: How We Handle Label Uncertainty

The Challenge

We want to identify patients at high risk for a COPD-related emergency department or inpatient visit in the next 6 months, so we can proactively send them emergency packs.

The easy part: When a patient has a COPD-related ED or inpatient visit at a Mercy facility, we see it clearly in the record.

The hard part: How do we confirm a patient *didn't* have a COPD-related ED or inpatient visit? For patients with a Mercy PCP, only about 20% of their ED and inpatient visits occur at Mercy facilities. The other 80% happen elsewhere—at non-Mercy hospitals, freestanding ERs, or urgent care centers we don't have visibility into.

This creates a labeling problem: when we mark a patient as “no ED/IP visit,” we’re really saying “no ED/IP visit *that we observed*.” Some of those patients did have visits—just not at our facilities.

Option A: Tiered Label Confidence

Rather than treating all “no visit” patients the same, we categorize them by how confident we are in that label:

Tier 1 – High Confidence Negative

The patient returned to Mercy for care 7-12 months after the observation date, and we see no COPD-related ED or inpatient visit documented during the 6-month prediction window.

Why this matters: The return visit confirms the patient is still engaged with our system. If they’d had a major COPD event requiring emergency care, there’s a reasonable chance it would appear in our records—either because they came to us, or because the outside visit was communicated back during their follow-up care.

Tier 2 – Medium Confidence Negative

The patient returned to Mercy 4-6 months after observation AND shows signs of ongoing COPD engagement with our system:

- Filled a maintenance inhaler (like Advair, Symbicort, or Breo) at a Mercy pharmacy in the last 90 days, OR
- Had a pulmonology visit at Mercy in the prior 12 months, OR
- Had two or more PCP visits at Mercy in the prior 12 months

Why this matters: The return visit covers most of the prediction window. The engagement signals suggest this patient manages their COPD through Mercy—they’re not just loosely affiliated. Patients actively engaged with us for COPD care are more likely to come to a Mercy facility when acutely ill, and more likely to have outside events documented when they follow up.

Tier 3 – Assumed Negative

The patient didn’t return to Mercy during the observation period, but has a Mercy PCP AND at least one of the engagement signals above.

Why this matters: We can’t confirm they stayed out of the ED. But their engagement pattern suggests they consider Mercy their healthcare home. We assume that if they’d had a serious COPD exacerbation, it would more likely appear in our system than for a completely disengaged patient.

Who We Exclude Entirely

Patients with no return visit AND no engagement signals. These patients may have left our system, switched to another health network, or simply aren't engaged enough for us to have meaningful visibility. Including them would add more noise than signal.

Why Include Tier 2 and Tier 3 at All?

A reasonable question: "*If we only see 20% of ED visits, isn't it risky to assume patients are negative?*"

Three reasons we include them:

1. **Statistically:** Even with some mislabeled patients, the overall signal is strong enough to learn from. A patient who has frequent rescue inhaler fills, multiple prior exacerbations, and poorly controlled comorbidities is high-risk regardless of which ED they go to. The clinical patterns that predict ED visits are the same whether the visit happens at Mercy or elsewhere.
2. **Pragmatically:** If we only used patients with perfect confirmation (Tier 1), we'd throw out a large portion of our training data. Worse, the patients we'd keep might be systematically different—perhaps healthier, more engaged, or living closer to Mercy facilities. That would bias the model in ways that could hurt its real-world performance.
3. **Clinically:** Patients who actively manage their COPD through Mercy—filling inhalers with us, seeing our pulmonologists, visiting their Mercy PCP regularly—are more likely to use Mercy facilities when they get sick. The engagement signals aren't perfect proxies for system loyalty, but they're meaningful. These patients aren't random; they've chosen to receive their COPD care from us.

Option B: Same Tiers, With Explicit Capture Rate Acknowledgment

This option uses the identical tier structure as Option A, but frames the methodology differently for clinical audiences.

The key reframe: We are not claiming to predict *all* COPD-related ED/IP visits. We are building a model that identifies patients at high clinical risk for COPD exacerbation severe enough to require emergency care.

What we observe: Approximately 20% of ED/IP visits for our Mercy PCP population occur at Mercy facilities.

What we assume: The clinical warning signs that predict a Mercy ED visit are the same warning signs that predict an outside ED visit. A patient with worsening shortness of breath, escalating rescue inhaler use, and a recent steroid burst is at high risk for an ED visit—it doesn't matter which ED they walk into.

How we validate this assumption: After the model is built, we examine what features it considers most important. If the top predictors are clinically meaningful—things like:

- Prior COPD exacerbations
- Rescue inhaler overuse
- Gaps in maintenance inhaler fills
- Comorbidities like heart failure or anxiety
- Recent oral steroid courses

...then we have confidence the model learned true clinical risk. If instead the top predictors are things like "distance to nearest Mercy hospital" or "number of prior Mercy ED visits," that would suggest the model learned *access patterns* rather than *clinical risk*—and we'd need to revisit our approach.

The bottom line for Option B: We're transparent that we only observe a fraction of the outcome. But we believe—and will validate—that the clinical signals are transferable. A high-risk patient is high-risk regardless of which door they walk through when they can't breathe.

Which Option to Choose?

Option A is the more conservative framing. It focuses on how we handle label uncertainty and why our tiered approach is methodologically sound. It doesn't dwell on the 20% capture rate.

Option B is the more transparent framing. It acknowledges the capture rate directly and makes an explicit argument for why the model should still work. It invites clinical validation of the assumption.

Both use the same underlying methodology. The difference is how much we emphasize the known limitation versus the reasoning for why it's acceptable.