Abstract Presentation 5

International Primary Care Respiratory Group
IPCRG work locally collaborate globally

Asthma/COPD Differentiation Classification (AC/DC):

Machine learning to aid physicians in diagnosing asthma, COPD and asthma-COPD overlap (ACO)

- Alan Kaplan, MD

Chair AC/DC Steering Committee

Breathing and feeling well through universal access to right care

AC/DC Steering Committee

Alan Kaplan (Chair)
University of Toronto, Canada

David Price
Observational and Pragmatic Research Institute, Singapore

Claus Vogelmeier
Philipps Universität Marburg, Germany

Helen Reddel
University of Sydney, Australia

Mark FitzGerald
The University of British Columbia, Canada

Janwillem Kocks
General Practitioners Research Institute, Groningen, Netherlands

Ioanna Tsiligianni
University of Crete, Greece

Kostas Kostikas
University of Ioannina Medical School, Greece

Disclosures: Alan Kaplan is a medical advisor or on the speakers' bureau for: AstraZeneca, Behring, Boehringer Ingelheim, Covis, Grifols, GSK, Merck Frosst, Pfizer, Novartis, Novo Nordisk, Sanofi, Teva, and Trudell

What is the problem?

Physicians misdiagnose

COPD and Asthma

So what?

- Up to **50%** of patients are misdiagnosed between asthma and COPD[1]
- This leads to inappropriate prescribing with possibility of harm[2–4]
- Basic Tenants:
  - No ICS Monotherapy for COPD
  - No LABA Monotherapy for asthma

1. Tinkelman et al. J Asthma 2006;43:75–80; 2.Hangaard et al. Respir Med. 2017;129:63-84. 3. Aaron et al. Am J Respir Crit Care Med. 2018;198(8):1012-1020. 4. Levy M et al. Why asthma still kills: the National Review of Asthma Deaths (NRAD) Confidential Enquiry report. London, Royal College of Physicians, 2014.

History of Artificial Intelligence (AI)

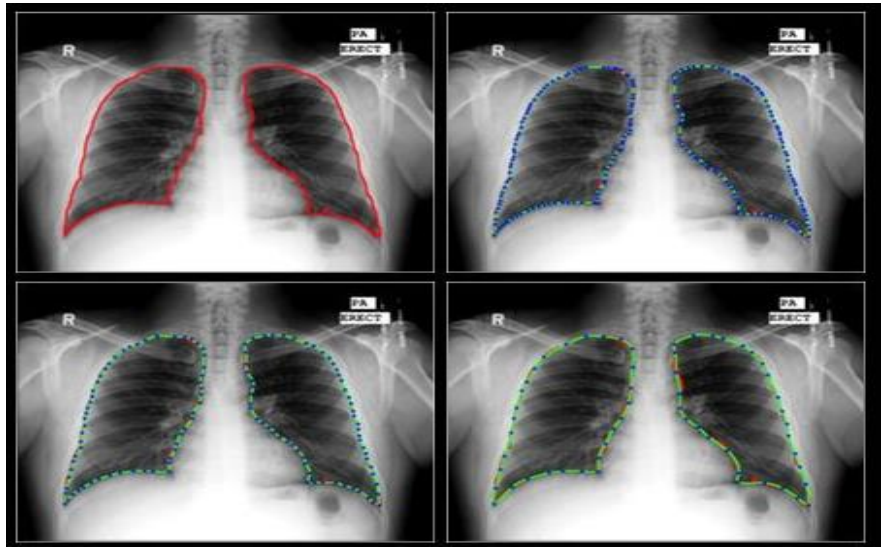
- Papers discussing AI in medicine first appeared in the literature as early as 1987[1]

- Discussion on the ability of computers to assist with difficult diagnoses

1. Schwartz et al. N Engl J Med 1987;316:685-688

AI evidence in chest imaging

- The AI team **outperforms all three Stanford radiologists** on the test set!
- demonstrating the role that AI can play in providing precise medical diagnostics, especially in underserved areas of the world,



AI vs. Humans: AI Solution Beats Stanford Radiologists in Chest X-ray Diagnostics Competition

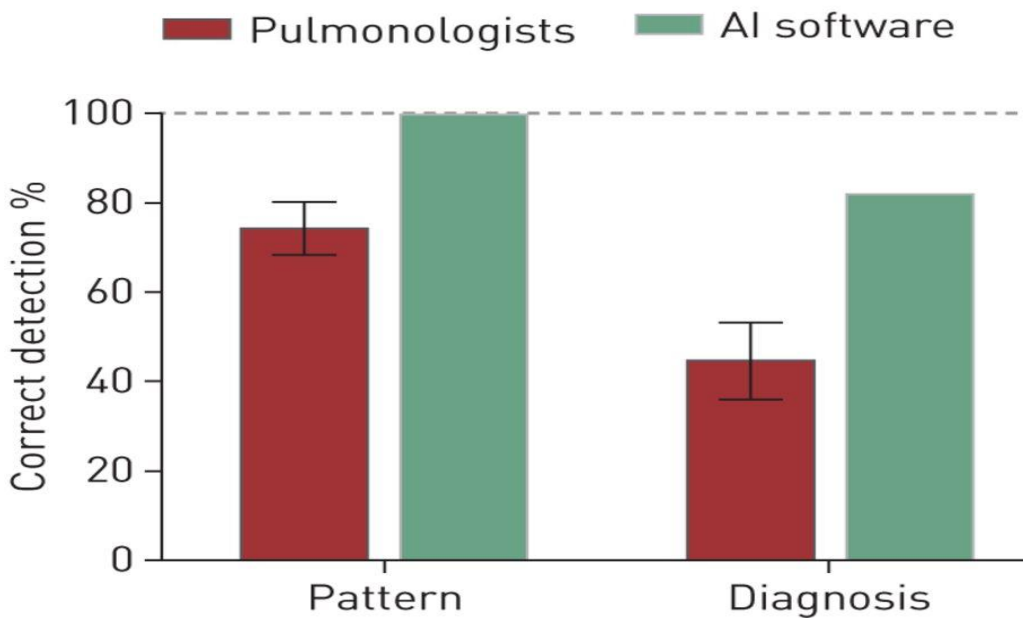
August 27, 2019

https://hitconsultant.net/2019/08/22/ai-tech-beats-radiologists-in-stanford-chest-x-ray-diagnostic-competition/#.XtjEKv9KiUk

Rajpurkar et al. PLoS Med 2018;15(11):e10026863.

AI outperforms pulmonologists in interpreting PFTs

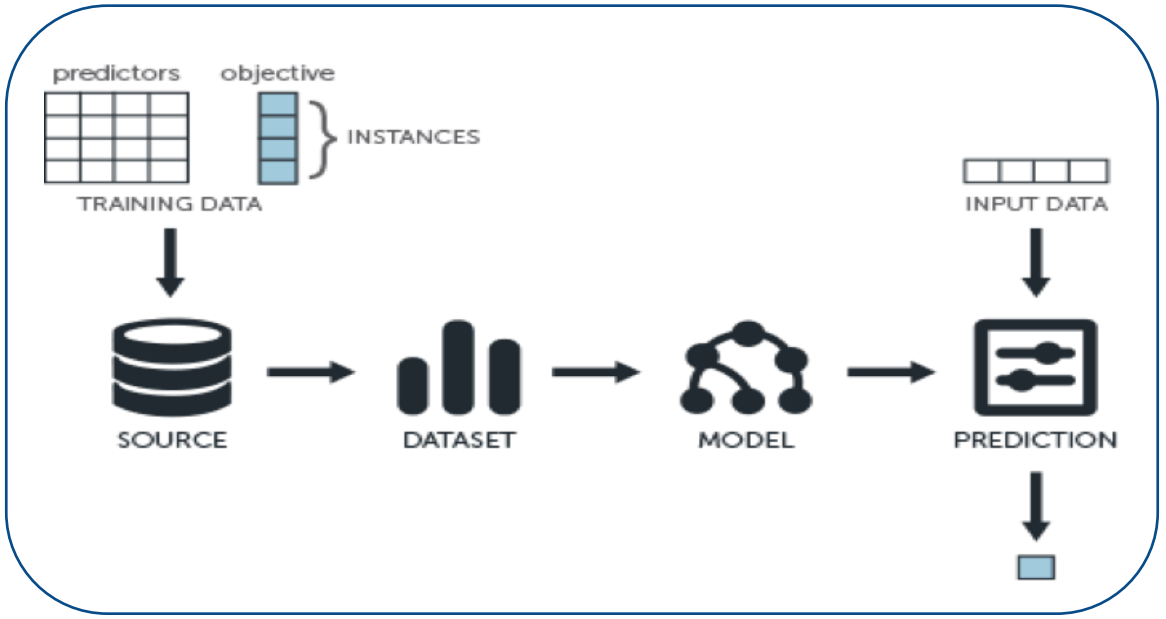
- 120 pulmonologists from 16 European hospitals evaluated 50 cases comprising PFT and clinical information resulting in 6000 independent interpretations.
  - ATS/ERS guidelines were used as the gold standard for PFT pattern interpretation. The gold standard for diagnosis was derived from clinical history, PFT and all additional tests
- Compared with AI-based software for PFT interpretation and diagnostic suggestion
- Pulmonologists made correct diagnoses in **44.6%** (±8.7) of the cases (range: 24−62%) .
- The AI-based software assigned a correct diagnosis in **82%** of all cases ($p<0.0001$ for both measures).



Topalovic et al Eur Respir J 2019;53:1801660

What is a machine learning model?

- A machine learning model is a statistical method to empirically relate **features** to **labels**.
- Once trained, **new data** of the same structure **can be fed** into the model **to predict the label**.
- Once the data is organized, the ML model can be employed!



What is the purpose of AC/DC?

Machine Learning-Assisted
Asthma **C**OPD **D**ifferentiation **C**lassification

On-the-ground support for **general practitioners** to help diagnose **Asthma from COPD**

Intended use statement for AC/DC

The intended use for the Novartis Asthma COPD Differentiation Classification (AC/DC) System is to assist clinicians in the differential diagnosis of asthma, COPD or asthma/COPD Overlap (ACO) in symptomatic patients age 35 years and above. These patients are either:

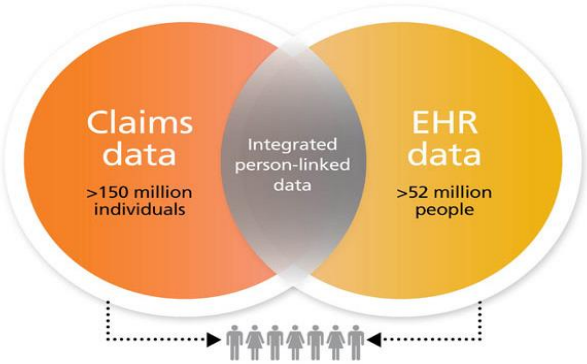
a) suspected of having one of these diseases; or b) were previously diagnosed with asthma, COPD or ACO but reconfirmation of the diagnosis is required.

Data source for ML – Optum® de-identified Electronic Health Record dataset

Staging area

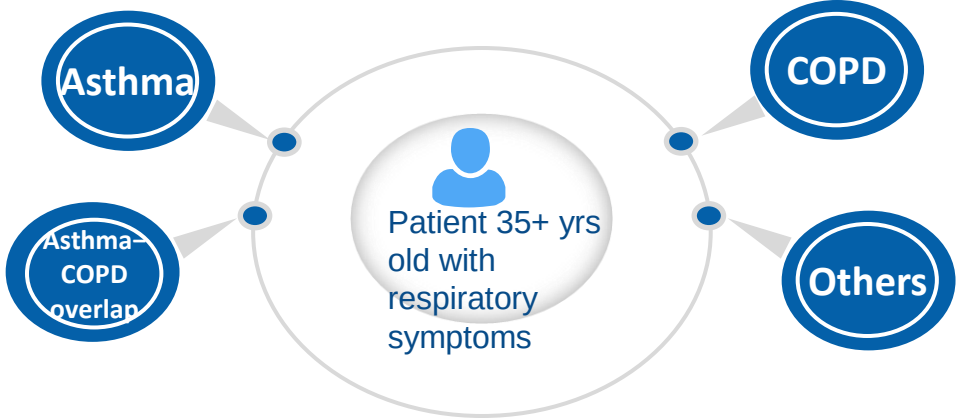
- Demographics
- Diagnosis
- Hospitalizations
- Lab results
- Medications
- Observations
- Provider notes (**NLP**)
- Outpatient visits
- Procedures
- Vital signs

Processing: validation, normalization, standardization, mapping



Longitudinal comprehensive clinical dataset

Case identification



Asthma patients w/o COPD

- A diagnosis of asthma in at least two outpatient visits given by a pulmonologist/ allergist in the identification period, *or* at least one hospitalization discharge with a primary diagnosis of asthma
- No COPD diagnosis in any occasions (outpatient visit, ER visit or hospitalization) in the identification period

COPD patients w/o asthma

- A diagnosis of COPD in at least two outpatient visits by a pulmonologist in the identification period, or at least one hospitalization discharge with a primary diagnosis of COPD
- No asthma diagnosis in any occasions (outpatient visit, ER visit or hospitalization) in the identification

What about overlap?

- It is challenging to identify true patients with overlap of two diagnoses
- Identify patients who were continuously cared for both conditions

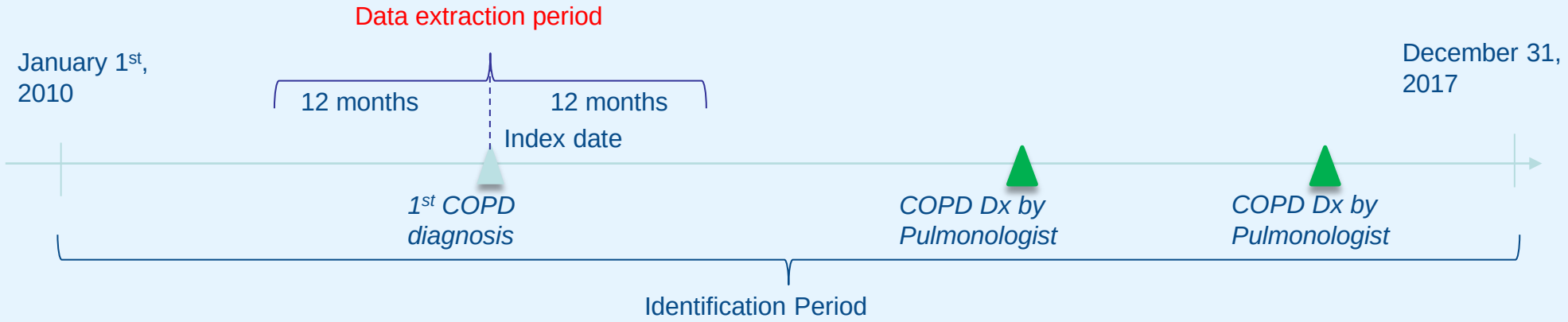
Patients with overlap of two diagnoses

1. Patients have at least two COPD diagnoses and two asthma diagnoses in the identification period (2010-2017)
2. The very first diagnosis of the later disease is the index date
3. Patients are at least 35 years old at index
4. In one year post index, patients have
   a. two COPD-related office visit and two asthma-related office visits, OR two office visits with both COPD and asthma diagnosis
      OR
   b. one hospitalization with both COPD and asthma diagnoses

Data extraction

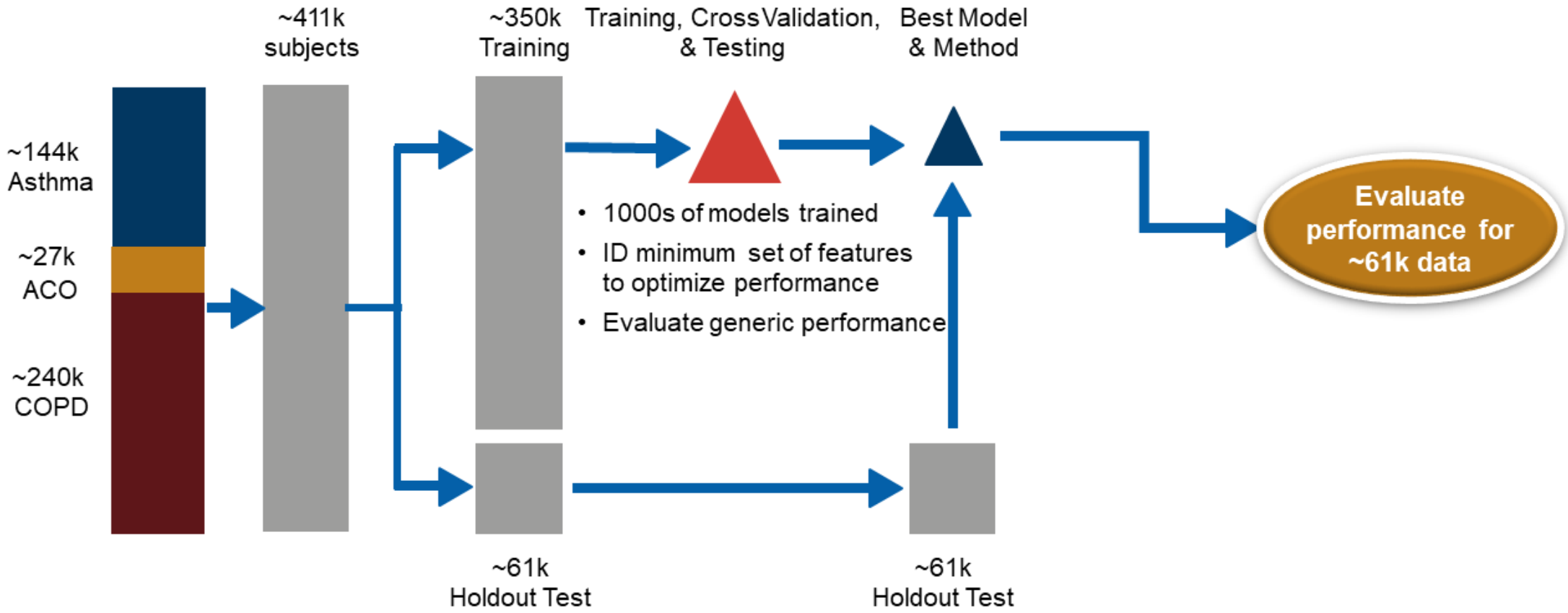
- It was agreed that the cohorts of newly diagnosed patients are suitable for this AC/DC project.
  - The index will be the first diagnosis date
  - The data extraction period is 12 months pre & post index
- For patients with overlap of two diagnoses
  - The index will be the first diagnosis date of 2nd condition
- For patients with respiratory signs/symptoms but no airway disease
  - The index will be the date of the first recorded respiratory signs/symptoms

Graphic illustration for data extraction – COPD as example



Validation of the AC/DC diagnostic tool

- The holdout data set of 61,000 case records (approx.) was subsequently used to evaluate the performance of the algorithm derived by the XGB model



AC/DC, Asthma/COPD Differentiation Classification; ACO, asthma-COPD overlap; COPD, chronic obstructive pulmonary disease; XGB, Extreme Gradient

What features did we use?

60+ features

Demographics

'Gender', 'Race', 'Ethnicity', 'Diagnosis Date', 'Age at Diagnosis', 'Cohort', 'BMI', 'Height', 'Weight', 'Pack Year',

PFTs and Symptoms

'FEV1', 'FEV1/FVC', 'FVC', 'Smoking Status', 'Wheeze Incidence', 'Cough Incidence', 'Dyspnea Incidence', 'Tight-chest Incidence', 'Sputum Incidence', 'Rhinitis Incidence'

Lab Paneling

'RBC', 'HGB', 'HCT', 'MCV', 'MCH', 'MCHC', 'PLT', 'MPV', 'RDW-CV', 'Basophils', 'Eosinophils', 'Lymphocytes', 'Monocytes', 'Neutrophils', 'Diagnosis Month'

Relevant Diagnosis Info

'Chronic Rhinitis Dx', 'GERD Dx', 'Hemoptysis Dx', 'Allergic Rhinitis Dx'

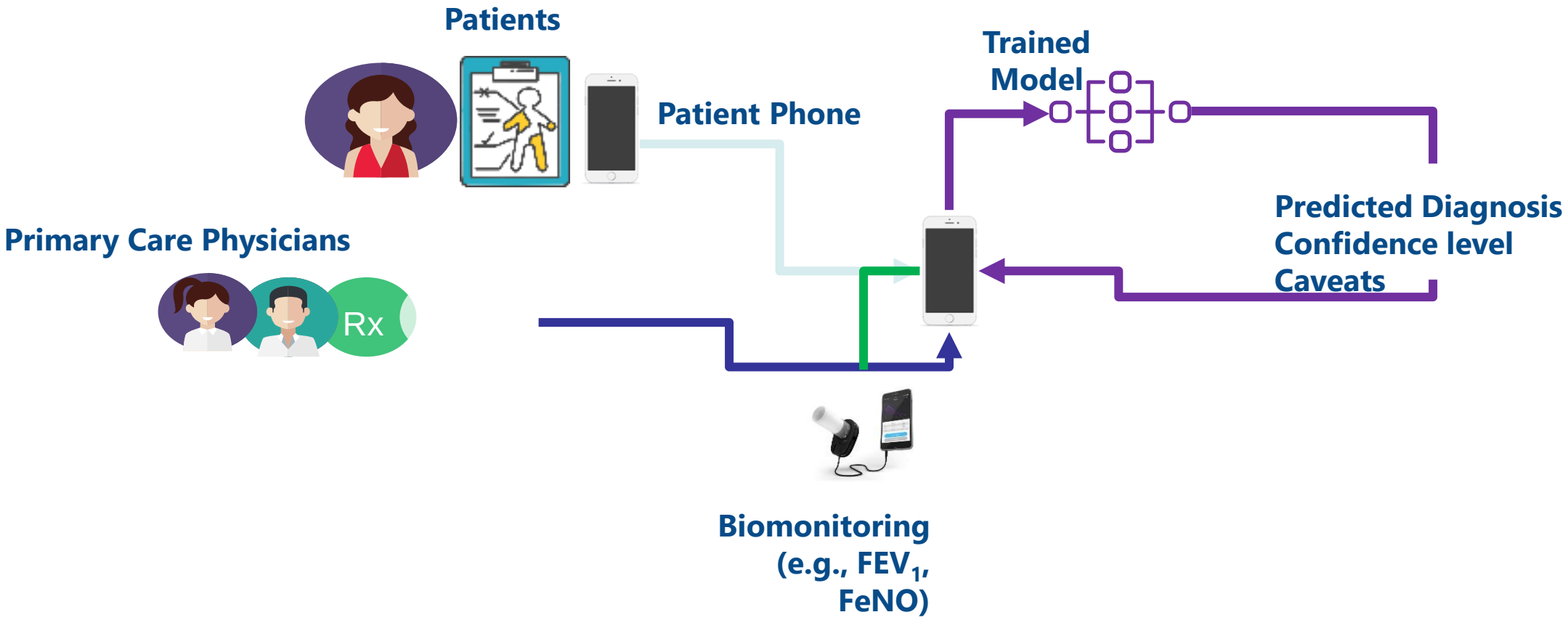
Comorbidities

'Myocardial Infarction', 'Congestive Heart Failure', 'Peripheral Vascular Disease', 'Cerebrovascular Disease', 'Dementia', 'Chronic Pulmonary Disease', 'Rheumatic Disease', 'Peptic Ulcer Disease', 'Mild Liver Disease', 'Diabetes Without Chronic Complications', 'Diabetes With Chronic Complications', 'Hemiplegia or Paraplegia', 'Renal Disease', 'Any Malignancy', 'Moderate or Severe Liver Disease', 'Metastatic Solid Tumor', 'AIDS/HIV', 'charlson_index',

Top features of importance by computer: Cohorts I, II, and III

- FEV_1
- FEV_1/FVC
- Pack Year
- Age
- BMI
- Dyspnea
- Wheeze
- Cough
- DX Allergic Rhinitis
- Current smoker
- Never smoked
- DX Chronic Rhinitis

How will the service work?

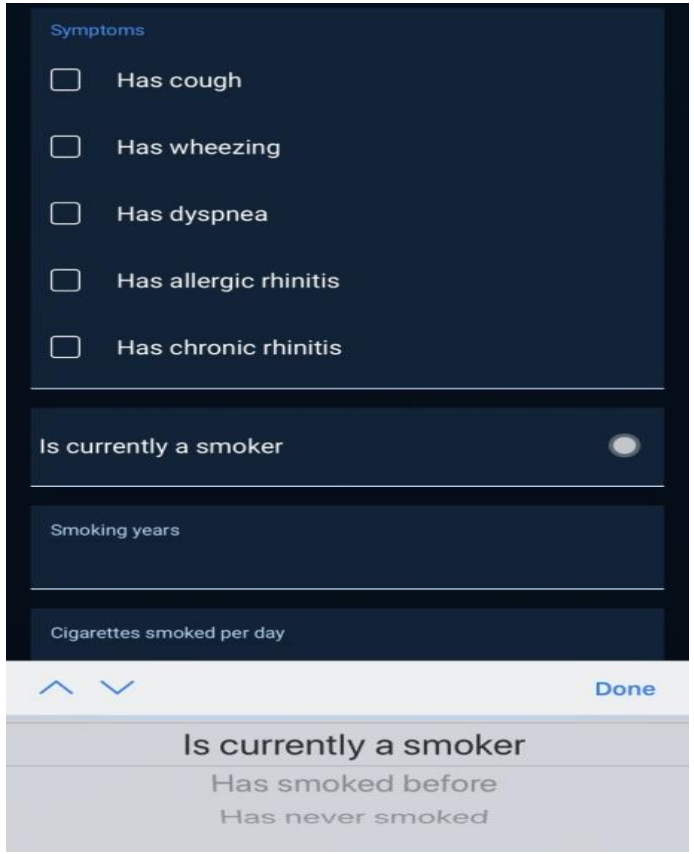


Spirometry Calculations

- Age
- Height
- Weight
- BMI (calculated)
- FEV_1
- FEV_1/FVC ratio

Questions for AC/DC

1. **Dyspnea incidence**
   - Do you have shortness of breath or did you have shortness of breath recently (yes/no)?
2. **Wheeze incidence**
   - Do you have wheezing (a whistling or rattling sound in the chest) or did you wheeze recently (yes/no)?
3. **Cough incidence**
   - Do you cough or did you cough recently (yes/no)?
4. **DX Allergic Rhinitis**
   - Have you been diagnosed with allergic rhinitis (yes/no)?
5. **DX Chronic Rhinitis**
   - Have you been diagnosed with chronic rhinitis (yes/no)?
6. **Smoking History**
   - Are you a current smoker (yes/no)?
   - Have you ever smoked (yes/no)?
     1. If yes, how many years have you smoked cigarettes?
     2. If yes, how many packs of cigarettes have you smoked per day?
   - Calculate Pack Years: packs per day X years smoked



Confusion Matrix: Cohorts I, II, and III

Deep optimization

	precision	recall	f1-score	support
aco	0.92	0.78	0.84	4116
asthma	0.97	0.98	0.98	21562
copd	0.97	0.98	0.98	36057
micro avg	0.97	0.97	0.97	61735
macro avg	0.95	0.91	0.93	61735
weighted avg	0.97	0.97	0.97	61735

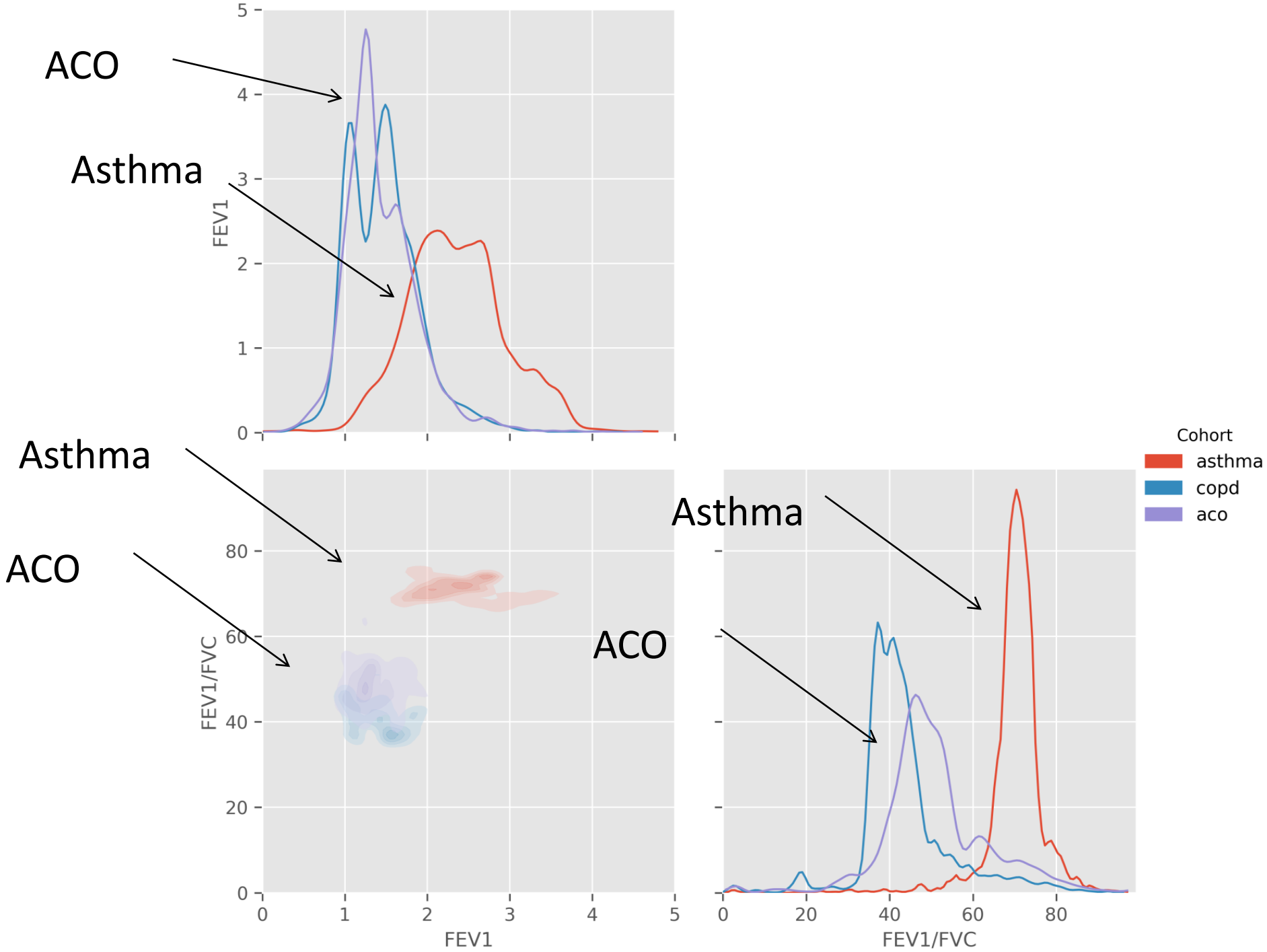
Precision (positive predictive value)

Recall (also known as sensitivity)

F1-score (a measure of a test's accuracy) is the weighted average of Precision and Recall;

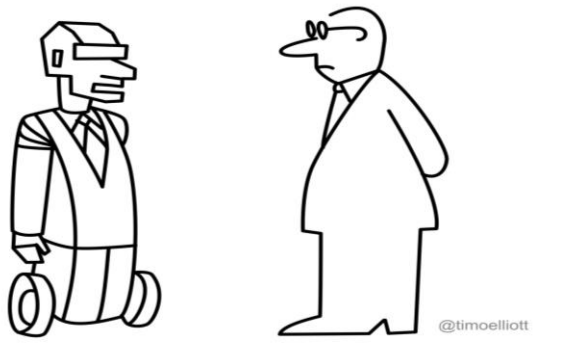
therefore, this score takes both false positives and false negatives into account.

Support = number of patients holdout


ACO
Asthma
Asthma
ACO
Asthma
ACO
FEV1
FEV1/FVC
FEV1
FEV1/FVC
Cohort
asthma
copd
aco

Summary

- Clinicians need help
- AC/DC can provide some **diagnostic assistance**
- Model is limited by what it was taught
  - Some things not available in Optum® database
    - Age of symptom beginning
    - FeNO (infrequently done, so not enough data entered)
- AI complements but does NOT replace clinician input
- Further validation trials being done currently



"The good news is I have discovered inefficiencies. The bad news is that you're one of them."