Using machine learning to screen for autism and other childhood ailments

(Ford Garberson, Cognoa Inc.)

Autism

- About 2-3% of children have autism
  - Severity ranges from mild to completely disabling

- Outcomes much better if you can diagnose early
  - Autism can be diagnosed at 1.5 years old, **but** median age of diagnosis in USA is between 4 to 5 years old depending on region
    - Why so late? Shortage of skilled specialists to examine and diagnose children.
    - Worried parents generally have to wait a full year to be seen by specialists after being referred by a pediatrician

- Cognoa's goal: give parents (and pediatricians) an App, backed by ML to identify if their kids are high risk so that they can be moved to the front of the line

Flagship instruments for autism identification

- Autism Diagnostic Interview ("ADI-R")
  - A questionnaire with 93 multiple choice questions, filled out by an expert clinician based on a few hour interview with parent

- Autism Diagnostic Observation Schedule ("ADOS")
  - A questionnaire with ~30 multiple choice questions, filled out by trained clinicians after interacting with child for ~an hour in a highly standardized setting

- Can we capture some benefits of these instruments in our App ?

Flagship instruments for autism identification

- Autism Diagnostic Interview ("ADI-R")
  - A questionnaire with 93 multiple choice questions, filled out by an expert clinician based on a few hour interview with parent

  What if we have the App just ask the parent a simplified version of of these questions directly?

- Autism Diagnostic Observation Schedule ("ADOS")
  - A questionnaire with ~30 multiple choice questions, filled out by trained clinicians after interacting with child for ~an hour in a highly standardized setting

  What if we have parents record a couple of 1 minute videos of kid playing or eating and have minimally trained clinicians watch them then fill out a questionnaire with simplified versions of these questions?

- Can we capture some benefits of these instruments in our App ?

Where does the machine learning come in?

- Normally instruments just add the severities of each answer together and look at how bad the sum is
  - A well trained model should be able to do better

- The app needs to be fast, easy and cheap
  - If you have to pick only 20 questions to ask, which 20 are best?

- We have plenty of training data available from clinical centers. So should be straightforward, right?
  - Cross validation for a model optimized to minimize training set error: ~90% accurate.
  - Real life with App: probably ~55% accurate. Why so much worse?

Systematic uncertainties/confounding factors (1/2)

- Parental questionnaire:
  - Parents may not understand questions as well as trained clinicians
  - Parents may be emotionally invested and give biased responses

- Video based questionnaire:
  - Technicians who watch videos have much less training than clinicians
  - Short, minimally controlled cell phone videos often do not show real symptoms clearly

- Both:
  - Types of children with/without autism may be very different in the clinical data we use for training than among those who use the App

Systematic uncertainties/confounding factors (2/2)

- These are the most important issues to solve
  - Most challenging part is engineering ML algorithms to be as robust as possible against these kinds of biases

- Example studies:
  - Which questions are most robust against confused responses injected into training data?
  - Which questions are most robust against missing observations being injected into training data?
  - How can algorithms be tweaked to learn more from an observation than from the lack of an observation?
  - What is distribution of other conditions besides autism present in our training data and how does it impact results?
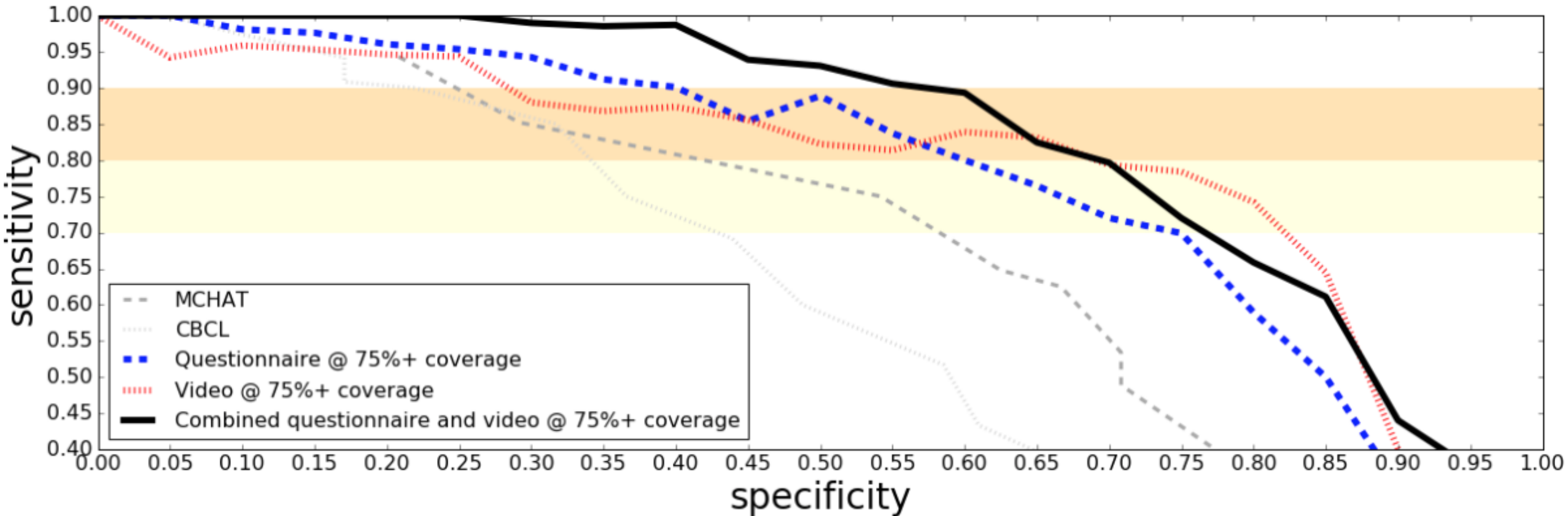
How'd we do?

- Compare with the two main alternative autism screeners in a real clinical trial
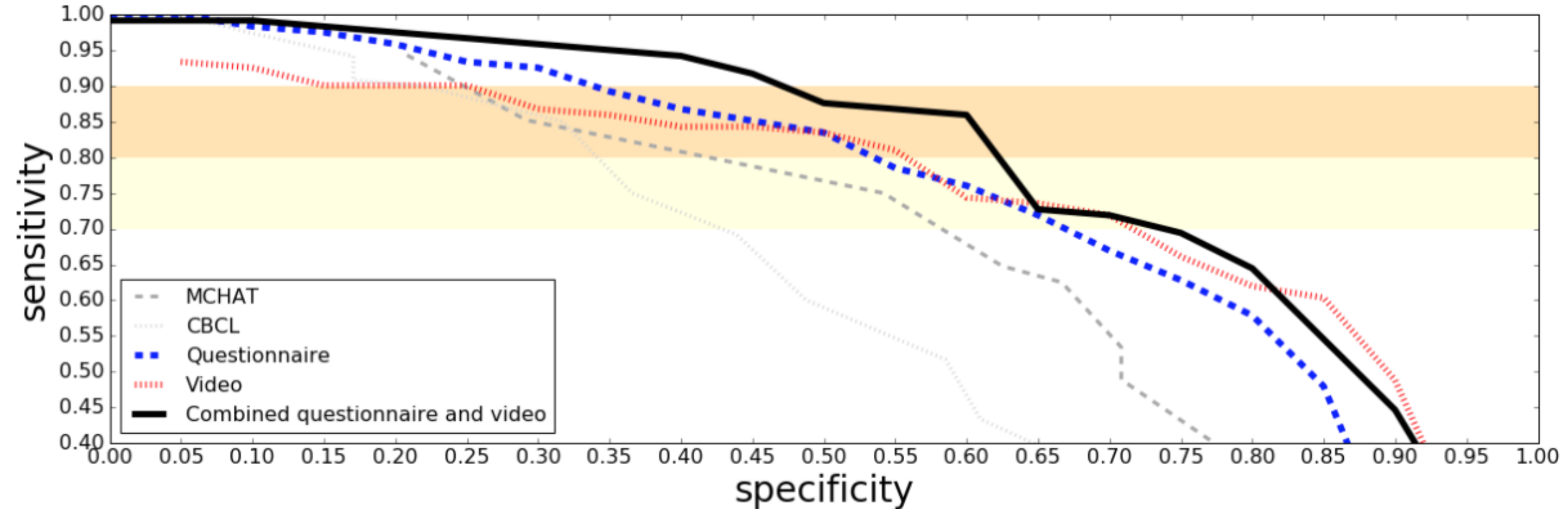
Allow an "inconclusive" determination

- What if we allow up to 25% of the ones we're less sure of to be classified as "inconclusive"?
  - Similar to what doctors normally do

The future

- ML on questionnaires is better any screener out there, but lots of other potential tools
    - ML on photos of child's expression?
    - ML on audio of child's voice?
    - ML on motion of child's movements?
    - ML on data about how child plays tablet games?

- Only ~2% of kids have autism. But about 13% of children have a diagnosable psychological condition
    - ADHD, OCD, severe learning delay, depression/bi-polar, …
    - Expand conditions we screen for

- Improve the activities we are recommending for children and progress tracking

Data science at Cognoa

- A small team working in a fast paced startup environment
  - ML is a deep part our product
  - We regularly submit our results to medical journals
  - Lots of interesting projects to work on

- We're hiring a data scientist who is good at both ML algorithms and independently driving R&D projects
  - https://cognoa.breezy.hr/#positions
  - If this position fills get in touch with me (ford@cognoa.com) or check back later. We intend to have other openings in the future.

- Try out our App:
  - https://www.cognoa.com/parents/dtc