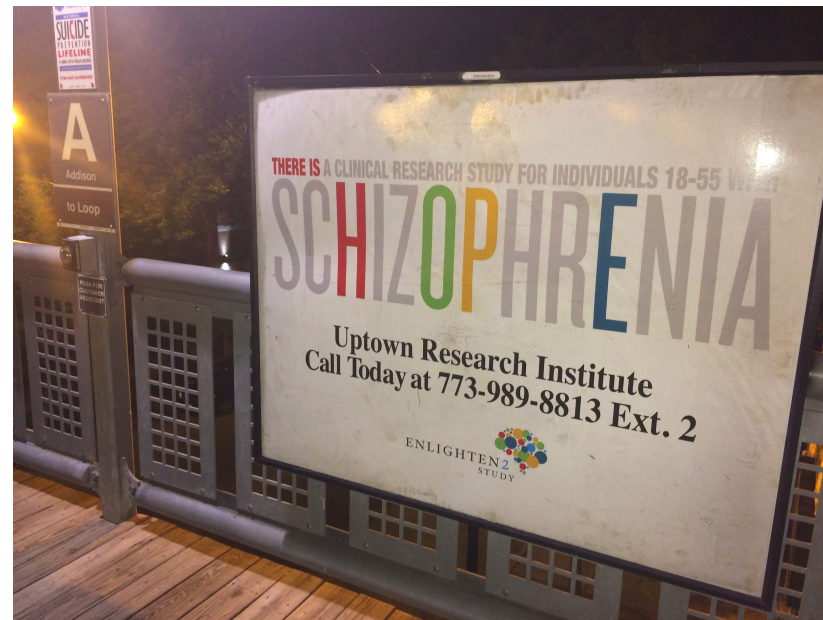# Matching Patients to Clinical Trials

Courtney Stepien

Clinical trial recruitment: largest barrier to success

- Despite heavy spending, up to 2/3rds of clinical trial sites don't enroll enough patients
- This leads to suspensions or termination of trials
- Recruiting is a big business

Patients have a hard time finding relevant trials

Showing: 1-100 of **339** studies [100 ▲▼] studies per page

Row	Saved	Status	Study Title	Conditions	Interventions
1	☐	Recruiting	Pediatric **ASTHMA**-Educator	• **Asthma**	• Other: Pediatric ASTHMA-Educator mobile application
2	☐	Recruiting	Pilot of Lifestyle and **Asthma** Intervention	• **Asthma**	• Behavioral: Weight loss intervention
3	☐	Recruiting **NEW**	Russian Severe **Asthma** Registry	• Severe **Asthma**	
4	☐	Recruiting	Study to Evaluate the Efficacy and Safety of Tezepelumab in Reducing Oral Corticosteroid Use in Adults With Oral Corticosteroid Dependent **Asthma**	• **Asthma**	• Biological: Tezepelumab • Other: Placebo

Business Objectives

- Analyze clinical trial eligibility using natural language processing to identify studies with similar criteria
- Recommend trials based on eligibility criteria of one trial

Data was retrieved from the CTTI database

- The NIH has records of 280,000 trials since the year 2000
- The Clinical Trials Transformation Initiative maintains a database of the trials

Topic modeling was applied to eligibility criteria

- Inclusion criteria was cleaned to remove common words like 'on', 'the', domain words like 'informed', 'consent'
- Split text into 2- and 3-word phrases
  - 'life threatening asthma', 'smoking history'

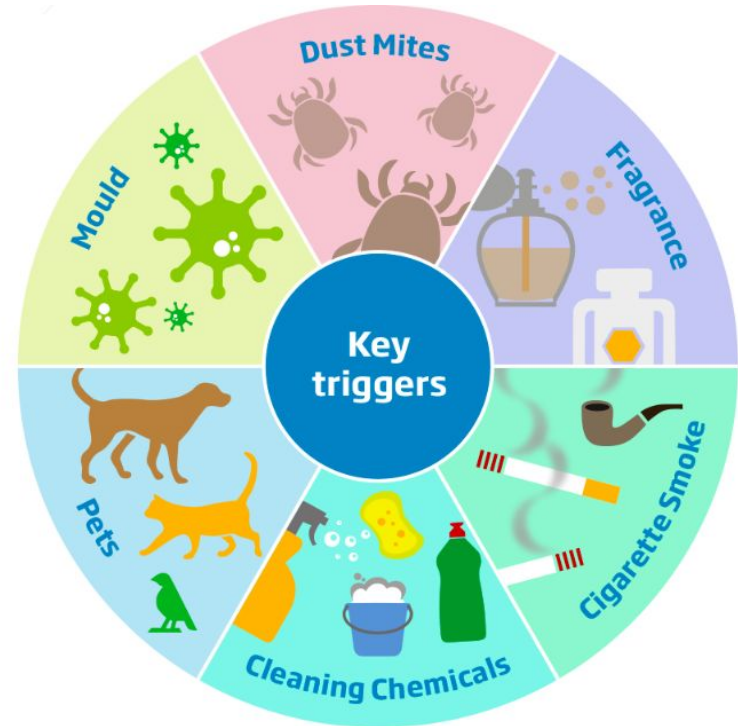# Topic modeling was applied to eligibility criteria

- Inclusion criteria was cleaned to remove common words like 'on', 'the', domain words like 'informed', 'consent'
- Split text into 2- and 3-word phrases
  - 'life threatening asthma', 'smoking history'
- Counted number of appearances of 2- and 3-word phrases in each inclusion criteria
- Modeled topics using Latent Dirichlet Allocation

Topic modeling was applied to eligibility criteria

- Inclusion criteria was cleaned to remove common words like 'on', 'the', domain words like 'informed', 'consent'
- Split text into 2- and 3-word phrases
  - 'life threatening asthma', 'smoking history'
- Counted number of appearances of 2- and 3-word phrases in each inclusion criteria
- Modeled topics using Latent Dirichlet Allocation
- **Topics were too broad/too many for all of medicine!**

Pilot work refocused on a single illness: asthma

- 2100 asthma trials were used for initial work
- Asthma inclusion criteria focuses on descriptors (asthma severity), environmental factors (pollen, mold) and comorbidities (smoking, infection, allergies)

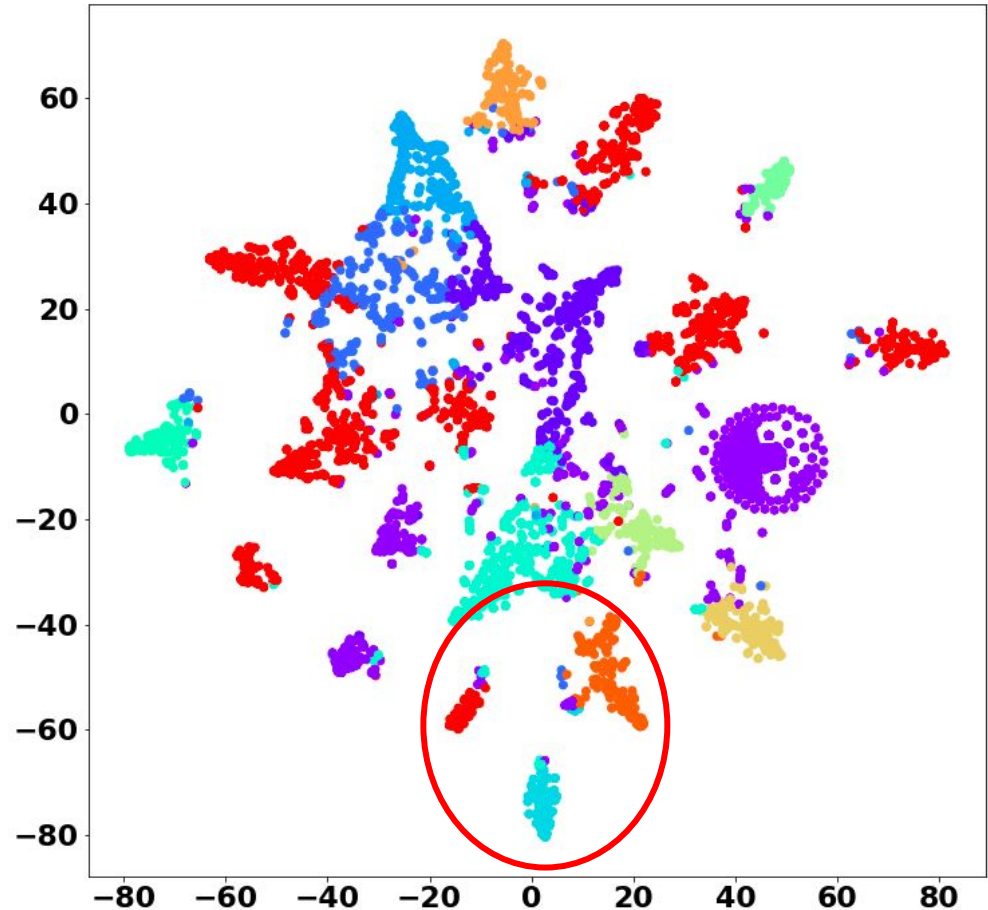# 20 topics gave strongest delineation of 5-20 tested

- **Allergies**: allergic rhinitis, history anaphylaxis, perennial allergic, seasonal allergic, self injectable, severe atopic, chronic sinusitis
- **Severity**: life threatening, asthma episode, respiratory arrest, asthma episode intubation

20 topics gave strongest delineation of 5-20 tested

- **Allergies**: allergic rhinitis, history anaphylaxis, perennial allergic, seasonal allergic, self injectable, severe atopic, chronic sinusitis
- **Severity**: life threatening, asthma episode, respiratory arrest, asthma episode intubation
- **Smoking History:** smoking history, smoking history pack, ex smoker
- **Comorbidities:** heart failure, liver disease, acute myocardial infarction, inflammatory bowel, disease inflammatory bowel, rheumatoid arthritis
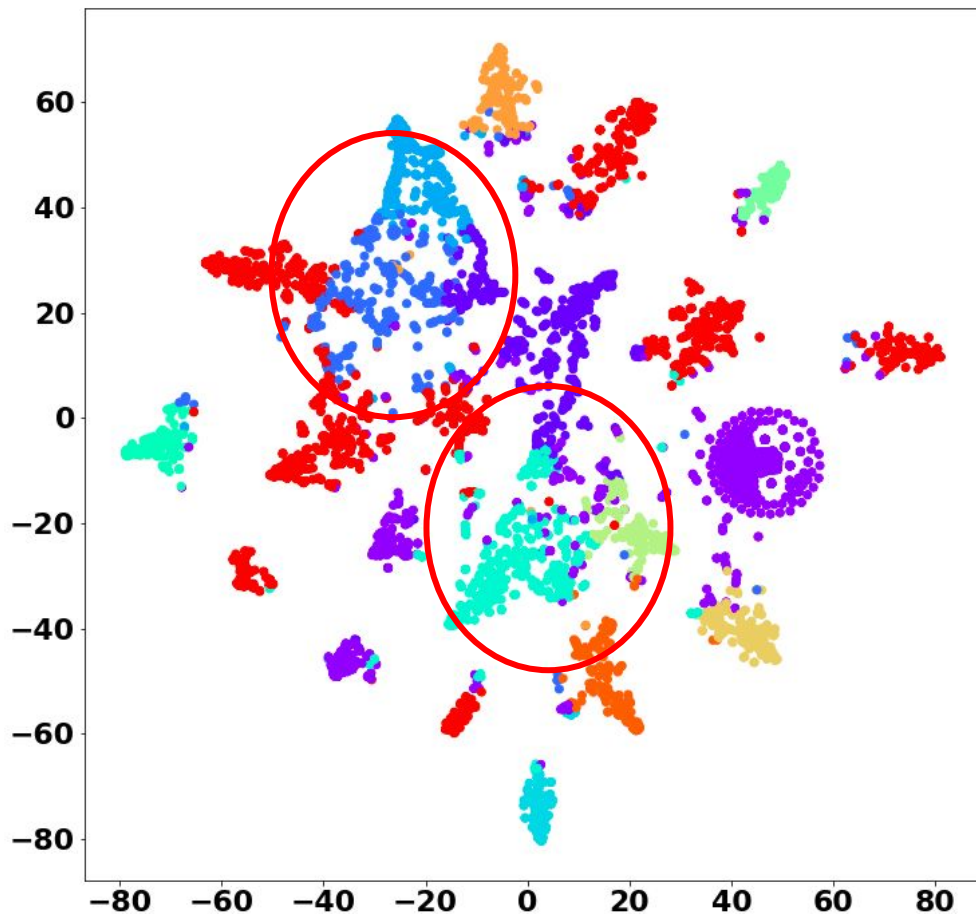
Visualizing eligibility

- Visualization of K-Means clustering shows distinct groups of inclusion criteria

Visualizing eligibility

- Visualization of K-Means clustering shows distinct groups of inclusion criteria
- Some groups have overlapping points - so inclusion criteria topic modeling could be more distinct
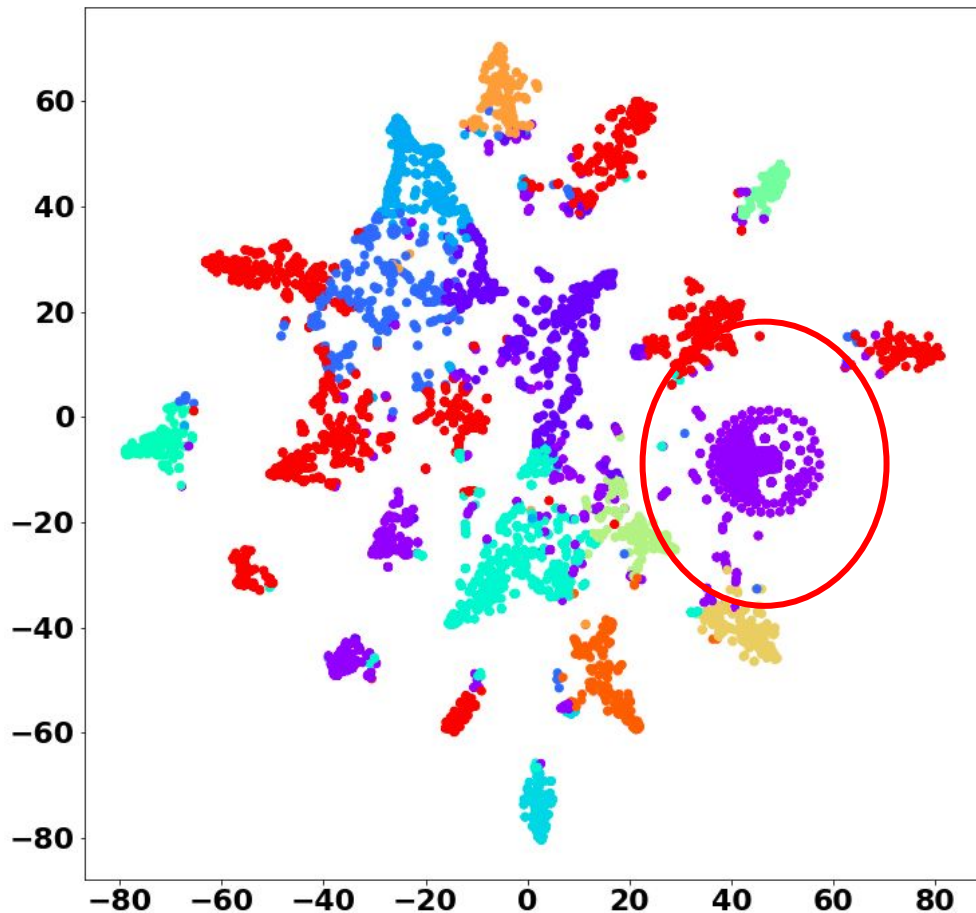
Visualizing eligibility

- Visualization of K-Means clustering shows distinct groups of inclusion criteria
- Some groups have overlapping points - so inclusion criteria topic modeling could be more distinct

Similar studies were identified with nearest neighbors

Input: 'Moderate-severe asthma', 'Stable asthma', '18-75 years old'

Similar studies were identified with nearest neighbors

Input: 'Moderate-severe asthma', 'Stable asthma', '18-75 years old'

Top 3 studies with the most similar inclusion criteria:
1. ['patients eligible for entry into study are the patients diagnosed with moderate to severe asthma, that the doctors have already decided to treat with Symbicort SMART within the last month, before inclusion in this program']
2. ['Patient has mild to moderate asthma', 'Patient is able to swallow pills', 'Patient is able to have have blood draws']
3. ['Moderate to severe asthma with suboptimal control']

Next steps

- Expand number of topics to further delineate comorbidities, age and prior treatment criteria
- Build an application out of the current recommender for patients and recruiters to use
- Return only results from currently recruiting studies
- Expand to other therapeutic areas

Thank you!

Now recruiting:

Row	Saved	Status	Study Title	Conditions	Interventions
1	☐	Recruiting **NEW**	Skydiving as a Model of Psychological Stress and Its Effect on Intestinal Barrier Function	• Psychological Stress Due to Skydiving	• Other: Skydiving
	Locations:		• Örebro University Örebro, Sweden		