# Major Points

## AudGenDB is a large database resource

## Need to develop computational approaches to data analysis

## Provides opportunities

### Large cohort sizes

### Comparison across institutions

### Compendium of data resources to bring to bear on a problem

### Quick testing of hypotheses

## Presents Challenges

### Observational studies

### Sparsity of data

### Significant amounts of data in the form of text, which isn’t readily accessible to computational methods without further work

## Work in Progress

### Demonstrate textmining approaches

### Discuss algorithms to identify progression

### Discuss mathematical modeling to facilitate our analyses

### Touch on how these approaches facilitate the longterm goal of predictive analytics

## HL Progression

### The larger question – EVA is a subset of the problem

### How to determine which patients will progress?

### What computational approaches can we bring to bear on the problem?

### Can machine learning techniques be used to predict HL progression?

#### What features are available for inclusion as features?

#### Can we generate ‘computational phenotypes’?

## HL Progression Intro

### Use Down’s slides

## EVA is an example of both the opportunity and the challenge

### Not identified by a single ICD code

### Textmining techniques required for cohort ascertainment

### However, large data store of radiology reports to mine

## Show textmining slides

## Progression

### Definitions

### Algorithm

### Stats

## Show graphs of samples

## Fluctuation

### Algorithm

### Show examples

### Show prevalence in EVA

## Can we identify characteristic graphs?

### 3-4 types

#### Slow HL progression

#### Fast HL progression

#### Sudden drop

## Show 4th degree polynomial vs. linear

## Show clustering

## Show Cluster 4

## Conclusion

### Work in progress

### Developing computational approaches

#### Ascertain cohorts

#### Characterize these cohorts

### Long term goal

#### Developing features for machine learning

#### Predictive analytic approaches to facilitate identification of patients who will progress