PREDICTION OF DRUG RESPONSE IN THE EHR

TNF-α

- TNF-α is a cytokine involved in the immune response.
- Cytokines are small protein signaling molecules.
- It stimulates inflammation.
- TNF-α is associated with autoimmune disorders such as rheumatoid arthritis.

Anti-TNF-α Agents

- As it is a cytokine, it isn't as easy to inhibit their pathways as it is with smaller signaling molecules.
- These drugs typically use fusion proteins to disable the TNF-α.
- Etanercept, for instance, is a modified antibody that uses the receptor to bind TNF-α.
- As one might imagine, this is expensive.

Studying Anti-TNF agents

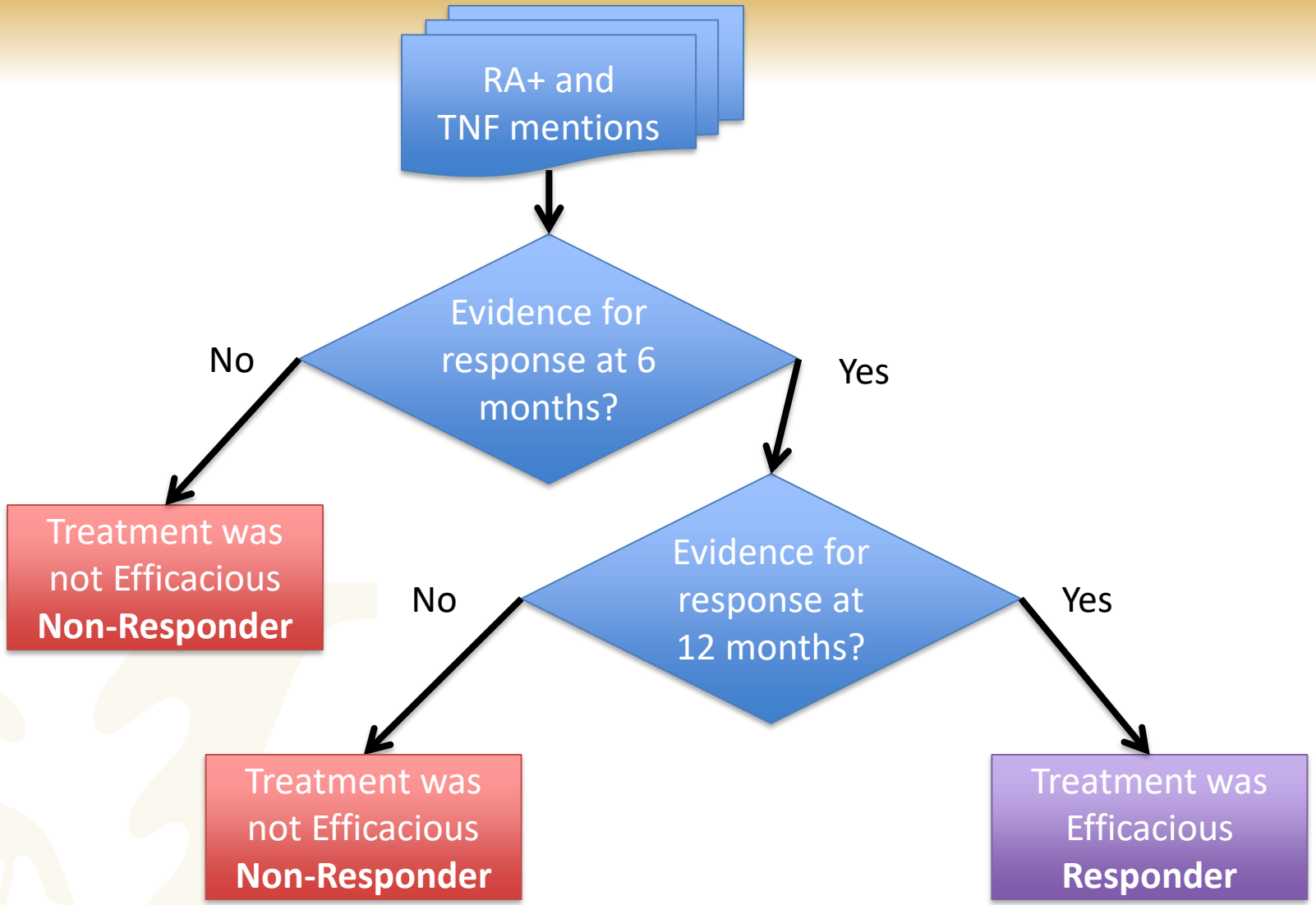
- Why investigate this class of drugs?
  - Expensive
  - Risk of infections
  - Not all individuals respond
- We need to be able to find these individuals in the EHR.
  - These drugs often get mentioned together
  - Difficult to determine response from the clinical record

Synthetic Derivative

- A deidentified version of the EHR
- Most data is available
- Dates are consistently shifted and other identifiers are removed
- Connected to the BioVU repository
- Genotyping results are collected and made available to other researchers

Finding Treated Individuals

- Our first cohort of patients was selected from individuals already reviewed for RA.
- We filtered to those who had a mention of an anti-TNF agent.
- After the first set of reviews, we found a Cohen's kappa of 0.56.
- On reconciling the reviews not in agreement, a more defined plan was established.


RA+ and
TNF mentions
Evidence for response at 6 months?
No
Yes
Treatment was not Efficacious
Non-Responder
Evidence for response at 12 months?
No
Yes
Treatment was not Efficacious
Non-Responder
Treatment was Efficacious
Responder

Records with no label

- We identified some records for which could not be classified with certainty
  - Evidence for treatment
  - May have been treated at another site
  - May have been treated in the past
  - May be inconclusive due to complications
- Reviewers labeled them as “unsure”
  - We trained models including and excluding these individuals
  - They meet our selection criteria

TNF reviews

Reviewer	Total	Responder	Non-Responder	Unsure
Original set (Reviewer 1)	**55**	43	12	0
Reviewer 1	**158**	105	25	28
Reviewer 2	**102**	64	13	25
Reviewer 3	**14**	3	9	2
Totals	**329**	**215**	**59**	**55**

Data Formulation

- Similar to methods for RA identification
- Focused our free text methods around drug mentions
  - Knowledge Map Concept Identifier
  - Ngrams
- Counts of drug mentions
  - Measures of estimated duration
  - Counts of prescriptions of other agents after the one under study
- Counts of PheWAS codes

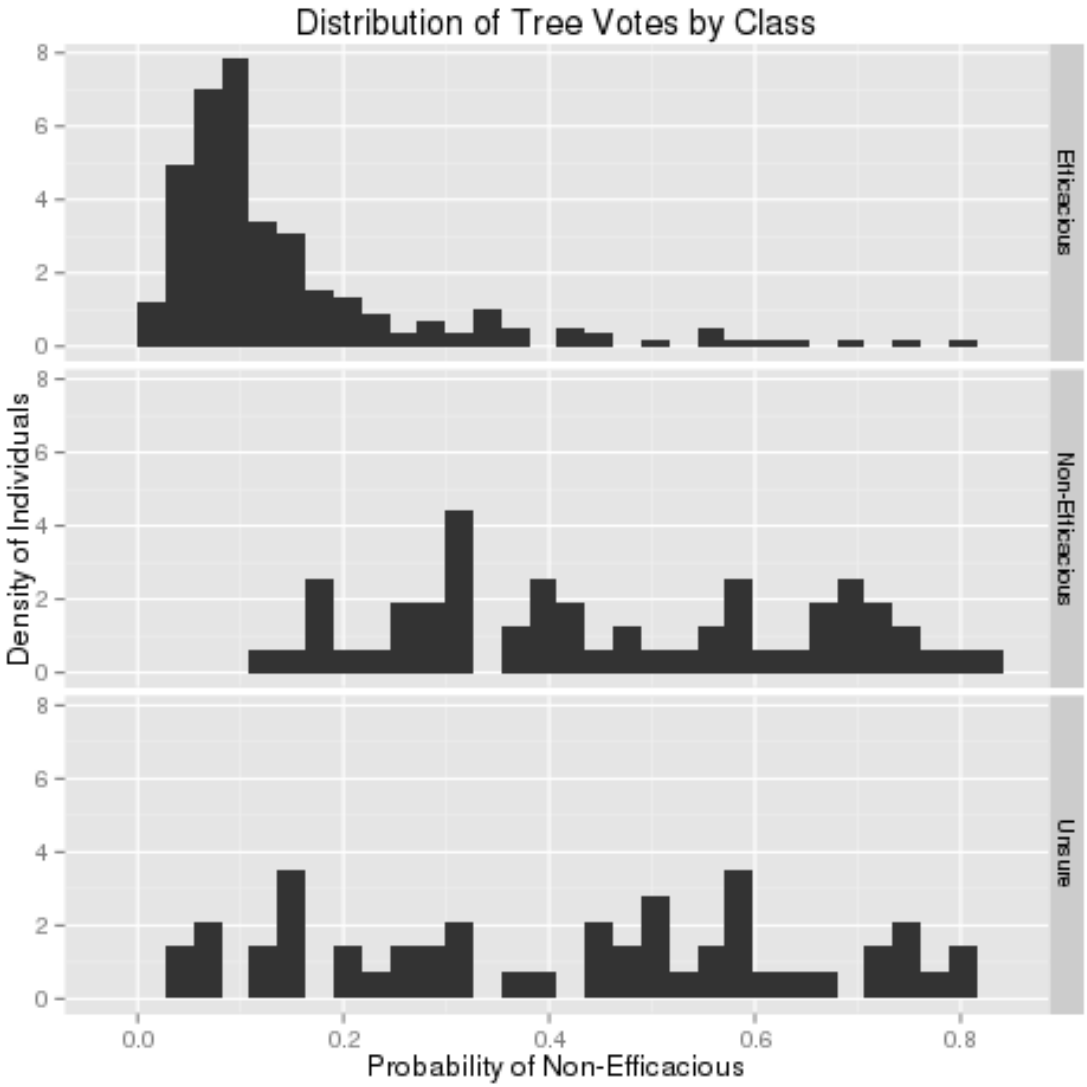
Predictive Models

- Support Vector Machines
  - Trained using cross validation
  - Included t-test based feature selection
  - Nested cross validation for hyperparameters
- Random Forests
  - 500 trees per forest
  - No feature selection
  - Default attributes per tree

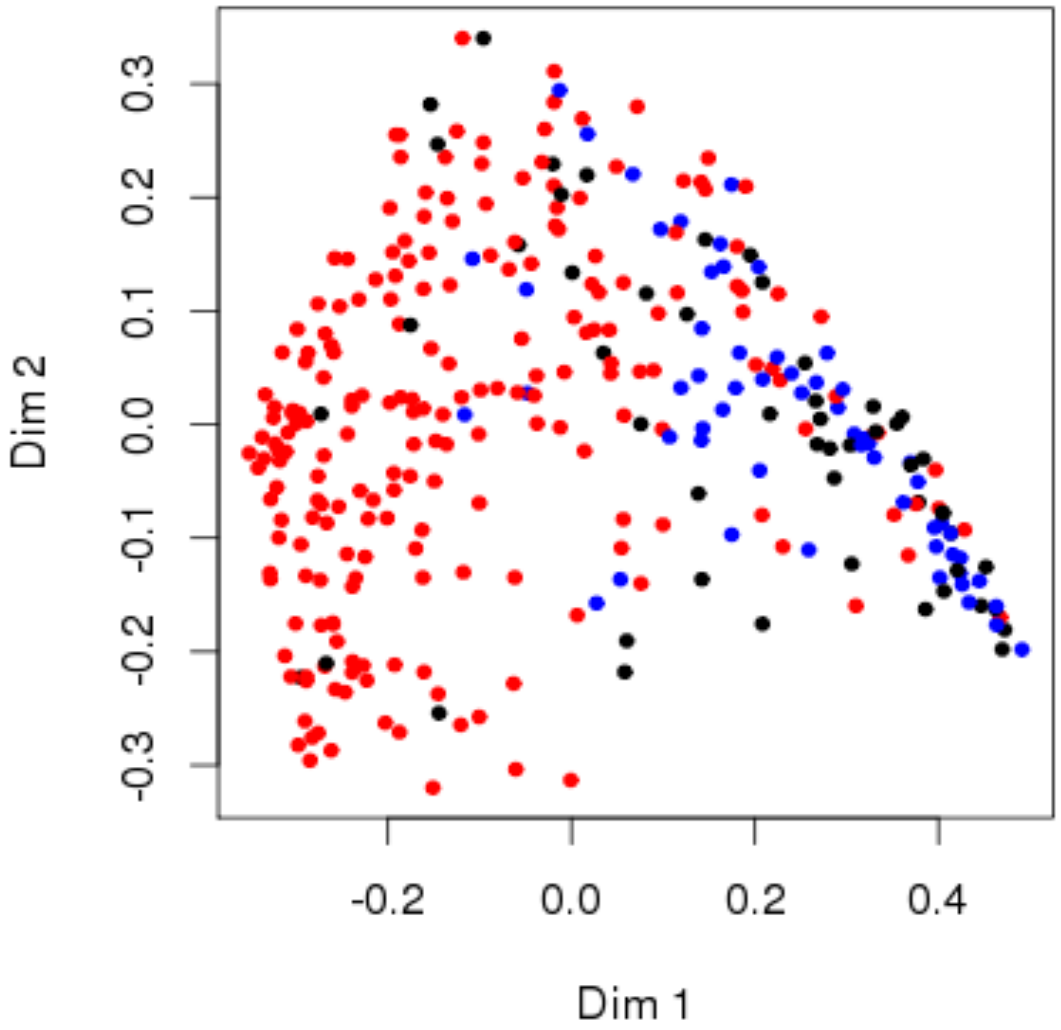
Two Class Results - AUC

Data Class	Data Types Included	SVM	RF
Combined	All available	93.76%	**92.38%**
	Ngrams, CUIs, Meds	**93.85%**	91.54%
	Ngrams, PheWAS, Meds	93.29%	91.66%
	Ngrams, Meds	92.96%	91.22%
	CUIs, Medications	87.34%	90.93%
	PheWAS, Medications	79.60%	84.79%
Free Text	Ngrams	**92.93%**	**91.96%**
	CUIs	88.17%	90.37%
PheWAS Codes	PheWAS code counts	57.09%	59.42%
Medications	Prescription counts	70.33%	77.31%
	Med summary measures	75.35%	84.71%
Demographics	Demographics only	48.39%	32.83%

Unsures?


Distribution of Tree Votes by Class
Density of Individuals
Probability of Non-Efficacious
Efficacious
Non-Efficacious
Unsure
0
2
4
6
8
0.0
0.2
0.4
0.6
0.8

Similarity of Records



Multiclass Prediction

		Non Efficacious		Efficacious		Unsure	
Data Class	Data Types Included	SVM	RF	SVM	RF	SVM	RF
Combined	All available	86.67%	85.76%	88.63%	87.78%	73.50%	71.35%
	Ngrams, CUIs, Meds	87.91%	85.14%	88.94%	88.71%	72.82%	70.45%
	Ngrams, PheWAS, Meds	87.78%	84.80%	88.46%	88.57%	72.04%	73.74%
	Ngrams, Meds	86.64%	85.53%	88.39%	88.00%	72.88%	72.11%
	CUIs, Medications	82.50%	84.86%	86.38%	86.52%	73.25%	70.04%
	PheWAS, Medications	75.93%	79.07%	82.62%	83.57%	73.92%	70.78%
Free Text	Ngrams	87.82%	84.79%	89.43%	87.67%	73.31%	71.25%
	CUIs	84.44%	85.11%	83.97%	85.40%	68.97%	67.50%
PheWAS Codes	PheWAS code counts	56.03%	59.26%	55.52%	52.90%	55.37%	44.45%
Medications	Prescription counts	68.11%	71.00%	73.13%	77.34%	67.74%	65.93%
	Med summary measures	73.89%	78.61%	81.63%	83.93%	74.22%	73.36%
Demographics	Demographics only	53.45%	35.76%	49.00%	40.93%	46.58%	52.40%

Multiclass Prediction

		Non Efficacious		Efficacious		Unsure	
Data Class	Data Types Included	SVM	RF	SVM	RF	SVM	RF
Combined	All available	86.67%	85.76%	88.63%	87.78%	73.50%	71.35%
	Ngrams, CUIs, Meds	87.91%	85.14%	88.94%	88.71%	72.82%	70.45%
	Ngrams, PheWAS, Meds	87.78%	84.80%	88.46%	88.57%	72.04%	73.74%
	Ngrams, Meds	86.64%	85.53%	88.39%	88.00%	72.88%	72.11%
	CUIs, Medications	82.50%	84.86%	86.38%	86.52%	73.25%	70.04%
	PheWAS, Medications	75.93%	79.07%	82.62%	83.57%	73.92%	70.78%
Free Text	Ngrams	87.82%	84.79%	89.43%	87.67%	73.31%	71.25%
	CUIs	84.44%	85.11%	83.97%	85.40%	68.97%	67.50%
PheWAS Codes	PheWAS code counts	56.03%	59.26%	55.52%	52.90%	55.37%	44.45%
Medications	Prescription counts	68.11%	71.00%	73.13%	77.34%	67.74%	65.93%
	Med summary measures	73.89%	78.61%	81.63%	83.93%	74.22%	73.36%
Demographics	Demographics only	53.45%	35.76%	49.00%	40.93%	46.58%	52.40%

Multiclass Prediction

		Non Efficacious		Efficacious		Unsure	
Data Class	Data Types Included	SVM	RF	SVM	RF	SVM	RF
Combined	All available	86.67%	85.76%	88.63%	87.78%	73.50%	71.35%
	Ngrams, CUIs, Meds	87.91%	85.14%	88.94%	88.71%	72.82%	70.45%
	Ngrams, PheWAS, Meds	87.78%	84.80%	88.46%	88.57%	72.04%	73.74%
	Ngrams, Meds	86.64%	85.53%	88.39%	88.00%	72.88%	72.11%
	CUIs, Medications	82.50%	84.86%	86.38%	86.52%	73.25%	70.04%
	PheWAS, Medications	75.93%	79.07%	82.62%	83.57%	73.92%	70.78%
Free Text	Ngrams	87.82%	84.79%	89.43%	87.67%	73.31%	71.25%
	CUIs	84.44%	85.11%	83.97%	85.40%	68.97%	67.50%
PheWAS Codes	PheWAS code counts	56.03%	59.26%	55.52%	52.90%	55.37%	44.45%
Medications	Prescription counts	68.11%	71.00%	73.13%	77.34%	67.74%	65.93%
	Med summary measures	73.89%	78.61%	81.63%	83.93%	74.22%	73.36%
Demographics	Demographics only	53.45%	35.76%	49.00%	40.93%	46.58%	52.40%

Multiclass Prediction

		Non Efficacious		Efficacious		Unsure	
Data Class	Data Types Included	SVM	RF	SVM	RF	SVM	RF
Combined	All available	86.67%	85.76%	88.63%	87.78%	73.50%	71.35%
	Ngrams, CUIs, Meds	87.91%	85.14%	88.94%	88.71%	72.82%	70.45%
	Ngrams, PheWAS, Meds	87.78%	84.80%	88.46%	88.57%	72.04%	73.74%
	Ngrams, Meds	86.64%	85.53%	88.39%	88.00%	72.88%	72.11%
	CUIs, Medications	82.50%	84.86%	86.38%	86.52%	73.25%	70.04%
	PheWAS, Medications	75.93%	79.07%	82.62%	83.57%	73.92%	70.78%
Free Text	Ngrams	87.82%	84.79%	89.43%	87.67%	73.31%	71.25%
	CUIs	84.44%	85.11%	83.97%	85.40%	68.97%	67.50%
PheWAS Codes	PheWAS code counts	56.03%	59.26%	55.52%	52.90%	55.37%	44.45%
Medications	Prescription counts	68.11%	71.00%	73.13%	77.34%	67.74%	65.93%
	Med summary measures	73.89%	78.61%	81.63%	83.93%	74.22%	73.36%
Demographics	Demographics only	53.45%	35.76%	49.00%	40.93%	46.58%	52.40%

Multiclass Prediction

		Non Efficacious		Efficacious		Unsure	
Data Class	Data Types Included	SVM	RF	SVM	RF	SVM	RF
Combined	All available	86.67%	85.76%	88.63%	87.78%	73.50%	71.35%
	Ngrams, CUIs, Meds	87.91%	85.14%	88.94%	88.71%	72.82%	70.45%
	Ngrams, PheWAS, Meds	87.78%	84.80%	88.46%	88.57%	72.04%	73.74%
	Ngrams, Meds	86.64%	85.53%	88.39%	88.00%	72.88%	72.11%
	CUIs, Medications	82.50%	84.86%	86.38%	86.52%	73.25%	70.04%
	PheWAS, Medications	75.93%	79.07%	82.62%	83.57%	73.92%	70.78%
Free Text	Ngrams	87.82%	84.79%	89.43%	87.67%	73.31%	71.25%
	CUIs	84.44%	85.11%	83.97%	85.40%	68.97%	67.50%
PheWAS Codes	PheWAS code counts	56.03%	59.26%	55.52%	52.90%	55.37%	44.45%
Medications	Prescription counts	68.11%	71.00%	73.13%	77.34%	67.74%	65.93%
	Med summary measures	73.89%	78.61%	81.63%	83.93%	74.22%	73.36%
Demographics	Demographics only	53.45%	35.76%	49.00%	40.93%	46.58%	52.40%

Lessons Learned

- Unclassifiable records need attention in the future.
- Narrowing the NLP used to be more specific to the topic under study was beneficial.
- Ngrams vs. CUIs
- Problem list occurrences were very unreliable.
- Chart reviews can be an important supplement to machine learning techniques.

APPLICATION OF DRUG RESPONSE PREDICTION METHODS AND SECONDARY ANALYSIS

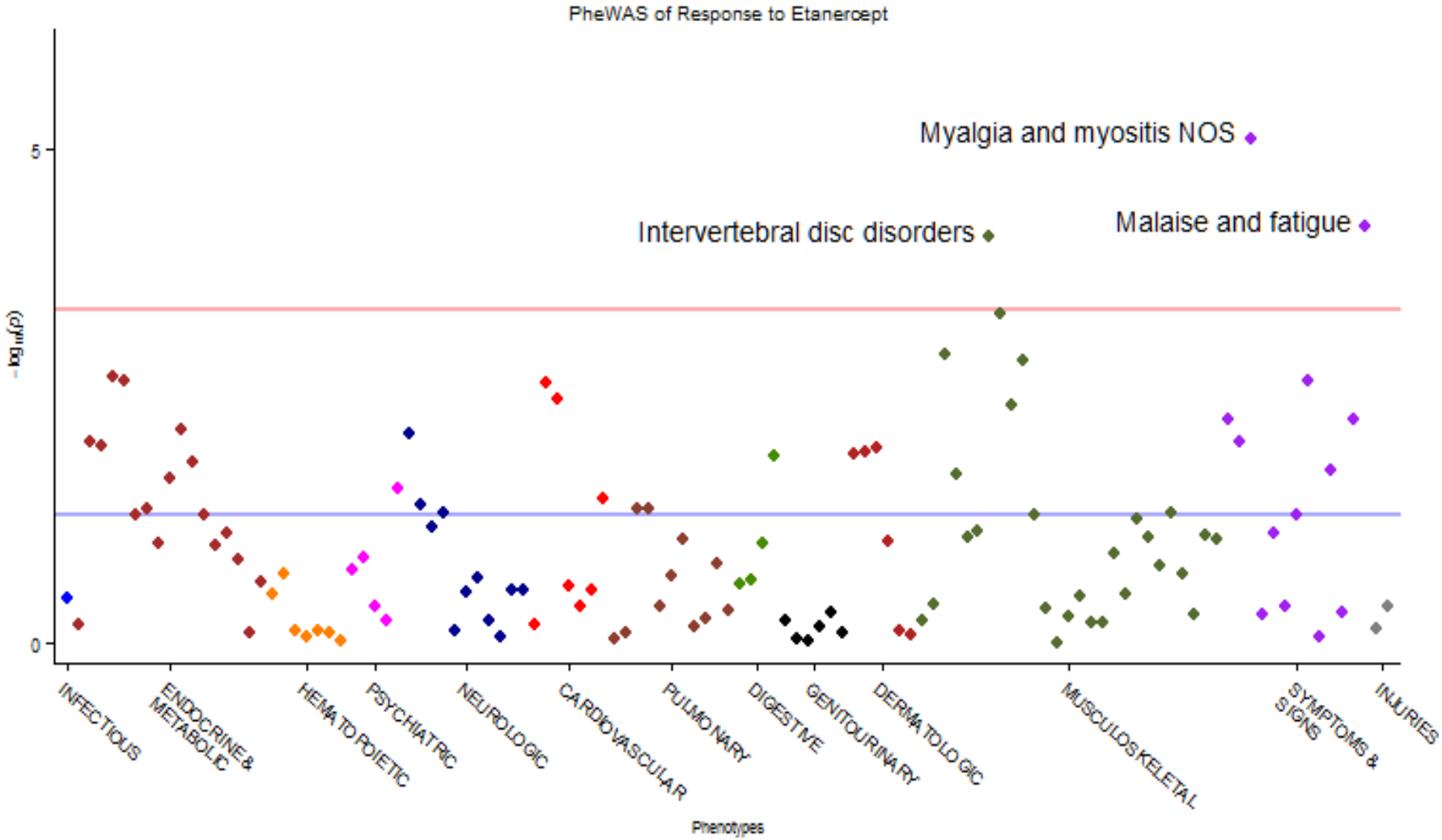
Data and Methods

- Synthetic Derivative
- RA phenotyping algorithm
- Anti-TNF response prediction algorithm
- PheWAS

Anti-TNF Response Predictions

- 752 individuals were identified that met the RA+ status and etanercept mention criteria.
- The RF algorithm was set to predict with more of a bias towards non-efficacious individuals.

		SVM		
		Efficacious	Non-Efficacious	Unsure
RF	Efficacious	**523**	1	1
	Non-Efficacious	63	**118**	1
	Unsure	4	8	33


PheWAS of Response to Etanercept
Myalgia and myositis NOS
Intervertebral disc disorders
Malaise and fatigue
5
0
INFECTIOUS
ENDOCRINE & METABOLIC
HEMATOPOIETIC
PSYCHIATRIC
NEUROLOGIC
CARDIOVASCULAR
PULMONARY
DIGESTIVE
GENITOURINARY
DERMATOLOGIC
MUSCULOSKELETAL
SYMPTOMS & SIGNS
INJURIES
Phenotypes

Top 5 PheWAS Associations

PheWAS Phenotype	OR	p	Cases	Controls
Myalgia and myositis NOS	3.09	7.90E-06	113	457
Malaise and fatigue	2.51	6.26E-05	230	309
Intervertebral disc disorders	3.30	7.76E-05	66	490
Degeneration of intervertebral disc	3.11	4.67E-04	56	490
Spinal stenosis	3.37	1.23E-03	41	490

Disc Disorders

- RA is not traditionally associated with axial skeleton disease.
- It can affect some cervical vertebrae.
- The strong association with these diseases could have several sources:
  - Poor disease control
  - Increased dosage over time of corticosteroids
  - A sub-phenotype of RA that does not respond to etanercept treatment and has axial skeleton involvement

Myalgia and myositis NOS

- Very generic code, but could be due to Fibromyalgia.
- Fibromyalgia individuals have been shown to have elevated TNF-alpha levels, and TNF-alpha plays a role in the inflammation and pain networks.
- It is feasible that individuals with elevated TNF-alpha are not managed with routine anti-TNF treatment.
- There is no determination of causality here.