INFORMATICS PROFESSIONALS. LEADING THE WAY.

Data-driven automated classification algorithms for acute health conditions: Applying PheNorm to COVID-19 disease

S39: Oral Presentations - Innovations in Informatics

Joshua C. Smith, PhD

Vanderbilt University Medical Center

#AMIA2022

Disclosure

I and my spouse/partner have no relevant relationships with commercial interests to disclose.

Disclaimer

- This project was supported by Task Order 75F40119F19002 under Master Agreement 75F40119D10037 from the U.S. Food and Drug Administration (FDA).
- The views expressed in this presentation represent those of the presenter and do not necessarily represent the official views of the U.S. FDA.

Learning Objectives

After participating in this session, the learner should be better able to:

- Understand the development and implementation of portable, automated phenotyping algorithms for use in post-market safety studies.

Introduction

- Sentinel is the U.S. FDA's medical product safety surveillance system utilizing electronic healthcare and claims data.
- One of the goals of the **Sentinel Innovation Center** is to develop, implement, and evaluate methods that incorporate unstructured EHR data to improve the performance of computable phenotype algorithms used to capture health outcomes relevant to medical product safety surveillance.
- In this study, we evaluated an **automated phenotyping** method (PheNorm) applied to an acute condition, COVID-19 disease, to investigate its feasibility for rapid phenotyping and use in post-market safety studies.

Rationale for automating phenotyping

Manual development

- *Expert*-driven
- *Manual* engineering
- Heavy reliance on *gold standard labels*
- Substantial operator dependence
- Slow

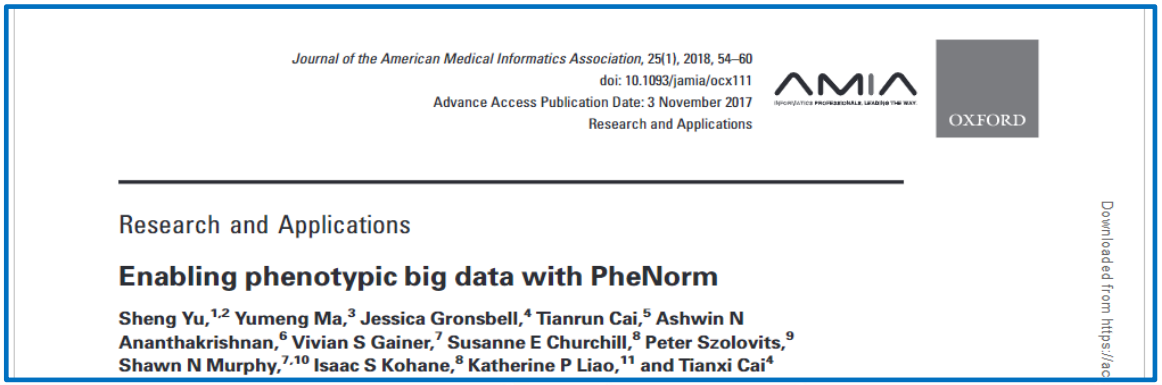
Automated development

- Data-driven
- Automated engineering
- Heavy reliance on silver standard labels
- Reduced operator dependence
- Fast

- Automated feature engineering (AFEP)
  Yu et al. Toward high-throughput phenotyping: unbiased automated feature extraction and selection from knowledge sources. JAMIA 2015.
- Surrogate-assisted feature extraction (SAFE)
  Yu et al. Surrogate-assisted feature extraction for high-throughput phenotyping. JAMIA 2017.
- Phenotype algorithm normalization (PheNorm)
  Yu et al. Enabling phenotypic big data with PheNorm. JAMIA 2018.
- Phenotyping common approach (PheCAP)
  Zhang et al. High-throughput phenotyping with EMR data using a common semi-supervised approach (PheCAP). Nature Protocols. 2019.

Introduction

- **PheNorm** is a general-purpose automated approach to creating computable phenotype algorithms based on natural language processing (NLP), machine learning, and (low- cost) silver-standard training labels.
- It has been demonstrated to perform well outside Sentinel for chronic health conditions, but little is known about its performance in acute conditions.
- https://pubmed.ncbi.nlm.nih.gov/29126253/



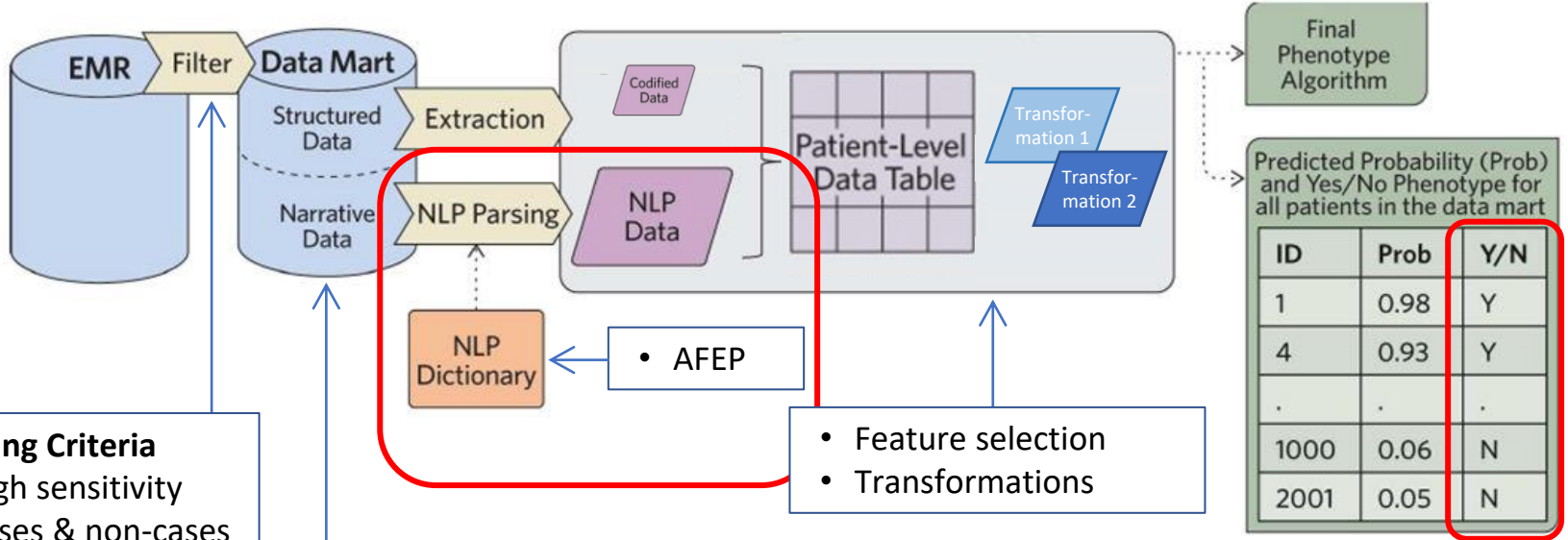
Journal of the American Medical Informatics Association, 25(1), 2018, 54–60
doi: 10.1093/jamia/ocx111
Advance Access Publication Date: 3 November 2017
Research and Applications

Research and Applications

Enabling phenotypic big data with PheNorm

Sheng Yu,[1,2] Yumeng Ma,[3] Jessica Gronsbell,[4] Tianrun Cai,[5] Ashwin N Ananthakrishnan,[6] Vivian S Gainer,[7] Susanne E Churchill,[8] Peter Szolovits,[9] Shawn N Murphy,[7,10] Isaac S Kohane,[8] Katherine P Liao,[11] and Tianxi Cai[4]

PheNorm Overview



Filtering Criteria
- High sensitivity
- Cases & non-cases

- Feature selection
- Transformations

- Catchment period
- Define silver label
- Minimal struct. data
- Heavy reliance on NLP-extracted data

- Predicted probability of being a case
- Phenotype classification (Yes/No)

- Evaluate using gold-labeled data (not included)

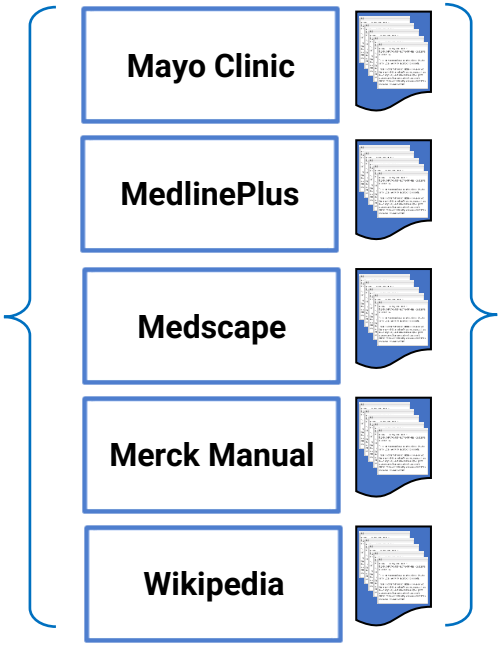
Methods

- This study was performed at Vanderbilt University Medical Center (VUMC) and Kaiser Permanente Washington (KPWA).
- We identified cohorts of potential COVID-19 patients from 4/2020-3/2021 at each site.
- Cohorts included all patients with encounters accompanied by structured EHR features found to be strongly associated with COVID-19, including diagnoses, problems, procedures, medications, and lab tests (described elsewhere). Each patient's earliest such encounter was used as index date.
- The VUMC cohort included both inpatient and outpatient encounters; the KPWA cohort included outpatient only.

Methods – NLP Dictionary Creation

Automating NLP dictionary creation (AFEP)

5 clinical knowledge base articles on a topic (*COVID-19*)



Corpus

MetaMap Lite NLP

	Source	CUI_Code	Term
1	SNOMEDCT_US	C0663655	abacavir
2	SNOMEDCT_US	C0000726	Abdomen
3	SNOMEDCT_US	C1122087	adalimumab
4	SNOMEDCT_US	C0001443	Adenosine
5	SNOMEDCT_US	C3536832	Air
6	SNOMEDCT_US	C0001927	Albuterol
7	SNOMEDCT_US	C0002055	Alkalies
8	SNOMEDCT_US	C0002092	Allergens
9	SNOMEDCT_US	C0002508	Amines
10	SNOMEDCT_US	C0002575	Aminophylline
11	SNOMEDCT_US	C0002667	Amphetamines
12	SNOMEDCT_US	C0002771	Analgesics
13	SNOMEDCT_US	C0002792	anaphylaxis
14	SNOMEDCT_US	C0002932	Anesthetics
15	SNOMEDCT_US	C0002994	Angioedema
16	SNOMEDCT_US	C0003018	Angiotensins
17	SNOMEDCT_US	C0003232	Antibiotics
18	SNOMEDCT_US	C0003241	Antibodies
19	SNOMEDCT_US	C0003320	Antigens
20	SNOMEDCT_US	C0003360	Antihistamines
21	SNOMEDCT_US	C0003445	Antitoxins
22	SNOMEDCT_US	C0003450	Antivenin
23	SNOMEDCT_US	C0003467	Anxiety
24	SNOMEDCT_US	C0003483	Aorta
25	SNOMEDCT_US	C0003564	Aphonia
26	SNOMEDCT_US	C0233485	apprehension
27	SNOMEDCT_US	C0003842	Arteries
28	SNOMEDCT_US	C0004044	Asphyxia
29	SNOMEDCT_US	C0004057	Aspirin
30	SNOMEDCT_US	C1510438	Assay
31	SNOMEDCT_US	C0004096	Asthma
32	SNOMEDCT_US	C0231221	Asymptomatic
33	SNOMEDCT_US	C0392707	Atopy
34	SNOMEDCT_US	C0004259	Atropine
35	SNOMEDCT_US	C0004268	Attention
36	SNOMEDCT_US	C0004271	Attitude
37	SNOMEDCT_US	C0004398	Autopsy
38	SNOMEDCT_US	C0004521	Aztreonam
39	SNOMEDCT_US	C0004827	Basophils
40	SNOMEDCT_US	C0005558	Biopsy
41	SNOMEDCT_US		

295 candidate UMLS Concepts (CUIs) appeared in *≥3 articles*

***158 CUIs** retained for the dictionary after manual review*

Yu et al. Toward high-throughput phenotyping: unbiased automated feature extraction and selection from knowledge sources. JAMIA 2015

Methods – Running PheNorm

- **Data/text catchment period**
  - Index date +/-30 days from index date
- **Input Data – Notes processed using MetaMap Lite**
  - KPWA: 143,584 notes from 8,329 patients
  - VUMC: Approximately 1.1 million notes from 24,304 patients
- **AFEP-Generated NLP Dictionary and Corresponding Features**
  - 158 CUIs extracted from five articles on COVID-19 yielding one feature per CUI

•Silver Labels

1. Structured Label – U07.1 Days
2. Structured Label – Six-ICD-Code Days (U07.1, J12.81, J12.82, B34.2, B97.21, B97.29)
3. NLP Label – Cumulative count mentions of COVID-19 in patients' charts
4. NLP Label – COVID-19 CUI Days (KPWA) or CUI Notes (VUMC)

Methods – Evaluation & Outcomes

- We used manual chart review to assign gold-standard labels for both phenotypes for stratified random samples of **483 VUMC** and **437 KPWA** patients.
- Subjects were initially reviewed by two reviewers at each site to assess inter-rater reliability (**kappa 0.951 at VUMC and 0.802 at KPWA**); subsequent reviews were performed by only one reviewer.
- We evaluated PheNorm performance at both sites on **two COVID-19 phenotype definitions** based on National Institutes of Health COVID-19 Treatment Guidelines:
  - **Symptomatic COVID-19 disease (mild or greater severity)**
  - **COVID-19 disease with at least moderate severity**

Methods – Evaluation & Outcomes

Evidence of COVID-19 infection

- **Definite or highly probable infection**
  - PCR-positive or explicit positive assertion
- **Probable or possible infection**
  - Symptoms are consistent with a diagnosis of COVID-19 and absence of an explicit *alternative* diagnosis
- **Unlikely infection**
  - Explicit *alternative* diagnosis or statement ruling-out COVID-19 and absence of relevant symptoms/labs
- ***Not infected***
  - No indication in the EHR of infection
- **Insufficient Information**

Severity of illness scale (NIH)

SEVERITY LEVEL	SIGN/SYMPTOM
Asymptomatic	No symptoms
Mild	Fever (>=100.4F)
	Cough
	Sore throat
	Malaise/fatigue
	Headache
	Muscle pain
	Nausea
	Vomiting
	Diarrhea
	Loss of sense of taste or smell
Moderate	Shortness of breath (SpO2 >=94%)
	Dyspnea (SpO2 >=94%)
	Abnormal chest imaging (SpO2 >=94%)
Severe	SpO2 <94%
	PaO2/FiO2* <300 mm Hg
	Respiratory freq >30 breaths/min
	Lung infiltrates >50%
Critical	Respiratory failure
	Septic shock
	Multiple organ dysfunction

Results

Sample demographics by Study Site.

	VUMC		KPWA	
	Patients	Percent	Patients	Percent
Gender Female				
no	10216	42%	3492	42%
yes	14088	58%	4837	58%
Ethnicity Hispanic				
no	23283	96%	7573	91%
yes	1021	4%	756	9%
Race White				
no	7840	32%	2994	36%
yes	16464	68%	5335	64%
Age Range				
18-29	5672	23%	1104	13%
30-49	8196	34%	2503	30%
50-69	7465	31%	3126	38%
70+	2971	12%	1596	19%
Total	**24304**	**100%**	**8329**	**100%**

Results

Gold standard chart review results by study site and COVID-19 phenotype definition				
Study site	**COVID-19 phenotype definition**	**Chart review result**	**Number of charts**	**Percent of charts**
VUMC (N=483)	Moderate+ severity	Non-case	334	69%
		Case	149	**31%**
	Mild+ severity	Non-case	188	39%
		Case	295	**61%**
KPWA (N=437)	Moderate+ severity	Non-case	315	72%
		Case	122	**28%**
	Mild+ severity	Non-case	168	38%
		Case	269	**62%**

*stratified by selection criteria

PheNorm Results - Symptomatic (Mild+) Phenotype

Site	Silver Standard	Phenotype	AUC	Sensitivity at PPV=0.8
KPWA	1 - U07.1 Days	Mild+	0.773	0.89
VUMC	1 - U07.1 Days	Mild+	0.901	0.99
KPWA	2 - Six-ICD Days	Mild+	0.766	0.88
VUMC	2 - Six-ICD Days	Mild+	0.899	0.95
KPWA	3 - COVID Mentions	Mild+	0.864	0.98
VUMC	3 - COVID Mentions	Mild+	0.887	0.94
KPWA	4A - CUI Days	Mild+	0.892	0.98
VUMC	4B - CUI Notes	Mild+	0.875	0.95

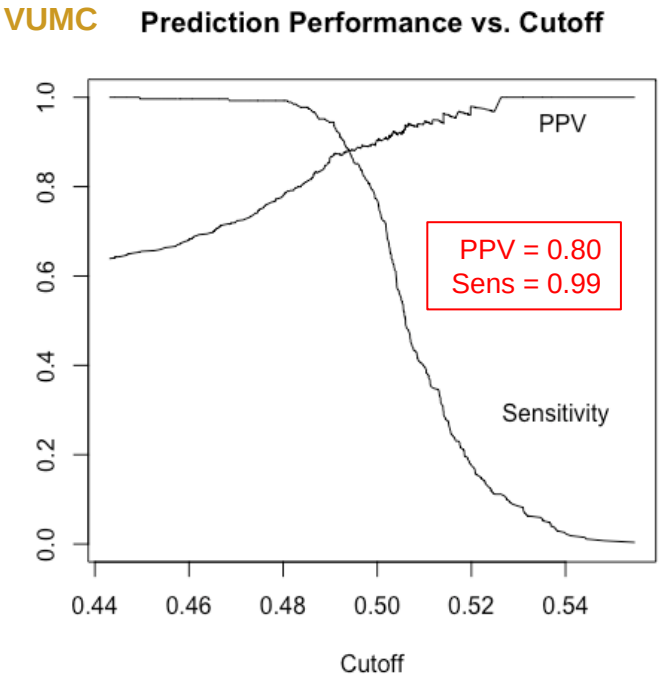
PheNorm Results - Moderate+ Phenotype

Site	Silver Standard	Phenotype	AUC	Sensitivity at PPV=0.8
KPWA	1 - U07.1 Days	Moderate+	0.700	0.07
VUMC	1 - U07.1 Days	Moderate+	0.814	0.29
KPWA	2 - Six-ICD Days	Moderate+	0.695	0.05
VUMC	2 - Six-ICD Days	Moderate+	0.841	0.47
KPWA	3 - COVID Mentions	Moderate+	0.674	0.00
VUMC	3 - COVID Mentions	Moderate+	0.775	0.29
KPWA	4A - CUI Days	Moderate+	0.695	0.00
VUMC	4B - CUI Notes	Moderate+	0.768	0.27

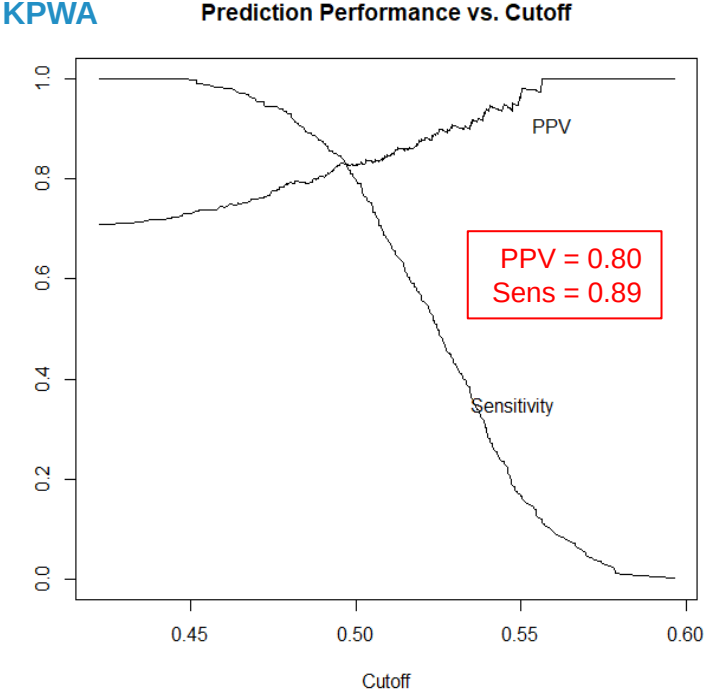
Prediction Performance

Symptomatic (Mild+) phenotype, Silver #1 – U07.1 Days

VUMC



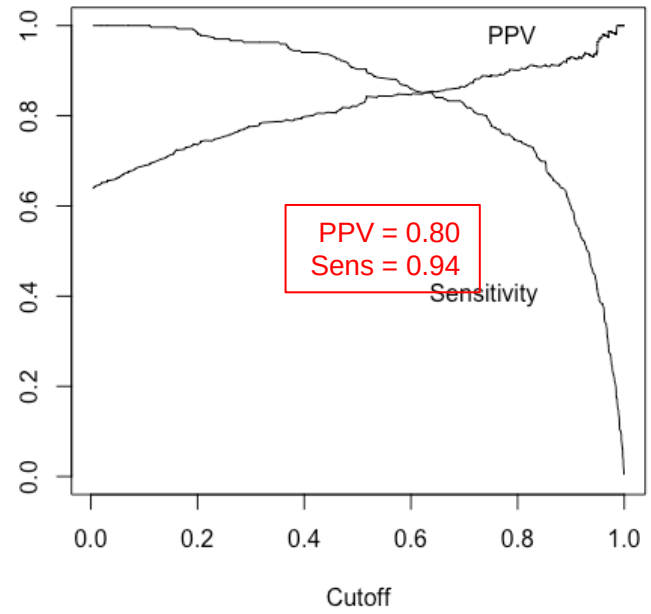
KPWA



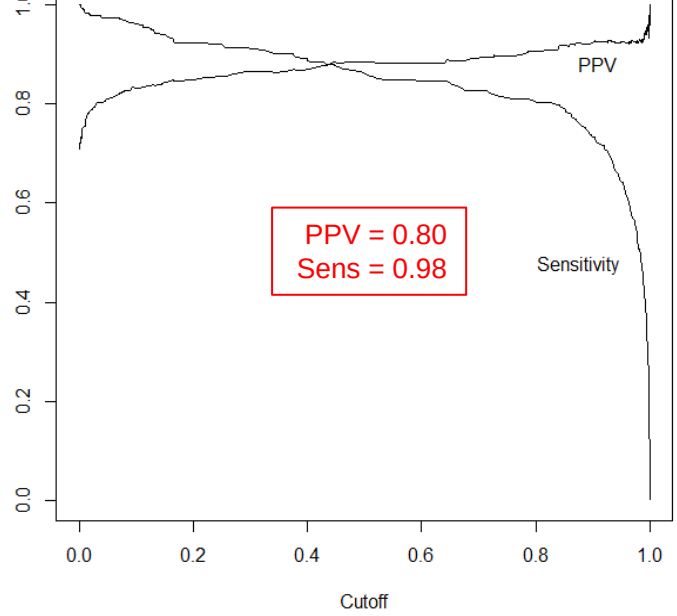
Prediction Performance

Symptomatic (Mild+) phenotype, Silver #3 – COVID-19 Mentions

VUMC



KPWA



Conclusions

Relevance to Sentinel safety surveillance

- *Relatively modest effort* was needed to implement this approach
- *Replication* in two heterogeneous settings was straightforward with (mostly) similar performance
- May be relevant for other *acute* health conditions

Performance of automated models

- While PheNorm performed very well for the symptomatic (Mild+) phenotype, the algorithm worked less well on the Moderate+ phenotype
    - We believe this was likely due to a mismatch between *phenotype definition* and *silver labels*, as well as *phenotype definition* and the *source data*

Limitations & Next Steps

- For NLP-processing and Dictionary generation, we ignored the negation status of mentioned concepts (a la PheCap)
  - Experimenting with including negation (only keeping non-negated concepts)
- Severity-specific model did not perform well
  - Experimenting with using severity-specific silver labels and dictionaries
- Only COVID-19; performance on other acute conditions is not well known
  - We will be continuing this work using PheNorm (and other automated phenotyping models) to explore performance on other acute conditions, as well as chronic phenotypes.
  - Working with other Sentinel partners to support the use of PheNorm within the Sentinel Common Data Model.

Acknowledgements

FDA Sentinel Initiative & Sentinel Innovation Center/Workgroup

My co-PI on this project: David S. Carrell

Our co-authors: Daniel Park, Jill M. Whitaker, Michael F. McLemore, Elizabeth E. Hanchrow, Dax Westerman, Joshua T. Osmanski, Robert Winter, Arvind Ramaprasan, Ann Kelley, Mary Shea, David Cronkite, Saranrat Wittayanukorn, Danijela Stojanovic, Yueqin Zhao, Darren Toh, Kevin B. Johnson, David M Aronoff

Thank you!

Email me at: joshua.smith@vumc.org