Protocol Details

Basic Info

Confirmation Number:	**cjbbjaef**
Protocol Number:	**833483**
Created By:	**WEISSMAN, GARY E**
Principal Investigator:	**WEISSMAN, GARY E**
Protocol Title:	**Natural language processing of clinical notes to determine risk phenotypes**
Short Title:	**NLP of clinical notes to identify risk**
Protocol Description:	**We will identify actionable risk phenotypes embedded in the clinical text of encounter notes to better target interventions to outpatients with chronic lung disease. We will also contribute to methodologic innovations in the processing of clinical text.**
Submission Type:	**Biomedical Research**
Application Type:	**EXPEDITED Category 5**

Resubmission*
No

Study Personnel

Principal Investigator

Name:	**WEISSMAN, GARY E**
Dept / School / Div:	**4257 - DM-Pulmonary, Allergy and Critical Care**
Campus Address Mail Code	**4283**
Address:	**HUP** **3400 SPRUCE ST**
City State Zip:	**PHILADELPHIA PA 19104-4283**
Phone:	**215-349-5488**
Fax:	**-**
Pager:	
Email:	**Gary.Weissman@uphs.upenn.edu**
HS Training Completed:	**Yes**
Training Expiration Date:	**05/30/2020**
Name of course completed :	**CITI Protection of Human Subjects Research Training - ORA**

Study Contacts

Name:	**SILVESTRI, JASMINE A**
Dept / School / Div:	**10724 - DM-Palliative and Advanced Illness Research Center**
Campus Address Mail Code	**4283**
Address:	**Floor 8 3400 Spruce Street** **HUP Maloney 9015**
City State Zip:	**PHILADELPHIA PA 19104-4283**
Phone:	**215-573-1832**
Fax:	**-**
Pager:	
Email:	**jasil@pennmedicine.upenn.edu**
HS Training Completed:	**Yes**
Training Expiration Date:	**08/07/2020**
Name of course completed :	**CITI Protection of Human Subjects Research Training - ORA**

Name:	**LAPITE, FOLASADE C**
Dept / School / Div:	**4009 - SM-DN-Biomedical Graduate Studies**
Campus Address Mail Code	
Address:	**1401 FRENCHMANS BEND RD** **310 S. 36TH STREET MB 0278**
City State Zip:	**MONROE LA 71203-0000**
Phone:	**318-855-8557**
Fax:	**-**
Pager:	
Email:	**flapite@sas.upenn.edu**
HS Training Completed:	**Yes**
Training Expiration Date:	**06/13/2020**
Name of course completed :	**CITI Protection of Human Subjects Research Training - ORA**

Other Investigator

Name:	**HALPERN, SCOTT D**
Dept / School / Div:	**4257 - DM-Pulmonary, Allergy and Critical Care**
Campus Address Mail Code	**6021**
Address:	**423 Guardian Drive** **301 Blockley Hall**
City State Zip:	**Philadelphia PA 19104-6021**
Phone:	**215-898-1462**
Fax:	**-**
Pager:	
Email:	**shalpern@upenn.edu**
HS Training Completed:	**Yes**
Training Expiration Date:	**05/18/2020**
Name of course completed :	**CITI Protection of Human Subjects Research Training - ORA**

Responsible Org (Department/School/Division):

4257 - DM-Pulmonary, Allergy and Critical Care

Key Study Personnel

Name:	**HIMES, BLANCA E**
Department/School/Division:	**BE-Informatics Division**
HS Training Completed:	**Yes**
Training Expiration Date:	**09/03/2020**
Name of course completed:	**CITI Protection of Human Subjects Research Training - ORA**

Name:	**MARTIN, JACOB**
Department/School/Division:	**Health System**
HS Training Completed:	**Yes**
Training Expiration Date:	**06/26/2021**
Name of course completed:	**CITI Protection of Human Subjects Research Training - ORA**

Name:	**CHAIYACHATI, KRISDA H**
Department/School/Division:	**DM-General Internal Medicine**
HS Training Completed:	**Yes**
Training Expiration Date:	**08/25/2021**
Name of course completed:	**CITI Protection of Human Subjects Research Training - ORA**

Name:	**SIMS, MICHAEL W**
Department/School/Division:	**DM-Pulmonary, Allergy and Critical Care**
HS Training Completed:	**Yes**
Training Expiration Date:	**02/28/2021**
Name of course completed:	**CITI Protection of Human Subjects Research Training - ORA**

Name:	**KANGOVI, SHREYA**
Department/School/Division:	**DM-General Internal Medicine**
HS Training Completed:	**Yes**
Training Expiration Date:	**11/17/2019**
Name of course completed:	**CITI Protection of Human Subjects Research Training - ORA**

Name:	**COURTRIGHT, KATHERINE R**
Department/School/Division:	**DM-Palliative and Advanced Illness Research Center**
HS Training Completed:	**Yes**
Training Expiration Date:	**02/28/2022**
Name of course completed:	**CITI Protection of Human Subjects Research Training - ORA**

Name:	**HART, JOANNA L**
Department/School/Division:	**DM-Pulmonary, Allergy and Critical Care**
HS Training Completed:	**Yes**
Training Expiration Date:	**11/07/2019**
Name of course completed:	**CITI Protection of Human Subjects Research Training - ORA**

Name:	**DIAMOND, JOSHUA**
Department/School/Division:	**DM-Pulmonary, Allergy and Critical Care**
HS Training Completed:	**Yes**
Training Expiration Date:	**07/17/2021**
Name of course completed:	**CITI Protection of Human Subjects Research Training - ORA**

Disclosure of Significant Financial Interests*
Does any person who is responsible for the design, conduct, or reporting of this research protocol have a **FINANCIAL INTEREST**?
No

Penn Intellectual Property*
To the best of the Principal Investigator's knowledge, does this protocol involve the testing, development or evaluation of a drug, device, product, or other type of intellectual property (IP) that is owned by or assigned to the University of Pennsylvania?
No

Certification
I have reviewed the *Financial Disclosure and Presumptively Prohibited Conflicts for Faculty Participating in Clinical Trials* and the *Financial Disclosure Policy for Research and Sponsored Projects* with all persons who are responsible for the design, conduct, or reporting of this research; and all required Disclosures have been attached to this application.
Yes

Biomedical Research

Clinical Trial*
Is this a clinical trial?
No

Investigator Initiated Trial*
Is this an investigator initiated trial?
No

Drugs or Devices*
Does this research study involve Drugs or Devices?
No

IND Exemption

For studies that fall under an IND exemption, please provide the number below

For studies including IND or IDE's, please provide the number(s) below

IDE Review*
NOTE: For research involving investigational devices, you are required to review the guidance on Managing Research Device Inventory. Consult the Penn Manual for Clinical Research: https://www.med.upenn.edu/pennmanual/secure/investigational-product-management-at-sites-not-using-investigational-drug-services-(ids).html Please check the box Yes if you have reviewed the guidance.
No

Research Device Management*
Please indicate how research device(s) will be managed.
Not Applicable (no investigational devices)

Drug, Herbal Product or Other Chemical Element Management *
Please indicate how drugs, herbal products or other chemical entities will be managed.
Not Applicable (no drugs, herbal products or other chemical entities)

Radiation Exposure*
Are research subjects receiving any radiation exposure (e.g. X-rays, CT, Fluoroscopy, DEXA, pQCT, FDG, Tc-99m, etc.) that they would not receive if they were not enrolled in this protocol?
No

Gene Transfer*
Does this research involve gene transfer (including all vectors) to human subjects?
No

Human Source Material*
Does this research include collection or use of human source material (i.e., human blood, blood products, tissues or body fluids)?
No

CACTIS and CT Studies*
Does the research involve Center for Advanced Computed Tomography Imaging Services (CACTIS) and CT studies that research subjects would not receive if they were not part of this protocol?
No

CAMRIS and MRI Studies*
Does the research involve Center for Advanced Magnetic Resonance Imaging and Spectroscopy (CAMRIS) and MRI studies that research subjects would not receive if they were not part of this protocol?
No

Investigational Agent or Device within the Operating Room*
Does the research project involve the use of an investigational agent or device within the Operating Room?
No

Cancer Related research not being conducted by an NCI cooperative group*
Does this protocol involve cancer-related studies in any of the following categories?
No

Processing of Materials*
Will the research involve processing (such as over encapsulating, or compounding)?
No

In-House Manufacturing of Materials*
Will the research involve processing (such as over encapsulating, or compounding)?
No

Medical Information Disclosure*
Does the research proposal involve the use and disclosure of research subject's medical information for research purposes?
No

If the answer is YES, indicate which items is is provided with this submission:

CTRC Resources*
Does the research involve CTRC resources?
No

Pathology and Laboratory Medicine Resources*
Will samples be collected by hospital phlebotomy and/or processed or analyzed by any of the clinical laboratories of the University of Pennsylvania Health System?
No

Research Involves Apheresis, Cell Collection, and/or Blood Product Collection*
Does this research involve collection of blood products in the Penn Donor Center and/or the use of apheresis for treatment or collection of cells or other blood components?
No

Research involving blood transfusion or drug infusions*
Will your research involve blood transfusion or infusion of study drug in 3 Ravdin Apheresis Unit for research purposes?
No

Trial in Radiation Oncology
Is this research a prospective trial being done in Radiation Oncology, and if so, has this protocol been approved by the Radiation Oncology Protocol committee?
N/A

Study in Radiation Oncology
Is this research a retrospective study being done in Radiation Oncology, and if so, has this project been reviewed by the Radiation Oncology Clinical Research Group?
N/A

Use of UPHS services*
Does your study require the use of University of Pennsylvania Health System (UPHS) services, tests or procedures*, whether considered routine care or strictly for research purposes?
No

Primary Focus*
Research on human data sets (e.g. medical records, clinical registries, existing research data sets, medical administrative data, etc.)

Protocol Interventions

	Sociobehavioral (i.e. cognitive or behavioral therapy)
	Drug
	Device - therapeutic
	Device - diagnostic (assessing a device for sensitivity or specificity in disease diagnosis)
	Surgical
	Diagnostic test/procedure (research-related diagnostic test or procedure)
	Obtaining human tissue for basic research or biospecimen bank
	Survey instrument
x	**None of the above**

The following documents are currently attached to this item:

There are no documents attached for this item.

Sponsors

Business Administrator

Name:	**SZERSZEN, SYLVIA M**
Dept / School / Div:	**4222 - DM-Department of Medicine**
Phone:	**-**
Fax:	**215-746-8931**
Pager:	
Email:	**sylvia.szerszen@uphs.upenn.edu**

Department budget code

000 - 000 - 0 - 000000 - 0000 - 0000 - 0000

Funding Sponsors

Funding sponsors billing address
If you have selected a commercial or industry sponsor, please provide the appropriate address and contact information for the Sponsor for the purposes of billing for IRB review fees (initial review, continuing review and convened modification fees apply here). If the Sponsor is not industry/ commercial, this information is not necessary to provide with your application.

Funding sponsors gift
Is this research being funded by a philanthropic gift?
No

Regulatory Sponsor

IND Sponsor

none

000 - 000 - 0 - 000000 - 0000 - 0000 - 0000

Industry Sponsor

None

Project Funding*
Is this project funded by or associated with a grant or contract?
Yes

Selected Proposals

Proposal No	Title
10062964	**Using natural language processing and machine learning to identify potentially preventable hospital**

Sponsor Funding
Is this study funded by an industry sponsor?
No

Status of contract

The following documents are currently attached to this item:

Grant Application (4b-research_approach.pdf)

Multi-Center Research

Penn as lead
1. Is this a multi-center study where Penn is serving as the Lead Site or the Penn PI is serving as the Lead Investigator?
No

Management of Information for Multi-Center Research

Penn irb of record
2. Is this a multi-center study where the Penn IRB will be asked to serve as the IRB of Record for other external study sites?
No

Other Sites

No other sites

Protocol

Abstract
The goal of this study is to build software tools (classifiers) that can identify important risk factors for clinical needs in an outpatient population by scanning all of the electronic health record data, including the text of clinical notes. This software will enable learning health systems, on a large scale, to proactively identify patients with particular needs rather than waiting for them to be admitted to the hospital or emergency room first. In the future, this approach will allow health systems not just to identify such needs, but also to make appropriate referrals or care decisions in real time to keep people at home, feeling well, and engaged in the healthcare system.

Objectives

Overall objectives

The overall objectives of the study are to produce text-based classification models for separate actionable clinical needs including frailty, depression, transportation need, unmet palliative care need, need for a community health worker, need for social support, and other clinical risk phenotypes to be determined. Each classification model will be trained and validated separately, and each will have its own final software product to be released as an open source software library.

Primary outcome variable(s)

The primary outcome variable for each risk phenotype will be the presence of text suggesting the existence of that phenotype at the sentence level of each document. These sentence-level predictions will then be combined with other sentences from the document, with other structured data elements, and other clinical information from the EHR to yield a final probability of having the clinical phenotype in question.

Secondary outcome variable(s)

Secondary analyses will be used to further validate the performance of the classifier. For example, for the frailty classifier will be measured against a cohort of about 300 patients who underwent the Short Physical Performance Battery (SPPB) and the Fried Frailty Phenotype in the Penn Lung Center. We will calibrate our predictions of frailty against these other measures and assess their correlation.

Background

While 15.7 million adults reported being diagnosed with chronic obstructive pulmonary disease (COPD) in the US, over 50% of adults with impaired pulmonary function were not aware of their diagnosis (2,3), suggesting a higher prevalence than has been formally documented. COPD is the third leading cause of death in the US (4) while COPD associated hospitalization account for $13.2 billion in direct costs each year (5). One in every five of these hospitalizations is associated with a hospital readmission within 30 days (6,7). Although many of these readmissions are thought to be preventable (8-10), despite identification of all-cause admission risk profiles for patients with COPD (11-15), national admission rates have stayed flat while costs have increased (16). Despite the large financial, personal, and health system burdens of COPD, little evidence exists to suggest actionable mechanisms through which COPD readmissions may be averted. Without a deep understanding of causal mechanisms underlying 30-day COPD readmissions, financial incentives have not produced meaningful changes in readmission rates. Similarly, clinical prediction models aimed at risk-stratifying patients are left with only moderate performance and minimal actionable insights. As previous attempts to design interventions have ignored heterogeneity in underlying mechanisms, or have lumped so many interventions together, their individual effects and which patients are likely to benefit remain unknown. Therefore, this proposal seeks to by 1) identify the root causes leading to preventable COPD readmissions, and 2) develop large-scale, automated approaches to discern the presence of these root causes using free-text data in the electronic health record (EHR). The traditional analysis of COPD readmissions has focused on structured data fields like the diagnosis list to identify acute respiratory viral infection or shortness of breath as an underlying cause. We recognize that the first in a chain of events leading a patient to seek care the iatrotropic stimulus (17) is likely more complex and embedded within the social, behavioral, geographic, clinical, and economic environment. If the problem is lack of transportation to get to the pharmacy to refill medications, or significant frailty in someone not receiving appropriate rehabilitation therapy, the diagnosis list in the administrative claims will not capture it. We hypothesize that such factors, while potentially actionable, are best elicited with deep conversations with patients, and best identified through incorporating subtle, narrative information from free-text EHR sources. References 1. Wheaton AG, Cunningham, TJ, Ford ES, Croft JB. Employment and activity limitations among adults with chronic obstructive pulmonary diseaseUnited States, 2013. MMWR. 2015:64 (11):290-295. 2. Mannino DM, Gagnon RC, Petty TL, Lydick E. Obstructive lung disease and low lung function in adults in the United States: data from the National Health and Nutrition Examination Survey 1988-1994. Arch Intern Med. 2000;160:1683-1689. 3. Lamprecht B, Soriano JB, Studnicka M, et al. Determinants of underdiagnosis of COPD in national and international surveys. CHEST Journal 2015;148(4):971985. 4. National Center for Health Statistics. Leading causes of death. Centers for Disease Control and Prevention Website. http://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm. Accessed February 20, 2016. 5. Wier LM, Elixhauser A, Pfuntner A, et al. Overview of hospitalizations among patients with COPD. AHRQ Statistical Brief #106, 2006. 6. Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for- service

program. N Engl J Med 2009;360(14):1418-1428 7. Shah T, Churpek MM, Coca Perraillon M, Konetzka RT. Understanding why patients with COPD get readmitted: a large national study to delineate the Medicare population for the readmissions penalty expansion. Chest 2015;147 (5):1219-1226 8. Billings J, Anderson GM, Newman LS. Recent findings on preventable hospitalizations. Health Aff 1996;15(3):239249. 9. Bindman AB, Grumbach K, Osmond D, et al. Preventable hospitalizations and access to health care. JAMA 1995;274(4):305311. 10. Feemster LC, Au DH. Penalizing hospitals for chronic obstructive pulmonary disease readmissions. Am J Resp Crit Care Med 2014;189(6):634639. 11. Kessler R, Faller L, Fourgaut G, et al. Predictive factors of hospitalization for acute exacerbation in a series of 64 patients with chronic obstructive pulmonary disease. Am J Resp Crit Care Med 1999;159(1):158164. 12. Montserrat-Capdevila J, Godoy P, Marsal JP, et al. Predictive model of hospital admission for COPD exacerbation. Resp Care 2015;60 (9):12881294. 13. Lemke KW, Weiner JP, Clark JM. Development and validation of a model for predicting inpatient hospitalization. Med Care 2012;50(2):131139. 14. Bhatt SP, Wells JM, Iyer AS, et al. Results of a Medicare bundled payments for care improvement initiative for chronic obstructive pulmonary disease readmissions. Ann Am Thorac Soc 2017;14(5):643 648. 15. Bashir B, Schneider D, Naglak MC, et al. Evaluation of prediction strategy and care coordination for COPD readmissions. Hosp Pract 2016;44(3):123128. 16. Torio CM, Elixhauser A, Andrews RM. Trends in potentially preventable hospital admissions among adults and children, 20052010. AHRQ Statistical Brief #151, 2006. 17. Feinstein AR. Clinical Judgment. Williams & Wilkins, Baltimore, 1967.

Study Design

Phase*
Not applicable

Design
This is a retrospective cohort study focused on the development, validation, and dissemination of a series of text classifiers to identify actionable clinical phenotypes in the electronic health record. For each phenotype, we will use an iterative sampling, training, and validation approach (see attached grant for more details). The first round of sampling will include a purposive sample of 50 notes (25 high- and 25-low risk for the phenotype) based on crude ICD-based definitions of the outcome. These 50 notes will be reviewed with the research team and with clinical experts in the domain and in clinical natural language processing from our study team. We will identify the complexity of the task, important vocabulary terms, and identify other relevant NLP tasks (e.g. negation, temporal dependence, normalization, etc) that are relevant for the task. This will inform the choice of model to build the classifier: a rules-based model using regular expressions, a logistic regression or SVM using word embeddings, or a more complex neural network-based model using the embeddings for word representation. We will train a set of word embeddings using local data from all outpatient notes in the timeframe of the study and test them against embeddings trained on non-Penn clinical notes. We will then iteratively train the models and update their performance to build a learning curve to identify a final sample size for each classifier. We estimate that to train each model we will need 500 to 1000 notes, and will need to annotate an additional 500 notes for a hold-out testing sample to capture performance. The classification model will rely on the content of the text of notes, structured data fields (age, gender, diagnostic and procedure codes, home address, and other demographics). Note types include home visits, outpatient clinical encounters, consults, telephone notes, and other electronic communications in the EHR.

Study duration
We estimate it will take 6 months to properly query and clean all of the relevant data, 12 months to annotate it, and 6 months to train and develop the classification models, for a total of 24 months.

Resources necessary for human research protection
Describe research staff and justify that the staff are adequate in number and qualifications to conduct the research. Describe how you will ensure that all staff assisting with the research are adequately informed about the protocol and their research related duties. Please allow adequate time for the researchers to conduct and complete the research. Please confirm that there are adequate facilities for the research.
All research team members have received updated certification in human subjects research and extensive in direct clinical care and/or research with human subjects and PHI, so are informed of the necessary precautions in handling EHR data. The facilities to conduct the proposed research are

adequate and are located within the PAIR Center at the University of Pennsylvania Perelman School of Medicine. The PAIR Center includes a data core, a statistical core, and numerous investigators with clinical informatics experience.

Characteristics of the Study Population

Target population
We will analyze all patients in the University of Pennsylvania Health System who had any outpatient encounters from January 1, 2017 through December 31, 2019. The analysis will focus on up to 12 months of historical data preceding any outpatient encounters (as far back as January 1, 2016), and up to 12 months following any outpatient encounter (through December 31, 2020) to determine future utilization patterns (inpatient or emergency care) associated with features of the clinical text. We will then examine subsets of patients with different characteristics, including those with chronic lung disease, heart failure, diabetes, and other chronic medical conditions.

Subjects enrolled by Penn Researchers
200000

Subjects enrolled by Collaborating Researchers
0

Accrual
NA

Key inclusion criteria
Any outpatient encounter in the University of Pennsylvania Health from January 1, 2017 through December 31, 2019.

Key exclusion criteria
We will only analyze documents for patients older than 18 years of age.

Vulnerable Populations

	Children Form
	Pregnant women (if the study procedures may affect the condition of the pregnant woman or fetus) Form
	Fetuses and/or Neonates Form
	Prisoners Form
	Other
x	**None of the above populations are included in the research study**

The following documents are currently attached to this item:

There are no documents attached for this item.

Populations vulnerable to undue influence or coercion
NA

Subject recruitment
NA

Will the recruitment plan propose to use any Penn media services (communications, marketing, etc.) for outreach via social media avenues (examples include: Facebook, Twitter, blogging, texting, etc.) or does the study team plan to directly use social media to recruit for the research?
No

The following documents are currently attached to this item:

There are no documents attached for this item.

Subject compensation*
Will subjects be financially compensated for their participation?
No

The following documents are currently attached to this item:

There are no documents attached for this item.

If there is subject compensation, provide the schedule for compensation per study visit or session and total amount for entire participation, either as text or separate document

None.

Study Procedures

Suicidal Ideation and Behavior
Does this research qualify as a clinical investigation that will utilize a test article (ie- drug or biological) which may carry a potential for central nervous system (CNS) effect(s)?
No

Procedures
None

The following documents are currently attached to this item:

There are no documents attached for this item.

Deception
Does your project use deception?
No

International Research
Are you conducting research outside of the United States?
No

Analysis Plan
Please see the Design section for analytic methods.

The following documents are currently attached to this item:

There are no documents attached for this item.

Data confidentiality

	Paper-based records will be kept in a secure location and only be accessible to personnel involved in the study.
x	**Computer-based files will only be made available to personnel involved in the study through the use of access privileges and passwords.**
	Prior to access to any study-related information, personnel will be required to sign statements agreeing to protect the security and confidentiality of identifiable information.
x	**Wherever feasible, identifiers will be removed from study-related information.**
	A Certificate of Confidentiality will be obtained, because the research could place the subject at risk of criminal or civil liability or cause damage to the subject's financial standing, employability, or liability.
x	**A waiver of documentation of consent is being requested, because the only link between the subject and the study would be the consent document and the primary risk is a breach of confidentiality. (This is not an option for FDA-regulated research.)**
x	**Precautions are in place to ensure the data is secure by using passwords and encryption, because the research involves web-based surveys.**
	Audio and/or video recordings will be transcribed and then destroyed to eliminate audible identification of subjects.

Subject Confidentiality
All data will be stored on password-protected computers locked in the PAIR office on the PMACS network with identifiers removed as soon as able for the analysis. Only those team members requiring access to the de-identified data will then have access to the raw information. Six months following publication of the final analysis all data files will be removed to allow for time to respond to any final reviewer or editor concerns.

Sensitive Research Information*
Does this research involve collection of sensitive information about the subjects that should be excluded from the electronic medical record?
No

Subject Privacy
Privacy refers to the person's desire to control access of others to themselves. Privacy concerns people, whereas confidentiality concerns data. Describe the strategies to protect privacy giving consideration to the following: The degree to which privacy can be expected in the proposed research and the safeguards that will be put into place to respect those boundaries. The methods used to identify and contact potential participants. The settings in which an individual will be interacting with an investigator. The privacy guidelines developed by relevant professions, professional associations and scholarly disciplines (e.g., psychiatry, genetic counseling, oral history, anthropology, psychology).
Full privacy can be expected. See above for protections regarding confidentiality. Additionally, participants will not be contacted directly and a request for waiver of consent is being requested to reduce any risk of exposing the identities of any participants. Given there are no interactions between participants and investigators in this purely retrospective cohort study, the risk for breach of privacy is minimal.

Data Disclosure
Will the data be disclosed to anyone who is not listed under Personnel?
None.

Data Protection*

- x **Name**
- x **Street address, city, county, precinct, zip code, and equivalent geocodes**
- x **All elements of dates (except year) for dates directly related to an individual and all ages over 89**
- **Telephone and fax number**
- **Electronic mail addresses**
- **Social security numbers**
- x **Medical record numbers**
- **Health plan ID numbers**
- **Account numbers**
- **Certificate/license numbers**
- **Vehicle identifiers and serial numbers, including license plate numbers**
- **Device identifiers/serial numbers**
- **Web addresses (URLs)**
- **Internet IP addresses**
- **Biometric identifiers, incl. finger and voice prints**
- **Full face photographic images and any comparable images**
- **Any other unique identifying number, characteristic, or code**
- **None**

Does your research request both a waiver of HIPAA authorization for collection of patient information and involve providing Protected Health Information ("PHI") that is classified as a "limited data set" (city/town/state/zip code, dates except year, ages less than 90 or aggregate report for over 90) to a recipient outside of the University of Pennsylvania covered entity?
No

Tissue Specimens Obtained as Part of Research*
Are Tissue Specimens being obtained for research?
No

Tissue Specimens - Collected during regular care*
Will tissue specimens be collected during regulator clinical care (for treatment or diagnosis)?
No

Tissue Specimens - otherwise discarded*
Would specimens otherwise be discarded?
No

Tissue Specimens - publicly available*
Will tissue specimens be publicly available?
No

Tissue Specimens - Collected as part of research protocol*
Will tissue specimens be collected as part of the research protocol?
No

Tissue Specimens - Banking of blood, tissue etc. for future use*
Does research involve banking of blood, tissue, etc. for future use?
No

Genetic testing
If genetic testing is involved, describe the nature of the tests, including if the testing is predicative or exploratory in nature. If predictive, please describe plan for disclosing results to subjects and provision

of genetic counseling. Describe how subject confidentiality will be protected Note: If no genetic testing is to be obtained, write: "Not applicable."
Not applicable.

Consent

1. Consent Process

Overview
A waiver of informed consent is being requested for this purely retrospective cohort analysis involving the use of electronic health record information.

Children and Adolescents
Not applicable. Only those 18 years or older are being included in the study.

Adult Subjects Not Competent to Give Consent
Not applicable. See below -- we are requesting a waiver of informed consent.

2. Waiver of Consent

Waiver or Alteration of Informed Consent*
Waiver of written documentation of informed consent: the only record linking the subject and the research would be the consent document and the principal risk would be potential harm resulting from a breach of confidentiality.

Minimal Risk*

Impact on Subject Rights and Welfare*

Waiver Essential to Research*

Additional Information to Subjects

Written Statement of Research*
No

If no written statement will be provided, please provide justificiation

We are requesting a waiver of informed consent because the consent process would represent the only risk to breach of privacy. All data used are already available in the electronic health record database. Therefore, no written statement of research will be provided.

The following documents are currently attached to this item:

There are no documents attached for this item.

Risk / Benefit

Potential Study Risks
There are no interventions in this study. The only potential risk of this study is a loss of confidentiality through the use of protected health information derived from the electronic health record. The likelihood of this loss of confidentiality is minimal given the protection in place to secure digital data and the transition to a de-identified version of the dataset as soon as all the available information has been extracted.

Potential Study Benefits
There are no potential benefits to be gained on behalf of individuals whose information is contained

within the dataset through the analysis itself. However, with a better understanding of risk stratification, future patients seen at Penn Medicine clinics, including those whose data were included in the analytic dataset, may benefit from better allocation of community services to promote higher-value allocation of care.

Alternatives to Participation (optional)
Not applicable.

Data and Safety Monitoring
Data privacy will be periodically assessed by all investigators on the study team to ensure minimization of PHI in any dataset use for analysis and to ensure safe practices in maintaining the dataset in password protected, secure locations behind the PMACS firewall at all times.

The following documents are currently attached to this item:

There are no documents attached for this item.

Risk / Benefit Assessment
The expected risk is extremely low and minimal, while the potential benefit to society is potentially moderate.

General Attachments

The following documents are currently attached to this item:

Grant Application (4b-research_approach.pdf)

Cover Letter (k23_aim_2_irb_cover_letter.docx)

HIPAA Authorization or Waiver (k23_aim2_requestforwaiverofhipaaauthorization2017.10_0.docx)