Title : Stanford PHS Data Protocol: Data Exploration, Cohort Identification Educational Activities and Research in Population Sciences

Approval Period: 01/25/2019 - 04/30/2019

Modification

1. **Summarize your proposed changes.**
   Clarifications on types of data and variables. Updated DMP. No substantive changes.
2. **Indicate Level of Risk**
   No Change
3. **Update the Conflict of Interest (COI) section if any changes in COI have been made since the last protocol submission.**

   N Is there a change in the conflicting interest status for any existing personnel on this protocol?

Protocol Director

Name		Degree (Program/year if student)		Position, e.g. Assistant Professor, Resident, etc.
Mark Cullen		MD		Professor
Department	**Mail Code**	**Phone**	**Fax**	**E-mail**
Medicine - Med/General Internal Medicine	5411	(650) 721-6209	(659) 723-8596	mrcullen@stanford.edu
CITI Training current				Y

Admin Contact

Name		Degree (Program/year if student)		Position, e.g. Assistant Professor, Resident, etc.
Isabella Chu		MPH		Data Core Manager
Department	**Mail Code**	**Phone**	**Fax**	**E-mail**
Medicine - Med/General Internal Medicine	5411	(650) 723-2513		isabella.chu@stanford.edu
CITI Training current				Y

Investigator

Name		Degree (Program/year if student)		Position, e.g. Assistant Professor, Resident, etc.
Lorene Nelson		Ph.D., M.S.		Associate Professor & Chief, Division of Epidemiology
Department	**Mail Code**	**Phone**	**Fax**	**E-mail**
Health Research and Policy - Epidemiology	5405	(650) 723-6854	(650) 725-6951	lnelson@stanford.edu
CITI Training current				Y

Other Contact

Name	Degree (Program/year if student)	Position, e.g. Assistant Professor, Resident, etc.
Valerie Carolina Meausoone		

Title : Stanford PHS Data Protocol: Data Exploration, Cohort Identification Educational Activities and Research in Population Sciences
Approval Period: 01/25/2019 - 04/30/2019

		BS		Assistant Director, Data Implementation
Department	**Mail Code**	**Phone**	**Fax**	**E-mail**
Population Health Sciences	5560	(650) 721-8460		vmeau@stanford.edu
CITI Training current				Y

Academic Sponsor

Name		**Degree (Program/year if student)**		**Position, e.g. Assistant Professor, Resident, etc.**
Department	**Mail Code**	**Phone**	**Fax**	**E-mail**
CITI Training current				

Other Personnel

Name Ivan Mejia Guevara		**Degree (Program/year if student)** PhD		**Position, e.g. Assistant Professor, Resident, etc.** Sr. Res Scientist-Basic Life
Department	**Mail Code**	**Phone**	**Fax**	**E-mail**
Biology	5020			imejia@stanford.edu
CITI Training current				Y

Name Suzanne Tamang		**Degree (Program/year if student)** PhD		**Position, e.g. Assistant Professor, Resident, etc.** Research Engineer
Department	**Mail Code**	**Phone**	**Fax**	**E-mail**
		(650) 497-4392		stamang@stanford.edu
CITI Training current				Y

Name Sanne Smith		**Degree (Program/year if student)** PhD		**Position, e.g. Assistant Professor, Resident, etc.** Research Scientist and Academic Program Professional
Department	**Mail Code**	**Phone**	**Fax**	**E-mail**
Graduate School of Education				sannesmith@stanford.edu
CITI Training current				Y

Name Pooja Loftus		**Degree (Program/year if student)** MS		**Position, e.g. Assistant Professor, Resident, etc.** Biostatistician 3
Department	**Mail Code**	**Phone**	**Fax**	**E-mail**
		(650) 721-6110		ploftus@stanford.edu
CITI Training current				Y

Name Ian Barry Mathews		**Degree (Program/year if student)**		**Position, e.g. Assistant Professor, Resident, etc.**

Title : Stanford PHS Data Protocol: Data Exploration, Cohort Identification Educational Activities and Research in Population Sciences
Approval Period: 01/25/2019 - 04/30/2019

		MS		
Department Medicine - Med/General Internal Medicine	**Mail Code** 3073	**Phone** (650) 761-1991	**Fax**	**E-mail** imathews@stanford.edu
CITI Training current				Y

Name Sean Francis Mcintyre		**Degree (Program/year if student)** MS		**Position, e.g. Assistant Professor, Resident, etc.**
Department Medicine - Med/General Internal Medicine	**Mail Code**	**Phone** (650) 761-1991	**Fax**	**E-mail** sean13@stanford.edu
CITI Training current				Y

Name Erin Kathleen Delaney		**Degree (Program/year if student)** BS		**Position, e.g. Assistant Professor, Resident, etc.**
Department Medicine - Med/General Internal Medicine	**Mail Code** 4101	**Phone** (650) 761-1991	**Fax**	**E-mail** delaney1@stanford.edu
CITI Training current				Y

Name Shahrzad Aminshahidy		**Degree (Program/year if student)** MS		**Position, e.g. Assistant Professor, Resident, etc.**
Department Population Health Sciences	**Mail Code**	**Phone** (650) 761-1991	**Fax**	**E-mail** sherriam@stanford.edu
CITI Training current				Y

Name Erika Tribett		**Degree (Program/year if student)**		**Position, e.g. Assistant Professor, Resident, etc.** Academic Prog Prof 1
Department Population Health Sciences	**Mail Code** 5411	**Phone**	**Fax**	**E-mail** etribett@stanford.edu
CITI Training current				Y

Name Emma Sofia Thonander Hallgren		**Degree (Program/year if student)**		**Position, e.g. Assistant Professor, Resident, etc.** Biostatistician 1
Department Pediatrics - Neonatology	**Mail Code** 5731	**Phone**	**Fax**	**E-mail** emma.hallgren@stanford.edu

Title : Stanford PHS Data Protocol: Data Exploration, Cohort Identification Educational Activities and Research in Population Sciences
Approval Period: 01/25/2019 - 04/30/2019

CITI Training current	Y

Name		**Degree (Program/year if student)**		**Position, e.g. Assistant Professor, Resident, etc.**
Gary Darmstadt				Professor-Teaching
Department	**Mail Code**	**Phone**	**Fax**	**E-mail**
Pediatrics - Neonatology	5731			gdarmsta@stanford.edu
CITI Training current				Y

Name		**Degree (Program/year if student)**		**Position, e.g. Assistant Professor, Resident, etc.**
Safa Abdalla				Biostatistician
Department	**Mail Code**	**Phone**	**Fax**	**E-mail**
Pediatrics - Neonatology	5731			sabdalla@stanford.edu
CITI Training current				Y

Name		**Degree (Program/year if student)**		**Position, e.g. Assistant Professor, Resident, etc.**
Irena Stepanikova				
Department	**Mail Code**	**Phone**	**Fax**	**E-mail**
Pediatrics - Neonatology	6074			irena.stepanikova@stanford.edu
CITI Training current				Y

Name		**Degree (Program/year if student)**		**Position, e.g. Assistant Professor, Resident, etc.**
Lesley Park				Instructor
Department	**Mail Code**	**Phone**	**Fax**	**E-mail**
Med/Primary Care and Population Health	5411	(650) 721-8410		lesley.park@stanford.edu
CITI Training current				Y

Name		**Degree (Program/year if student)**		**Position, e.g. Assistant Professor, Resident, etc.**
Lisa Chamberlain				Assoc Prof-Med Ctr Line
Department	**Mail Code**	**Phone**	**Fax**	**E-mail**
Pediatrics - General Pediatrics	5459	(650) 725-8314	(650) 498-5684	lchamberlain@stanford.edu
CITI Training current				Y

Participant Population(s) Checklist	**Yes/No**
• Children (under 18)	Y
• Pregnant Women and Fetuses	N
• Neonates (0 - 28 days)	N

Title : Stanford PHS Data Protocol: Data Exploration, Cohort Identification Educational Activities and Research in Population Sciences
Approval Period: 01/25/2019 - 04/30/2019

• Abortuses	N
• Impaired Decision Making Capacity	N
• Cancer Subjects	N
• Laboratory Personnel	N
• Healthy Volunteers	N
• Students	N
• Employees	N
• Prisoners	N
• Other (i.e., any population that is not specified above)	Y

Study Location(s) Checklist

	Yes/No
• Stanford University	Y
• Clinical & Translational Research Unit (CTRU)	Y
• Stanford Hospital and Clinics	Y
• Lucile Packard Children's Hospital (LPCH)	Y
• VAPAHCS (Specify PI at VA)	
• Other (Click ADD to specify details)	

General Checklist

Multi-site

	Yes/No
• Is this a multi-site study? A multi-site study is generally a study that involves one or more medical or research institutions in which one site takes a lead role.(e.g., multi-site clinical trial)	N

Collaborating Institution(s)

	Yes/No
• Are there any collaborating institution(s)? A collaborating institution is generally an institution that collaborates equally on a research endeavor with one or more institutions.	N

Cancer Institute

	Yes/No
• Cancer-Related Studies (studies with cancer endpoints), Cancer Subjects (e.g., clinical trials, behavior/prevention) or Cancer Specimens (e.g., blood, tissue, cells, body fluids with a scientific hypothesis stated in the protocol).	N

Clinical Trials

	Yes/No
• Investigational drugs, biologics, reagents, or chemicals?	N
• Commercially available drugs, reagents, or other chemicals administered to subjects (even	N

Title : Stanford PHS Data Protocol: Data Exploration, Cohort Identification Educational Activities and Research in Population Sciences
Approval Period: 01/25/2019 - 04/30/2019

if they are not being studied)?

- Investigational Device / Commercial Device used off-label? N
- IDE Exempt Device (Commercial Device used according to label, Investigational In Vitro Device or Assay, or Consumer Preference/Modifications/Combinations of Approved Devices) N

- Will this study be registered on# clinicaltrials.gov? (See Stanford decision tree) N
- Is Stanford responsible for ClinicalTrials.gov registration? (See Stanford decision tree) N

NCT#

Tissues and Specimens **Yes/No**

- Human blood, cells, tissues, or body fluids (tissues)? N
- Tissues to be stored for future research projects? N
- Tissues to be sent out of this institution as part of a research agreement? For guidelines, please see https://sites.stanford.edu/ico/mtas N

Biosafety (APB) **Yes/No**

- Are you submitting a recombinant DNA vector or Human Gene Transfer investigation using biological agents? If yes, please complete and attach the Gene Transfer Protocol Application Supplemental Questions to section 16 of the eProtocol application. N
- Are you submitting a Human study using biohazardous/infectious agents? If yes, refer to the http://www.stanford.edu/dept/EHS/prod/researchlab/bio/index.html Administrative Panel on BioSafety website prior to performing studies. N
- Are you submitting a Human study using samples from subjects that are known or likely to contain biohazardous/infectious agents? If yes, refer to the http://web.stanford.edu/dept/EHS/prod/researchlab/bio/index.html Administrative Panel on BioSafety website prior to performing studies. N

Human Embryos or Stem Cells **Yes/No**

- Human Embryos or Gametes? N
- Human Stem Cells (including hESC, iPSC, cancer stem cells, progenitor cells) N

Veterans Affairs (VA) **Yes/No**

- The research recruits participants at the Veterans Affairs Palo Alto Health Care System(VAPAHCS). N
- The research involves the use of VAPAHCS non-public information to identify or contact human research participants or prospective subjects or to use such data for research purposes. N
- The research is sponsored (i.e., funded) by VAPAHCS. N
- The research is conducted by or under the direction of any employee or agent of N

Title : Stanford PHS Data Protocol: Data Exploration, Cohort Identification Educational Activities and Research in Population Sciences
Approval Period: 01/25/2019 - 04/30/2019

VAPAHCS (full-time, part-time, intermittent, consultant, without compensation (WOC), on-station fee-basis, on-station contract, or on-station sharing agreement basis) in connection with her/his VAPAHCS responsibilities.

- The research is conducted using any property or facility of VAPAHCS. N

Equipment	Yes/No
• Use of Patient related equipment? If Yes, equipment must meet the standards established by Hospital Instrumentation and Electrical Safety Committee (650-725-5000)	N
• Medical equipment used for human patients/subjects also used on animals?	N
• Radioisotopes/radiation-producing machines, even if standard of care? http://www.stanford.edu/dept/EHS/prod/researchlab/radlaser/Human_use_guide.pdf More Info	N

Payment	Yes/No
• Subjects will be paid/reimbursed for participation? See payment considerations.	N

Funding	Yes/No
• Training Grant?	N
• Program Project Grant?	N
• Federally Sponsored Project?	Y
• Industry Sponsored Clinical Trial?	N

Funding

Funding - Grants/Contracts

Funding - Fellowships

Gift Funding

Dept. Funding

Other Funding

Other Fund Name : Stanford CTRU (Spectrum)

Expedited Form

A protocol must be no more than minimal risk (i.e., "not greater than those ordinarily encountered in daily life") AND must only involve human subjects in one or more of the following paragraphs.

Title : Stanford PHS Data Protocol: Data Exploration, Cohort Identification Educational Activities and Research in Population Sciences
Approval Period: 01/25/2019 - 04/30/2019

Select one or more of the following paragraphs:

1. N **Clinical studies of drugs and medical devices only when condition (a) or (b) is met.**
- a) Research on drugs for which an investigational new drug application (21 CFR Part 312) is not required. (Note: Research on marketed drugs that significantly increases the risks or decreases the acceptability of the risks associated with the use of the product is not eligible for expedited review.)
- b) Research on medical devices for which
  - i) an investigational device exemption application (21 CFR Part 812) is not required; or
  - ii) the medical device is cleared/approved for marketing and the medical device is being used in accordance with its cleared/approved labeling.

2. N **Collection of blood samples by finger stick, heel stick, ear stick, or venipuncture as follows:**
- a) from healthy, nonpregnant adults who weigh at least 110 pounds. For these subjects, the amounts drawn may not exceed 550 ml in an 8 week period and collection may not occur more frequently than 2 times per week; or
- b) from other adults and children, considering the age, weight, and health of the subjects, the collection procedure, the amount of blood to be collected, and the frequency with which it will be collected. For these subjects, the amount drawn may not exceed the lesser of 50 ml or 3 ml per kg in an 8 week period and collection may not occur more frequently than 2 times per week.

3. N **Prospective collection of biological specimens for research purposes by non invasive means.**

4. N **Collection of data through non invasive procedures (not involving general anesthesia or sedation) routinely employed in clinical practice, excluding procedures involving x-rays or microwaves. Where medical devices are employed, they must be cleared/approved for marketing. (Studies intended to evaluate the safety and effectiveness of the medical device are not generally eligible for expedited review, including studies of cleared medical devices for new indications.)**

Examples:
- a) physical sensors that are applied either to the surface of the body or at a distance and do not involve input of significant amounts of energy into the subject or an invasion of the subject's privacy;
- b) weighing or testing sensory acuity;
- c) magnetic resonance imaging;
- d) electrocardiography, electroencephalography, thermography, detection of naturally occurring radioactivity, electroretinography, ultrasound, diagnostic infrared imaging, doppler blood flow, and echocardiography;
- e) moderate exercise, muscular strength testing, body composition assessment, and flexibility testing where appropriate given the age, weight, and health of the individual.

5. Y **Research involving materials (data, documents, records, or specimens) that have been collected, or will be collected solely for nonresearch purposes (such as medical treatment or diagnosis). (NOTE: Some research in this paragraph may be exempt from the HHS regulations for the protection of human subjects. 45 CFR 46.101(b)(4). This listing refers only to research that is not exempt.)**

6. N **Collection of data from voice, video, digital, or image recordings made for research purposes.**

7. N **Research on individual or group characteristics or behavior(including, but not limited to, research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices, and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies. (NOTE: Some research in this category may be exempt from the HHS regulations for the protection of human subjects. 45 CFR 46.101(b)(2) and (b)(3). This listing refers only to research that is not exempt.)**

Title : Stanford PHS Data Protocol: Data Exploration, Cohort Identification Educational Activities and Research in Population Sciences
Approval Period: 01/25/2019 - 04/30/2019

Resources :

a) Qualified staff.

Please state and justify the number and qualifications of your study staff.

Mark Cullen is a Professor of Medicine and has more than 20 years experience in data analyses studies for health research. Lorene Nelson is an epidemiologist and Stanford faculty with extensive methodological and data acquisition experience. Lesley Park is an Epidemiologist and Instructor in the Department of Medicine with over 10 years of research experience. Isabella Chu is a research administrator with over 10 years experience. All personnel have been extensively trained in both research methods, data security and data policies.

b) Training.

Describe the training you will provide to ensure that all persons assisting with the research are informed about the protocol and their research-related duties and functions.

All personnel approved as SPHS personnel or approved to use analytic data sets will have completed required HIPAA and human subject protection training. We will also require all personnel to sign a DUA which outlines the use of the data for a specified project and attests that they will not attempt to download data.

c) Facilities.

Please describe and justify.

We have described our Data Management Plan in great detail in the attachment in section 16.

In brief:
Sensitive and restricted data sets are received by the Stanford Principal Investigator (PI) via encrypted disk or whatever mode is preferred by the distributor of the data, and are saved to a secure server that abides by Stanford's computer and network usage policy (http://adminguide.stanford.edu/62.pdf) and information security policy (http://adminguide.stanford.edu/63.pdf). The server is behind a firewall with access restricted to authorized investigators only. Two step authentication is required at each sign on to access the data. Intrusion monitoring software is handled through the central university network office and locally as a backup. Data are encrypted and backed up nightly onsite for both onsite and offsite data storage. The server is housed in a secure location with very limited access. This location is monitored 24/7 via camera and has limited keycard access. In case of power failure both the server and network switches are tied into the natural gas generator to eliminate any down time and/or disruption in service.

A log of data files is maintained by the Manager of the PHS Data Core. The Manager will record the date and files received. For analytic data sets, the Manager will record: Project or study PI, project title, IRB approval date, date DUA was signed and received, variables included in the analytic data set and date data were made available on the server. Data access is routinely audited to ensure that there is no inappropriate access.

2. Shared File System:

All computers and workstations used to access and process analytic data sets including laptops and desktops (both office and home computers) that may be used to access participant data will be password protected, encrypted, and backed-up per Stanford policy.

No person will be permitted to download data onto a computer, laptop or workstation. All analytic data sets will be kept on the server. Investigators will access the server via a secure VPN which requires an active authorized account and two-step authentication.

Title : Stanford PHS Data Protocol: Data Exploration, Cohort Identification Educational Activities and Research in Population Sciences
Approval Period: 01/25/2019 - 04/30/2019

d) Sufficient time.

Explain whether you will have sufficient time to conduct and complete the research. Include how much time is required.

Data Core staff have 100% of their time protected for research and management of these data. Faculty (Mark Cullen, Lorene Nelson) devote about 25% of their time to Stanford PHS Data Core activities. We are adequately staffed to protect these data and accomplish the research goals.

e) Access to target population.

Explain and justify whether you will have access to a population that will allow recruitment of the required number of participants.

We have ample resources to obtain large data sets for use by Stanford researchers.

For example, to date we have obtained Truven, Optum and a 20% CMS Medicare Sample.

f) Access to resources if needed as a consequence of the research.

State whether you have medical or psychological resources available that participants might require as a consequence of the research when applicable. Please describe these resources.

This is a data analysis study. Not applicable.

g) Lead Investigator or Coordinating Institution in Multi-site Study.

Please explain (i) your role in coordinating the studies, (ii) procedures for routine communication with other sites, (iii) documentation of routine communications with other sites, (iv) planned management of communication of adverse outcomes, unexpected problems involving risk to participants or others, protocol modifications or interim findings.

1. Purpose

a) In layperson's language state the purpose of the study in 3-5 sentences.

The goal of the Stanford Center for Population Health Sciences (PHS) Data Core is to create a library of data assets that is so compelling and easy to use that it spurs trans-disciplinary population health sciences' projects across the university and between other institutions.

The purpose of this protocol is to give our investigators the ability to explore high value datasets, identify variables of interest, formulate research questions and determine potential viability of projects. In the case where datasets contain identifiers, the investigators would be required to submit an IRB Chart Review protocol upon completion of their feasibility or pilot study.

For the many datasets we host and vend which are de-identified, we will review research projects prior to granting access and then again when individuals transition from the samples (1 and 5% samples for our largest datasets) the full datasets.

b) State what the Investigator(s) hope to learn from the study. Include an assessment of the importance of this new knowledge.

We hope to learn whether providing a mechanism for data discovery prior to completing the steps for obtaining access to the full datasets will increase the use of these valuable datasets.

It is also our expectation that investigators will gain valuable insights from the use of our data. We will provide the IRB with a list of these projects, a summary of datasets used and publications on an annual

Title : Stanford PHS Data Protocol: Data Exploration, Cohort Identification Educational Activities and Research in Population Sciences
Approval Period: 01/25/2019 - 04/30/2019

basis.

c) Explain why human subjects must be used for this project. (i.e. purpose of study is to test efficacy of investigational device in individuals with specific condition; purpose of study is to examine specific behavioral traits in humans in classroom or other environment)

This research necessarily involves data derived from interactions with human beings as the research questions will focus on questions of human health and well-being.

2. Study Procedures

a) Please SUMMARIZE the research procedures, screening through closeout, which the human subject will undergo. Refer to sections in the protocol attached in section 16, BUT do not copy the clinical protocol. Be clear on what is to be done for research and what is part of standard of care.

Each dataset has different inclusion/exclusion criteria. Most of the datasets are administrative data that have been passively collected for other purposes (eg, claims). However, some datasets have been collected expressly for research purposes and we will make an effort to make these readily available to researchers (eg, HRS).

Examples of data Sets:

1. 20% Medicaid Sample or Research Identifiable files. I have attached the specifications worksheet in Section 16 which describes which data sets and years we are requesting.

2. Optum data: The Optum data are a large dataset of commercial claims for over 150 million unique members.

3. Truven data: Truven Health Analytics MarketScan® Databases are the largest of their kind in the industry with data on more than 200 million unique patients since 1995. These data contain claims from large commercially insured populations.

4. Registries from professional medical societies. These datasets will vary widely but we have requested them with full identifiers so that they can be linked to other datasets.

5. Datasets available to researchers are displayed on our data portal: phsdata.stanford.edu. In some cases, datasets in preparation are unpublished. In almost all cases, we will make the existence of the dataset possible to Stanford personnel.

Our Data Management Plan template is attached in Section 16. This outlines the general procedures for obtaining access to data and the nature of that access.

A full list of available datasets can be found on our portal at:
phsdata.stanford.edu

Procedures:
Prior to accessing any moderate or high risk dataset, every researcher associated with a project must complete:
1. A study form which will be reviewed by the Administrative Manager of the PHS Data Core.
2. The PHS Data Security Training
3. Answer questions wrt COI
4. Offer proof of encryption of all machines used to access the PHS data.
5. Have completed all other dataset specific trainings and other trainings as appropriate. This is outlined in

Title : Stanford PHS Data Protocol: Data Exploration, Cohort Identification Educational Activities and Research in Population Sciences
Approval Period: 01/25/2019 - 04/30/2019

our Data Management Plan attached in Section 16.

1. For most of our largest datasets, we will extract a large, random, representative sample (1 or 5%) for data exploration, variable identification, code refinement, feasibility and, where applicable, preliminary results.
2. Individuals who have completed the internal requirements for access to that dataset (eg, CITI, HIPAA, encryption and security training as applicable) will be given access to the sample or data as appropriate.
4. The researcher can then explore the sample for the following purposes:

A. Confirmation of the suitability of the dataset for the research question including preliminary results
B. Variable identification.
C. Code development.
D. Other activities preparatory to research.

At the point the researcher has determined that the dataset is appropriate and that they wish to access the full dataset, they will need to complete the steps for accessing that dataset at that time. These additional steps may include but are not limited to:
A. Complete a reuse agreement (eg, CMS).
B. Complete dataset specific training (eg, HCUP).
C. Full IRB or Determination if the data contain PHI or PII*
D. Payment or fees associated with data access or reuse agreements.

* For datasets containing PHI or PII, researchers will often be required to complete an IRB prior to exploration.

Procedures for trainees, students and those unfamiliar with programming, data types hosted by PHS or who otherwise require a training period are equivalent to those used by all other researchers to assess the data samples. In the case that a trainee or student has a project of sufficient merit that it is likely to result in a research product, they will be treated like any other researcher and access the full dataset through the usual channels. In almost all cases, they must have a faculty PI in order to obtain approval.

In some cases, datasets are anonymized to a sufficient degree that trainees will be permitted access to the full dataset. This designation will be made on a case by case basis and only for datasets which can be classified as low risk.

b) Explain how the above research procedures are the least risky that can be performed consistent with sound research design.

The greatest risk is in this study is data breach. Confidentiality of data will be ensured as described in the Data Management Plan attached in Section 16.

Beyond breach of confidentiality, there is no risk to subjects as these are data analysis studies on existing (mostly administrative) data.

c) State if deception will be used. If so, provide the rationale and describe debriefing procedures. Since you will not be fully informing the participant in your consent process and form, complete an alteration of consent (in section 13). Submit a debriefing script (in section 16).

Data analysis studies. Not applicable.

d) State if audio or video recording will occur. Describe what will become of the recording after use, e.g., shown at scientific meetings, erased. Describe the final disposition of the recordings.

Data analysis studies. Not applicable.

e) Describe alternative procedures or courses of treatment, if any, that might be advantageous to the participant. Describe potential risks and benefits associated with these. Any standard treatment that is being withheld must be disclosed in the consent process and form. (i.e. standard-of-care drug, different interventional procedure, no procedure or treatment, palliative care, other research studies).

Title : Stanford PHS Data Protocol: Data Exploration, Cohort Identification Educational Activities and Research in Population Sciences
Approval Period: 01/25/2019 - 04/30/2019

Data analysis studies. Not applicable.

f) Will it be possible to continue the more (most) appropriate therapy for the participant(s) after the conclusion of the study?

Data analysis studies. Not applicable.

g) Study Endpoint. What are the guidelines or end points by which you can evaluate the different treatments (i.e. study drug, device, procedure) during the study? If one proves to be clearly more effective than another (or others) during the course of a study, will the study be terminated before the projected total participant population has been enrolled? When will the study end if no important differences are detected?

Data analysis studies. Not applicable.

3. Background

a) Describe past experimental and/or clinical findings leading to the formulation of the study.

Stanford PHS hosts many valuable and proprietary datasets including Optum and Truven claims data, which combined, represent a large percentage of the commercially insured US population. We also have a 20% CMS sample and are continually working towards obtaining new datasets.

These datasets often require considerable investment in terms of trainings, applications and financial investment to access. As these data represent over a million dollar investment by the University, we want them to be as highly utilized as possible.

By allowing investigators to explore the data and conduct pilot testing of their code on a small sample of these datasets, we believe the data will be more highly utilized. This will also allow investigators wishing to do data linkage or overlay to test the feasibility of their proposed projects before investing large financial or administrative energy into obtaining data access.

In all cases, at the point the investigator decides the study will go forward and requests access to the full dataset, all of the usual requirements will be implemented at that time.

b) Describe any animal experimentation and findings leading to the formulation of the study.

Not applicable.

4. Radioisotopes or Radiation Machines

a) List all standard of care procedures using ionizing radiation (radiation dose received by a subject that is considered part of their normal medical care). List all research procedures using ionizing radiation (procedures performed due to participation in this study that is not considered part of their normal medical care). List each potential procedure in the sequence that it would normally occur during the entire study. More Info

Identify Week/Month of study	Name of Exam	Identify if SOC or Research

b) For research radioisotope projects, provide the following radiation-related information:

Identify the radionuclide(s) and chemical form(s).

For the typical subject, provide the total number of times the radioisotope and activity will be administered (mCi) and the route of administration.

Title : Stanford PHS Data Protocol: Data Exploration, Cohort Identification Educational Activities and Research in Population Sciences
Approval Period: 01/25/2019 - 04/30/2019

If not FDA approved provide dosimetry information and reference the source documents (package insert, MIRD calculation, peer reviewed literature).

c) For research radiation machine projects, provide the following diagnostic procedures:

For well-established radiographic procedures describe the exam.

For the typical subject, identify the total number of times each will be performed on a single research subject.

For each radiographic procedure, provide the setup and technique sufficient to permit research subject dose modeling. The chief technologist can usually provide this information.

For radiographic procedures not well-established, provide FDA status of the machine, and information sufficient to permit research subject dose modeling.

d) For research radiation machine projects, provide the following therapeutic procedures:

For a well-established therapeutic procedure, identify the area treated, dose per fraction and number of fractions. State whether the therapeutic procedure is being performed as a normal part of clinical management for the research participants's medical condition or whether it is being performed because the research participant is participating in this project.

For a therapeutic procedure that is not well-established, provide FDA status of the machine, basis for dosimetry, area treated, dose per fraction and number of fractions.

5. Devices

a) Please list in the table below all Investigational Devices (including Commercial Devices used off-label) to be used on participants

b) Please list in the table below all IDE Exempt Devices (Commercial Device used according to label, Investigational In Vitro Device or Assay, or Consumer Preference/Modifications/Combinations of Approved Devices) to be used on participants.

6. Drugs, Reagents, or Chemicals

a) Please list in the table below all investigational drugs, reagents or chemicals to be administered to participants.

b) Please list in the table below all commercial drugs, reagents or chemicals to be administered to participants.

7. Medical Equipment for Human Subjects and Laboratory Animals

If medical equipment used for human patients/participants is also used on animals, describe such equipment and disinfection procedures.

Data analysis studies. Not applicable.

8. Participant Population

Title : Stanford PHS Data Protocol: Data Exploration, Cohort Identification Educational Activities and Research in Population Sciences
Approval Period: 01/25/2019 - 04/30/2019

a) State the following: (i) the number of participants expected to be enrolled at Stanford-affiliated site(s); (ii)the total number of participants expected to enroll at all sites; (iii) the type of participants (i.e. students, patients with certain cancer, patients with certain cardiac condition) and the reasons for using such participants.

This is a data repository and the datasets vary in terms of their size and scope.

Marquee datasets include:
20% CMS sample (11 million lives)
Truven: 149 million lives, 8 years, commercial claims data
Optum: 58 million lives, 12 years, commercial claims data
IPUMS: Full census data from 1790 - 1940

For details on our datasets, please see the PHS Data Portal at phsdata.stanford.edu.

Data will come from a wide variety of sources including but not limited to:

1. Data which are purchased or obtained explicitly for PHS (ie, Truven, Optum).
2. Data arrangements where the Director of PHS or his representative reaches an agreement with the data owner to either allow access to the data for PHS members or for PHS to import the data. (Danish Biobank, Indian Railroad, Taiwanese health claims).
3. Data which are already on campus where an investigator wishes to make the data more broadly available. (H-CUP, CHIS, etc).

Data Core Vision and plan attached in Section 16.

b) State the age range, gender, and ethnic background of the participant population being recruited.

All ages, sexes, genders and, we hope, races and ethnicities.

c) State the number and rationale for involvement of potentially vulnerable subjects in the study (including children, pregnant women, economically and educationally disadvantaged, decisionally impaired, homeless people, employees and students). Specify the measures being taken to minimize the risks and the chance of harm to the potentially vulnerable subjects and the additional safeguards that have been included in the protocol to protect their rights and welfare.

Vulnerable subjects are either included incidentally as these are large claims datasets (eg, Truven) or data are collected on vulnerable populations explicitly (eg, March of Dimes) to address risks and needs specific to that population.

d) If women, minorities, or children are not included, a clear compelling rationale must be provided (e.g., disease does not occur in children, drug or device would interfere with normal growth and development, etc.).

Not applicable. No excluded populations.

e) State the number, if any, of participants who are laboratory personnel, employees, and/or students. They should render the same written informed consent. If payment is allowed, they should also receive it. Please see Stanford University policy.

Not applicable. Inclusion of employees in any datasets will be incidental.

f) State the number, if any, of participants who are healthy volunteers. Provide rationale for the inclusion of healthy volunteers in this study. Specify any risks to which participants may possibly be exposed. Specify the measures being taken to minimize the risks and the chance of harm to the volunteers and the additional safeguards that have been included in the protocol to protect their rights and welfare.

Health status varies widely. Many of these datasets are representative of the general population so have large percentage of data from healthy individuals included.

g) How will you identify participants for recruitment? (E.g., by: chart review; referral from treating

physician; response to ad). Attach recruitment materials in Section #16 (Attachments). All Final or revised recruitment materials, flyers, etc. must be submitted to the IRB for review and approval before use. You may not contact potential participants prior to IRB approval. See Advertisements: Appropriate Language for Recruitment Material.

Almost all data will arrive at Stanford stripped of most identifiers.

Analysts in the Stanford PHS, who regularly work with the type of data to be requested, will fulfill an "honest broker" role. They will enrich and improve the utility of datasets via data overlay and linkage, risk scores (where applicable), harmonization, organization, cleaning, and, if necessary, de-identify the pertinent data.

The Stanford PHS web portal includes data exploration, carpentry and visualization tools. For our largest datasets, individuals may select either a random, representative sample (usually 1 or 5%) which will allow investigators to explore available datasets, conduct these cursory analysis and visualization and decide which data are most appropriate for their study question.

Access to datasets will be granted to researchers who have completed the procedures outlined in Section 1.3 of the data management plan attached in Section 16.

h) Inclusion and Exclusion Criteria.

Identify inclusion criteria.

The study population will be determined by the original owner of the data. In some cases, the datasets include the entire population of a country so there are virtually no inclusion or exclusion criteria. In other cases, datasets have very specific criteria.

Identify exclusion criteria.

As above. The study population will be determined by the original owner of the data. In some cases, the datasets include the entire population of a country so there are virtually no inclusion or exclusion criteria. In other cases, for example disease specific registries, datasets have very specific criteria.

i) Describe your screening procedures, including how qualifying laboratory values will be obtained. If you are collecting personal health information prior to enrollment (e.g., telephone screening), please request a waiver of authorization for recruitment (in section 15).

Analysts in the Stanford Center for Population Health Sciences, who regularly work with the type of data to be requested, will fulfill an "honest broker" role. They will overlay, link, organize, and, if necessary, de-identify the pertinent data as well as cut the random, representative samples. Data will not be shared outside the personnel designated for each project and will only be used for approved research projects. All personnel approved as Stanford-PHS personnel or approved to use analytic data sets will have a) completed required HIPAA and human subject protection training as applicable as well as a PHS specific security training; and b) have "need to know" status with regard to the data and cannot practically work on the project without it and c) if applicable, have been approved to access analytic data sets by the original owner of the data and the Stanford IRB.

In the event of an employee leaving Stanford of their own volition or being terminated their access to the data will be terminated.

DATA REPORTING AND PUBLICATION

Stanford does not disclose any non-aggregated information derived from analyses including listings, or information derived from the file(s) with or without direct identifiers, if such findings, listings, or information can, by themselves or in combination with other data, be used to deduce an individual's identity. Additionally, Stanford will not identify or report any identifiable pharmacy, provider or prescriber in any publication. All findings will be reported in aggregate. Tables or outputs with cell sizes less than 10 will not be published or made public.

Title : Stanford PHS Data Protocol: Data Exploration, Cohort Identification Educational Activities and Research in Population Sciences
Approval Period: 01/25/2019 - 04/30/2019

Stanford agrees that any use of data in the creation of any document (manuscript, table, chart, study, report, etc.) will adhere to the distributor of the data's current cell size suppression policy or Stanford's cell suppression policy, whichever is more restrictive. No cell (e.g. admittances, discharges, patients, services) 10 or less may be displayed or used in any publication. Also, no use of percentages or other mathematical formulas will be used if they result in the display of a cell 10 or less.

COMPLETION OF RESEARCH TASKS AND DATA DESTRUCTION

The Stanford Principal Investigator must be included as Other Personnel on all IRB applications (e-Protocol) for projects which wish to use the data.

The Stanford e-Protocol system keeps records on study personnel, which role they have, protocol or personnel changes (on an annual basis or as they occur, whichever is more frequent), and conflict of interest. In addition to listing relevant personnel on IRB protocols, the e-protocol system also confirms that required HIPAA and privacy trainings are completed and up to date. In the event research projects are completed, the Stanford PI records the completion of the project and the disposition of data access as terminated. The Principal Investigator also notes protocol closures in the e-protocol system.

j) Describe how you will be cognizant of other protocols in which participants might be enrolled. Please explain if participants will be enrolled in more than one study.

Data analysis studies. Not applicable.

k) Payment/reimbursement. Explain the amount and schedule of payment or reimbursement, if any, that will be paid for participation in the study. Substantiate that proposed payments are reasonable and commensurate with the expected contributions of participants and that they do not constitute undue pressure on participants to volunteer for the research study. Include provisions for prorating payment. See payment considerations

Data analysis studies. Not applicable.

l) Costs. Please explain any costs that will be charged to the participant.

Data analysis studies. Not applicable.

m) Estimate the probable duration of the entire study. Also estimate the total time per participant for: (i) screening of participant; (ii) active participation in study; (iii) analysis of participant data.

This is a protocol for exploratory work and data preparation. i) we anticipate data updates at least annually for most datasets. ii) We anticipate that the repository will remain at Stanford for a minimum of five years. iii) Protocols for analyses will continue for a minimum of one year and will be submitted separately to the IRB.

9. Risks

a) For the following categories include a scientific estimate of the frequency, severity, and reversibility of potential risks. Wherever possible, include statistical incidence of complications and the mortality rate of proposed procedures. Where there has been insufficient time to accumulate significant data on risk, a statement to this effect should be included. (In describing these risks in the consent form to the participant it is helpful to use comparisons which are meaningful to persons unfamiliar with medical terminology.)

The risks of the Investigational devices.

Data analysis studies. Not applicable.

The risks of the Investigational drugs. Information about risks can often be found in the Investigator's brochure.

Data analysis studies. Not applicable.

Title : Stanford PHS Data Protocol: Data Exploration, Cohort Identification Educational Activities and Research in Population Sciences
Approval Period: 01/25/2019 - 04/30/2019

The risks of the Commercially available drugs, reagents or chemicals. Information about risks can often be found in the package insert.

Data analysis studies. Not applicable.

The risks of the Procedures to be performed. Include all investigational, non-investigational and non-invasive procedures (e.g., surgery, blood draws, treadmill tests).

Analysts in the Stanford Center for Population Health Sciences, who regularly work with the type of data to be requested, will fulfill an "honest broker" role. They will cut the random samples, link, organize, and, if necessary, de-identify the pertinent data. Data will not be shared outside the personnel designated for this project and will only be used for cohort discovery and code and tool testing as described in section 2. All personnel approved as Stanford-PHS personnel or approved to use analytic data sets will have a) completed required HIPAA and human subject protection training; and b) have "need to know" status with regard to the data and cannot practically work on the project without it and c) if applicable, have been approved to access analytic data sets by the original owner of the data and the Stanford IRB.

In the event of an employee leaving Stanford of their own volition or being terminated their access to the data will be terminated.

DATA REPORTING AND PUBLICATION

Stanford does not disclose any non-aggregated information derived from analyses including listings, or information derived from the file(s) with or without direct identifiers, if such findings, listings, or information can, by themselves or in combination with other data, be used to deduce an individual's identity. Additionally, Stanford will not identify or report any identifiable pharmacy, provider or prescriber in any publication. All findings will be reported in aggregate. Tables or outputs with cell sizes less than 10 will not be published or made public.

Stanford agrees that any use of data in the creation of any document (manuscript, table, chart, study, report, etc.) will adhere to the distributor of the data's current cell size suppression policy or Stanford's cell suppression policy, whichever is more restrictive. No cell (e.g. admittances, discharges, patients, services) 10 or less may be displayed or used in any publication. Also, no use of percentages or other mathematical formulas will be used if they result in the display of a cell 10 or less.

COMPLETION OF RESEARCH TASKS AND DATA DESTRUCTION

The Stanford Principal Investigator must be included as Other Personnel on all IRB applications (e-Protocol) for projects which wish to use data containing PHI or PII.

The Stanford e-Protocol system keeps records on study personnel, which role they have, protocol or personnel changes (on an annual basis or as they occur, whichever is more frequent), and conflict of interest. In addition to listing relevant personnel on IRB protocols, the e-protocol system also confirms that required HIPAA and privacy trainings are completed and up to date. In the event research projects are completed, the Stanford PI records the completion of the project and the disposition of data access as terminated. The Principal Investigator also notes protocol closures in the e-protocol system.

For research on de-identified data, PHS will track the study team, study title and research question(s), datasets and other relevant information.

The risks of the Radioisotopes/radiation-producing machines (e.g., X-rays, CT scans, fluoroscopy) and associated risks.

Data repository protocol. Not applicable.

The risks of the Physical well-being.

We do not anticipate that this study will impact the physical well-being of participants.

The risks of the Psychological well-being.

We do not anticipate that this study will impact the psychological well-being of participants.

The risks of the Economic well-being.

We do not anticipate that this study will impact the economic well-being of participants.

The risks of the Social well-being.

We do not anticipate that this study will impact the social well-being of participants.

Overall evaluation of Risk.

Low - innocuous procedures such as phlebotomy, urine or stool collection, no therapeutic agent, or safe therapeutic agent such as the use of an FDA approved drug or device.

b) If you are conducting international research, describe the qualifications/preparations that enable you to both estimate and minimize risks to participants. Also complete the International Research Form and attach it in the Attachments section. If not applicable, enter N/A.

In the case where we have international data partners, we will adhere to all data governance laws and policies of the host country and institution.

c) Describe the planned procedures for protecting against and minimizing all potential risks. Include the means for monitoring to detect hazards to the participant (and/or to a potential fetus if applicable). Include steps to minimize risks to the confidentiality of identifiable information.

Please see the detailed description provided in the data protection plan attached in Section 16. In brief:

1. For the most part, data will be de-identified before being transmitted to Stanford. 2. Storage and distribution of data will adhere to Stanford's restricted data policies: http://adminguide.stanford.edu/62.pdf and information security policy http://adminguide.stanford.edu/63.pdf.

d) Explain the point at which the experiment will terminate. If appropriate, include the standards for the termination of the participation of the individual participant Also discuss plans for ensuring necessary medical or professional intervention in the event of adverse effects to the participants.

Data analysis studies. Not applicable.

e) Special Participant Populations

Children's Findings OHRP. As children are involved in your research, please select one or more regulatory categories (46.404 through 46.407) below that your research falls under and provide the necessary rationale for each determination. See full regulation citation.

Y 46.404 Research not involving greater than minimal risk. The research must present no greater than minimal risk to children and adequate provisions must be made for soliciting the assent of the children and the permission of their parents or guardians. Please provide rationale for the above statement.

Rationale for category selected above

We intend to collect and store datasets which may contain information about children. We will make limited datasets available to researchers who have an IRB approval to do the research and have completed all required encryption and trainings as outlined in the Data Management Plan in section 1.3.

10. Benefits

a) Describe the potential benefit(s) to be gained by the participants or by the acquisition of important knowledge which may benefit future participants, etc.

No direct benefits will accrue to the study subjects. These analyses may lead to a greater understanding of health and well-being.

11. Privacy and Confidentiality

Title : Stanford PHS Data Protocol: Data Exploration, Cohort Identification Educational Activities and Research in Population Sciences
Approval Period: 01/25/2019 - 04/30/2019

Privacy Protections

a) Describe how the conditions under which interactions will occur are adequate to protect the privacy interests of participants (e.g., privacy of physical setting for interviews or data collection, protections for follow-up interactions such as telephone, email and mail communications).

Data analysis studies. Not applicable. For data protection, please see the Data Management Plan in Section 16.

Confidentiality Protections

b) Specify PHI (Protected Health Information). PHI is health information linked to HIPAA identifiers (see above). List BOTH health information AND HIPAA identifiers. If you are using STARR, use the Data Privacy Attestation to ensure that your request will match your IRB-approved protocol. Be consistent with information entered in section 15a.

Datasets will vary widely in the PHI or PII fields available. PHI/PII fields include but are not limited to:

Datasets will vary widely in the PHI fields they make available. PHI/PII fields include but are not limited to:
· Names
· Social Security numbers
· Telephone numbers
· Geographic subdivisions smaller than a State, including street address, city, county, precinct, zip code, and their equivalent geocodes, census tracts and similar, geographic units
· Dates directly related to an individual, including birth date, admission date, discharge date, date of death; ages over 89 other elements of dates (including year) indicative of such age
· Fax numbers (though this will be exceedingly rare)
· Electronic mail addresses
· Medical record numbers
· Social Security numbers
· Health plan beneficiary numbers
· Account numbers
· Certificate/license numbers
· Vehicle identifiers and serial numbers, including license plate numbers
· Device identifiers and serial numbers
· Web Universal Resource Locations (URLs).
· Internet Protocol (IP) address numbers
· Biometric identifiers, including finger and voice prints
· Full face photographic images and any comparable images; and
· Other unique identifying number, characteristics or codes

Demographic PHI data listed above. Where available we will also include or incorporate sociodemographic variables and neighborhood characteristics.

Health encounter data including but not limited to: ICD-9 codes, dates of service (month and year), location of service, specialty of physician, pharmacy data). Claims data include dental, inpatient and outpatient encounters.

Electronic Health records including but not limited to: lab values, procedures, vital signs and flow sheet variables, assessments, clinic and inpatient notes, medications including dose and timing, symptoms, radiology images and reports etc.

Where available, we will also collect information on practice patterns including referral patterns and cost.

Where available, we will also incorporate behavioral data.

c) You are required to comply with University Policy that states that ALL electronic devices: computers (laptops and desktops; OFFICE or HOME); smart phones; tablets; external hard disks, USB drives, etc. that may hold identifiable participant data will be password protected, backed up, and encrypted.

Title : Stanford PHS Data Protocol: Data Exploration, Cohort Identification Educational Activities and Research in Population Sciences
Approval Period: 01/25/2019 - 04/30/2019

See http://med.stanford.edu/datasecurity/ for more information on the Data Security Policy and links to encrypt your devices.

Provide any additional information on ALL data security measures you are taking. You must use secure databases such as RedCap https://med.stanford.edu/researchit/infrastructure/redcap.html https://med.stanford.edu/researchit/infrastructure/redcap.html. If you are unsure of the security of the system, check with your Department IT representative. Please see http://med.stanford.edu/irt/security/ for more information on IRT Information Security Services and http://www.stanford.edu/group/security/securecomputing/mobile_devices.html for more information for securing mobile computing devices. Additionally, any PHI data on paper must be secured in an locked environment.

By checking this box, You affirm the aforementioned. Y

For a detailed description, please see the Data Management Plan attached in Section 16.

Sensitive and restricted data sets are received by the Stanford Principal Investigator (PI) via encrypted disk, the mode preferred by the distributor of the data, and are saved to a secure server that abides by Stanford's computer and network usage policy (http://adminguide.stanford.edu/62.pdf) and information security policy (http://adminguide.stanford.edu/63.pdf). The server is behind a firewall with access restricted to authorized investigators only. Two step authentication is required at each sign on to access the data. Intrusion monitoring software is handled through the central university network office and locally as a backup. Data are encrypted and backed up nightly onsite for both onsite and offsite data storage. The server is housed in a secure location with very limited access. This location is monitored 24/7 via camera and has limited keycard access. In case of power failure both the server and network switches are tied into the natural gas generator to eliminate any down time and/or disruption in service.

Detailed records will be logged as described in the attached Data Management Plan.

d) Describe how data or specimens will be labeled (e.g. name, medical record number, study number, linked coding system) or de-identified. If you are de-identifying data or specimens, who will be responsible for the de-identification? If x-rays or other digital images are used, explain how and by whom the images will be de-identified.

Most data will come to PHS de-deidentified by the data vendor. In the case where PHS receives data with identifiers, where practical, PHS will construct deidentified analytic datasets for distribution.

e) Indicate who will have access to the data or specimens (e.g., research team, sponsors, consultants) and describe levels of access control (e.g., restricted access for certain persons or groups, access to linked data or specimens).

Only the research team or approved investigators will have access to the data.

f) If data or specimens will be coded, describe the method in which they will be coded so that study participants' identities cannot be readily ascertained from the code.

The data will be deidentified either by the data vendor or by PHS staff in whatever manner is appropriate for the dataset and research project.

g) If data or specimens will be coded, indicate who will maintain the key to the code and describe how it will be protected against unauthorized access.

In the event we have access to personal identifiers such as name or MRN and data are coded, the data manager(s) will hold the codes and the keys. No other individuals will have access to the codes.

For the most part, the data owners (such as Optum, Truven and CMS) will hold the identities and data will arrive deidentified. In such cases, the data owners will retain the code.

h) If you will be sharing data with others, describe how data will be transferred (e.g., courier, mail) or transmitted (e.g., file transfer software, file sharing, email). If transmitted via electronic networks, describe how you will secure the data while in transit.See http://www.stanford.edu/group/security/securecomputing/.

http://www.stanford.edu/group/security/securecomputing/. Additionally, if you will be using or sharing PHI see https://uit.stanford.edu/security/hipaa https://uit.stanford.edu/security/hipaa.

Please see the detailed description in the Data Management Plan attached in Section 16. In brief:

No person will be permitted to download data onto a computer, laptop or workstation. All analytic data sets will be kept on the server. Investigators will access the server via a secure VPN which requires an active authorized account and two-step authentication.

i) How will you educate research staff to ensure they take appropriate measures to protect the privacy of participants and the confidentiality of data or specimens collected (e.g. conscious of oral and written communications, conducting insurance billing, and maintaining paper and electronic data)?

All researchers are required to complete HIPAA and CITI training in order to access data containing PHI or PII. Additionally, in order to obtain data sets, researchers will be required to sign a DUA indicating they understand they are not permitted to download data. Ever.

12. Potential Conflict of Interest

Investigators are required to disclose any financial interests that reasonably appear to be related to this protocol.

Financial Interest Tasks

Investigators	Role	Email	Has Financial Interest?	Date Financial Interest Answered	Date OPACS Disclosure Submitted	Date OPACS Review Completed
Mark Cullen	PD	mrcullen@stanford.edu	N	04/06/2017		
Lorene Nelson	COPD	lnelson@stanford.edu	N	03/10/2017		
Gary Darmstadt	OP	gdarmsta@stanford.edu	N	09/28/2018		
Lisa Chamberlain	OP	lchamberlain@stanford.edu	N	09/29/2018		

13. Consent Background

13. 1 Waiver of Consent Waiver of Consent

1) Y **The research involves no more than minimal risk to the participants.**

This protocol is for preliminary analyses of existing administrative, EMR and claims data. As these data have already been collected are often deidentified or anonymized.As these are data analysis studies (exploratory) we believe these are minimal risk studies.

2) Y **The waiver or alteration will not adversely affect the rights and welfare of the participants.**

There are study procedures are in place to protect confidentially as described in our data management

Title : Stanford PHS Data Protocol: Data Exploration, Cohort Identification Educational Activities and Research in Population Sciences
Approval Period: 01/25/2019 - 04/30/2019

plan in section 16. Information learned during the study will not affect the treatment of the participants who had infections in the pasts and thus will not adversely affect their welfare.

3a) Y **The research could not practicably be carried out with out the requested waiver or alteration.**

3b) **For research using identifiable private information or identifiable biospecimens, the research could not practicably be carried out without using such information or biospecimens in an identifiable format.**

If the IRB required informed consent of participants, this research would be impossible because the data are generally deidentified and reidentification (and certainly patient contact) are prohibited. In the rare cases where data are not deidentified, participant contact is prohibited or impractical due to the size of datasets.

4) Y **Whenever appropriate, the participants or legally authorized representatives will be provided with additional pertinent information after participation.**

We do not expect that any participants will be contacted directly but we hope these research studies will inform best practices and policies.

14. Assent Background (less than 18 years of age)

14. 1 Waiver of Assent Waiver of Assent

Address the following four regulatory criteria for a waiver of assent and provide a protocol-specific reasons for each:

Y **The research involves no more than minimal risk to the participants.**

This protocol is for preliminary analyses of existing administrative, EMR and claims data. As these data have already been collected are often deidentified or anonymized and these are exploratory analyses, we believe these are minimal risk studies.

Y **The waiver will not adversely affect the rights and welfare of the participants.**

There are study procedures are in place to protect confidentially as described in our data management plan in section 16. Information learned during the study will not affect the treatment of the participants who had infections in the pasts and thus will not adversely affect their welfare.

Y **The research could not practicably be carried out without the requested waiver.**

For research using identifiable private information or identifiable biospecimens, the research could not practicably be carried out without using such information or biospecimens in an identifiable format

If the IRB required informed consent of participants, this research would be impossible because the data are generally deidentified and reidentification (and certainly patient contact) are prohibited. In the rare cases where data are not deidentified, participant contact is prohibited or completely impractical.

Y **Whenever appropriate, the participants or legally authorized representatives will be provided with additional pertinent information after participation.**

We do not expect that any participants will be contacted directly but we hope these research studies will inform best practices and policies.

15. HIPAA Background

15. 1 Waiver of Authorization waiver of authorization

a) **Describe the Protected Health Information (PHI) needed to conduct the research. PHI is health information linked to HIPAA identifiers. List BOTH health information AND HIPAA identifiers. If**

you are using STRIDE, use the Data Privacy Attestation to ensure that your request will match your IRB-approved protocol.

This is a protocol for data hosting and exploration - activities preparatory to research. We will have hundreds of databases each with different fields. PHI may include, but is not limited to: claims, dates of service, demographics, health conditions, procedure codes and in some cases identifiable data such as name and address (eg, Census).

b) Please Answer:

Y **Do you certify that the use or disclosure of protected health information involves no more than a minimal risk to the privacy of individuals?**

Y **Do you certify that the research could not practically be conducted with out the waiver?**

Y **Do you certify that you have adequate written assurances that the protected health information will not be reused or disclosed to any other person or entity, except as required by law, for authorized oversight of the research project, or for other research for which the use or disclosure of protected health information would be permitted?**

Y **Do you certify that the research could not practically be conducted with out access to and use of the protected health information?**

c) **Please describe an adequate plan to protect any identifiers from improper use and disclosure.**

In almost all cases data will arrive at Stanford deidentified. We have attached a detailed description of how we will protect the data and deal with identifiers and de-identification in our DMP in Section 16.

d) **Please describe an adequate plan to destroy the identifiers at the earliest opportunity consistent with conduct of the research, unless there is a health or research justification for retaining the identifiers or such retention is otherwise required by law.**

As the data are maintained by a data repository, we hope to obtain and curate datasets valuable and useful enough that they would not be destroyed.

16. Attachments

Attachment Name	Attached Date	Attached By	Submitted Date
Stanford_DMP_General_01Jan19	01/22/2019	itaylor	

Obligations

The Protocol Director agrees to:

- Adhere to principles of sound scientific research designed to yield valid results
- Conduct the study according to the protocol approved by the IRB
- Be appropriately qualified to conduct the research and be trained in Human Research protection, ethical principles, regulations, policies and procedures
- Ensure all Stanford research personnel are adequately trained and supervised
- Ensure that the rights and welfare of participants are protected including privacy and confidentiality of data
- Ensure that, when de-identified materials are obtained for research purposes, no attempt will be made to re-identify them.

Title : Stanford PHS Data Protocol: Data Exploration, Cohort Identification Educational Activities and Research in Population Sciences
Approval Period: 01/25/2019 - 04/30/2019

- Disclose to the appropriate entities any potential conflict of interest
- Report promptly any new information, modification, or unanticipated problems that raise risks to participants or others
- Apply relevant professional standards.

Any change in the research protocol must be submitted to the IRB for review prior to the implementation of such change. Any complications in participants or evidence of increase in the original estimate of risk should be reported at once to the IRB before continuing with the project. Inasmuch as the Institutional Review Board (IRB) includes faculty, staff, legal counsel, public members, and students, protocols should be written in language that can be understood by all Panel members. The investigators must inform the participants of any significant new knowledge obtained during the course of the research.

IRB approval of any project is for a maximum period of one year. For continuing projects and activities, it is the responsibility of the investigator(s) to resubmit the project to the IRB for review and re-approval prior to the end of the approval period. A Notice to Renew Protocol is sent to the Protocol Director 7 weeks prior to the expiration date of the protocol.

Department Chair must approve faculty and staff research that is not part of a sponsored project. VA applicants must have Division Chief or Ward Supervisor approval. E-mail the Department Chair approval to IRBCoordinator@lists.stanford.edu.

All data including signed consent form documents must be retained for a minimum of three years past the completion of the research. Additional requirements may be imposed by your funding agency, your department, or other entities. (Policy on Retention of and Access to Research Data, Research Policy Handbook, http://doresearch.stanford.edu/policies/research-policy-handbook/conduct-research/retention-and-access-research-data)

PLEASE NOTE: List all items (verbatim) that you want to be reflected in your approval letter (e.g., Amendment, Investigator's Brochure, consent form(s), advertisement, etc.) in the box below. Include number and date when appropriate.

Protocol, Data Management Plan

Y By checking this box, I verify that I, as the Protocol Director (PD) responsible for this research protocol, have read and agree to abide by the above obligations, or that I have been delegated authority by the PD to certify that the PD has read and agrees to abide by the above obligations.