

Colorado All Payer Claims Database Data Release Application

Thank you for your interest in obtaining data from the CO APCD. As you fill out this application, please let us know if you have any questions or concerns by reaching out to ColoradoAPCD@civhc.org. We are here to help!

Also, please be aware that if you are requesting Protected Health Information (PHI), your request requires a recommendation for approval by the Data Release Review Committee (DRRC). Data elements that are considered PHI under HIPAA are indicated below. If PHI is requested, a CIVHC Account Executive will help you successfully complete an application and navigate the DRRC process.

Please use this application to submit information regarding your request for data from the Colorado All Payer Claims Database (CO APCD). This information will help the Center for Improving Value in Health Care (CIVHC), the Administrator of the CO APCD, answer any questions you have regarding your data request and assist us in helping you complete the data application form.

Note: Please reference the CO APCD Data Elements Request Form found at <http://www.civhc.org/get-data/data-release/> when completing this form.

Introduction: Section 10 CCR 2505-5-1.200.5 describes how the CO APCD Administrator addresses Requests for Data and Reports:

1.200.5.A. A state agency or private entity engaged in efforts to improve health care or public health outcomes for Colorado residents may request a specialized report from the CO APCD by submitting to the administrator a written request detailing the purpose of the project, the methodology, the qualifications of the research entity, and by executing a Data Use Agreement (DUA), to comply with the requirements of HIPAA.

1.200.5. B. A data release review committee shall review the request and advise the administrator on whether release of the data is consistent with the statutory purpose of the CO APCD, will contribute to efforts to improve health care for Colorado residents, and complies with the requirements of HIPAA. The administrator shall include a representative of a physician organization, hospital organization, non-physician provider organization and a payer organization on the data release review committee.

This Data Release Application serves as the written request for information noted in section 1.200.5.A.

PART ONE

Project Information	
Project Title:	21.116 COVReD: COVID-19 Real World Data – Unraveling causes, inequities and risk for death and disability in the era of COVID-19
Date:	April 1, 2021
Organization Requesting Data:	The Board of Trustees of the Leland Stanford Junior University
Contact Person:	Michala Welch
Title:	Senior Contract and Grant Officer
E-mail:	welchmi6@stanford.edu
Phone Number:	(o) 650-736-7736
Person Responsible for the Project (if different than above):	PI David Rehkopf, ScD
Title:	Faculty Director, Stanford Center for Population Health Sciences Associate Professor, Epidemiology and Population Health Associate Professor, Medicine - Primary Care and Population Health
E-mail:	drehkopf@stanford.edu
Phone Number:	650-725-0356

Project Purpose:

Project questions to be discussed with client representative:

- Please describe your project and project goals/objectives.

Introduction

SARS-CoV-2 (“COVID-19”) infection has diverse disease trajectories with tremendous variation in incidence, morbidity, mortality, symptomatology, length of disease,^{1–3} and symptoms emerging after apparent resolution. There are still many unknowns with regard to disease epidemiology, disparities and long-term outcomes for both direct and indirect impacts of COVID-19.

A concerning development is the large number of patients (between 10 - 30%) reporting symptoms for a significant length of time after acute COVID-19 infection.⁴ These symptoms can range from mild to debilitating. The course of **Post-Acute Sequelae of SARS-CoV-2 (“PASC”)** or **“long haul COVID-19”** remains unknown. There is an urgent need to characterize the trajectory of long haul COVID-19, from its risk factors, incidence and prevalence to its phenotypic signatures and prognosis.

Over and above the sharp rise in deaths due to COVID-19, another concerning phenomenon is a steep increase in all-cause mortality *over and above the large surge in* COVID-19 deaths. National estimates are generally around 20% +/-2%. The etiology of this increase in mortality is of yet, unknown. It may be due to an undercount COVID-19 deaths, the results of avoidance of care, an increase in events related to mental health or some other combination of factors. We seek to both

clearly understand what the actual increase in mortality is, characterize the epidemiology of these deaths and to try and untangle potential causes and factors which may put Colorado residents at increased risk of death.

Project Objectives:

There are two broad based objectives associated with this research project. First, our research will investigate the epidemiology and clinical course of COVID-19 and long haul COVID-19 in Colorado. This will include an in depth investigation of COVID-19 variants and vaccines and both COVID-19 and long haul COVID-19 treatments and therapeutic agents.

Second, our research will investigate the observed spike in all-cause mortality, the epidemiology and etiology of these deaths and attempt to disentangle their relationship to COVID-19. Estimates of the increase in all-cause mortality during the COVID pandemic are around 20% nationally.⁵⁻⁷ Of yet, there are not stable estimates of the increased mortality in Colorado, nor is it clear what the etiology of these deaths is, what factors put Colorado residents at increased risk nor is there an understanding of the relationship between these deaths and COVID-19.

Data Requested:

In order to accomplish these aims, Stanford will require access to the Colorado APCD data including person level identifiers for both individual level linkages and imputation of some variables. We are interested in all available years of data and all lines of business.

The reasons for requiring the full cohort are twofold.

First, COVID-19 is, of yet, still poorly understood with reported involvement of nearly every organ system. Long haul COVID-19 has followed an even more unusual trajectory with many patients who did not initially have serious disease who later report debilitating symptoms with a very wide range of systems. We need comprehensive data in order to properly investigate the clinical course and epidemiology of COVID-19.

The second reason for requesting all data relates to the large and unexplained spike in all-cause mortality. In order to understand the clinical course and epidemiology of COVID-19 and the excess deaths, it will be necessary to construct multiple control groups. We are, of yet, unsure what these will be and it is likely that we will employ unsupervised machine learning methods to identify patterns. In order to accomplish this, we will need the full cohort.

The reason for requiring patient and physician identifiers (e.g., name) will be the imputation of race and ethnicity. As noted above, there is large variation in the incidence and outcomes of COVID-19, long COVID-19 and excess mortality. In order to accurately characterize these patterns it will be necessary to impute race and ethnicity where a more reliable source of this information is unavailable.

As noted, there has been a substantial increase in all-cause mortality in 2020 and 2021. One hypothesis is that individuals were less likely to seek care for ongoing serious medical conditions. Concerns around COVID-19 exposure is a likely explanation for some of this. However, concerns

around cost of care may also play an important role due to financial stress related to job loss or economic uncertainty. We are requesting cost and financial data to disentangle the role that concerns over ability to pay may play in avoidance of necessary medical care.

We are also seeking to characterize the costs of care for COVID-19 and long haul COVID-19 for individuals suffering from these diseases as they may play an important role in the future financial wellbeing of impacted individuals. Research on potential treatments will focus on both efficacy and value for patient wellbeing.

It will be essential to enhance the CO APCD datasets in the domains of demographics, socioeconomic status (race/ethnicity) and cause and manner of death. We describe in this CO APCD Application our proposed methods and the data sources we intend to link with the CO APCD.

Stanford also has access to American Family Cohort (AFC) national data set. This data set is complete with very detailed and rich EMR information (e.g. vital signs, medical histories, patient demographics, lab and test results, etc.). We have identified 214,000 Coloradans as a part of the national AFC data set. We plan to provide a Finder File to CIVHC to complete a match for these 214,000 Coloradans. Enhancing the data on these 214,000 Coloradans with CO APCD will provide a very rich data source which we believe will be significant in addressing our specific research questions.

Data hosted at Stanford will be managed in accordance with the attached data management plan (Appendix 3).

- What specific research question(s) are you trying to answer or problem(s) are you trying to solve with this data request? (Please list and number the individual questions.)

Specific Research Questions

1. What is the incidence, prevalence, disease characteristics, and trajectory of COVID-19 and long haul COVID-19 disease in Colorado?
 - a. What are the associated risk factors and prognosis?
2. What is the incidence, prevalence, disease characteristics, and trajectory of COVID-19 variants in Colorado?
 - a. What are the associated risk factors and prognosis for these variants?
3. What is the impact of the COVID vaccine(s) on outcomes including around morbidity, long haul COVID, and mortality?
 - a. Will include an examination of vaccine type and date(s) of administration as well as variants
4. What is the epidemiology and etiology of an increase in deaths among certain populations (for example, deaths associated with cardiovascular disease, which increased 5% during the pandemic) and is the observed increase associated with or correlated to COVID-19?
5. What treatments and therapeutic agents are being utilized with COVID-19 and long haul COVID-19 patients?
 - a. Will include a review of safety and efficacy of off-label (and fast track approved) therapies

6. What are the factors that have contributed to the increase in all-cause mortality in Colorado during the recent pandemic? How has COVID-19 contributed to the increase in all-cause mortality in Colorado?

Note: For the above research questions, it is essential to pay close attention to disparities, mitigating factors, and immediate as well as long-term impacts in vulnerable populations including an examination of all research questions by race and ethnicity. Enhancing the CO APCD data set with socioeconomic data is a critical part of our project plan.

Additional Supporting Information Relative to the Specific Research Questions Above:

Part I: Data Enhancements: Race, Ethnicity and Socioeconomic Data

Obtaining accurate and complete information on race and ethnicity is hugely important for the evaluation of disparities in disease incidence, trajectory, outcomes and mitigation as has been recently demonstrated in the COVID-19 pandemic. As a practical matter, the absence of race, ethnicity and other protected attributes can often make it hard to estimate the rate of disparities or evaluate equity in the administration of programs. President Biden's Executive Order⁸, for instance, mandates assessments of racial differences in all federal programs and policies, but protected attributes are often not collected by state and federal agencies. Consequently, this information is most frequently obtained through record linkage or Bayesian Improved Surname Geocoding.⁹

It is also probable information on race and ethnicity is not missing at random and demographics at higher risk of adverse outcomes may also be either less likely to report race and ethnicity or have more complex demographic makeup.¹⁰⁻¹² For example, Labgold et al reported that racial disparities of SARS CoV-2 case rates in Georgia were an order of magnitude higher when information on race and ethnicity was imputed.¹¹ Ensuring that this information is complete and accurate is an important step towards measuring and addressing disparities in not just COVID-19 outcomes, but all manner of health disparities.

Aim 1: Stanford will use a combination of imputation and linkage to append demographic information to the CO APCD dataset. **Aim 1A.** Stanford will use natural language processing (NLP) and machine learning (ML) methods to impute and append race and ethnicity to each record based on a number of factors. This information will be appended and tagged as imputed. Outcomes using this information will be compared with data where race/ethnicity is based on linkages to datasets where race and ethnicity is reliably reported (e.g. the **American Family Cohort - AFC**). **Aim 1B.** Stanford will use external datasets, which contain reliable information on race, ethnicity and socioeconomic status to append this information to the CO APCD. The first of these linkages will be the American Family Cohort. **Aim 1C.** We will append a small area measure of social deprivation (eg, PHATE or other similar index as mutually decided on with the state of Colorado). Details on the AFC and the proposed linkage can be found in Appendix 2.

Aim 2: SARS CoV-2 has directly and indirectly resulted in a large number of deaths, with recent work showing COVID-19 was the third leading cause of death in 2020.¹³ In addition, there have been substantial increases in all cause mortality with estimates of around 20% nationally. some

other causes of death, including an almost 5% increase in Cardiovascular disease deaths. As of yet, it is unclear what the epidemiology and etiology of these additional deaths is, whether due to underreporting of SARS CoV-2, deaths related to deferment of needed care or other factors. In order to reduce disparities in excess mortality it is first necessary to understand which populations are disproportionately impacted and what other information may be available to help explain the elevated number of deaths. In order to accomplish this aim, Stanford will enhance the existing CO APCD linkage to vital records by extracting information from obituaries and using ML techniques to identify patterns in causes and manner of death. Obituaries will be used to both enhance the sociodemographic data available to the research team (Aim 1) as well as potentially clarify causes of death where none is given. To accomplish Aim 2, Stanford will require both birth and death records linked to the CO APCD for probabilistic linkages.

Aim 2A. Stanford will either purchase or scrape obituary data and extract dates of birth, death, cause and manner of death, demographic data. These variables will be appended to the data and their provenance noted. Concordance and discordance between dates, causes and manner of death between obituaries and official records will be noted. **Aim 2B.** Stanford will describe the demographics of populations which are likely to be included in the obituary or similar records and note the degree to which these records are a reliable source of mortality information and where they have explanatory power for excess mortality. It is likely that populations which are excluded from other types of administrative data, including all payer claims, are less likely to receive obituaries. As such, other sources of information on these populations must be explored. **Aim 2C.** We will characterize progression and patterns of cause mortality including distribution by age, race, ethnicity and socioeconomic status and map causes where available. **Aim 2D:** Where data are available, we will attempt to determine likely cause of death. For example, previously healthy individuals diagnosed with COVID-19 Like Illness early in the pandemic who later expired may represent undercounts of COVID-19 deaths.

Part II: Proposed Aims SARS CoV-2 and long haul COVID-19

Specific Aims SARS CoV-2 and Post Acute SARS CoV-2 (PASC)

SARS-CoV-2 (COVID-19) infection has diverse disease trajectories with tremendous variation in incidence, morbidity, mortality, symptomatology, length of disease,¹⁻³ and symptoms emerging after apparent resolution. There are still many unknowns with regard to disease epidemiology, disparities and long term outcomes for both direct and indirect impacts of COVID-19.

A concerning development is the large number of patients (between 10 - 30%) reporting symptoms for a significant length of time after acute COVID-19 infection.⁴ These symptoms can range from mild to debilitating. The course of Post-Acute Sequelae of SARS-CoV-2 (PASC) or, long haul COVID-19 remains unknown. There is an urgent need to characterize the trajectory of long haul COVID-19, from its risk factors, incidence and prevalence to its phenotypic signatures and prognosis.

Real-world data such as that contained in the CO APCD can complement clinical trial evidence to address novel questions identified by the NIH as critical for understanding long haul COVID-19 trajectory, course, and risks.

Our study proposes investigation of three aspects of COVID-19 and long haul COVID-19:

Aim 3: Characterize the epidemiology of COVID-19 and long haul COVID-19 with particular attention to disparities, mitigating factors and immediate and long-term impacts in vulnerable populations.

Aim 3A. Stanford will explore essential data on SARS-CoV-2 epidemiology, variants, vaccination (including vaccine type and data of administration) and outcomes around morbidity, particularly long haul COVID-19 and mortality. This will aid in understanding the frequency and severity of new variants and the effectiveness of different vaccines in protecting against COVID-19 infection. **Aim 3B.** Create sample weights so that findings from the CO APCD data are more generalizable to the Colorado and U.S. population. This will be at the individual and census tract level and will be standardized to distributions of the Colorado and U.S. population. Will use individual data on age, race (imputed or linked) and Ethnicity and census district, along with quintiles of the census tract social deprivation index to construct variance stabilized inverse-probability weights. Preliminary descriptive analysis suggests that distributions of the population of CO APCD and the U.S. population are similar enough across these variables that valid weights can be constructed.

Aim 4: The course of long haul COVID-19 and pathway of persistent infection with COVID-19 remain unknown. There is an urgent need to characterize the trajectory of long haul COVID-19, from its risk factors, incidence and prevalence to its phenotypic signatures and prognosis. **Aim 4A.** Stanford will clearly define disease characteristics, incidence and trajectory. **Aim 4B.** Stanford will investigate the impact of vaccinations, courses of treatment or other mitigating factors on the incidence, severity and trajectory long haul COVID-19. Strategies to deal with biases in the decision/ability to get vaccinated will include propensity matching and instrumental variables to control for these unmeasured selection effects.

Aim 5: The identification of therapeutic targets and prophylaxis for COVID-19 and long haul COVID-19. Initially, retrospective chart review for conditions with a similar symptom profile or suspected biological pathway will be used. We will refine this symptom based algorithm for use on other datasets. Data transformation into OMOP should enable the exchange of code to enable replication studies on a wide array of databases.

Aim 6: Observation and monitoring of safety and efficacy of off-label (and fast track approved) therapies and prophylaxis for COVID-19 and long haul COVID-19. As patients either acquire or are prescribed medications or medications are approved for use, or various treatment regimes are attempted, we will identify which therapies appear to be the most effective and note trends with regard to safety and efficacy.

- How will this project benefit Colorado or Colorado residents? (this is a statutory requirement for all non-public releases of CO APCD data)

We expect that this project will benefit Colorado residents in three important ways:

First, the proposed work will both expand and enrich the variables available in the Colorado All Payer Claims data, particularly information on race, ethnicity and socioeconomic status. Although the appended variables are of immediate practical value in answering the proposed

aims, they can be reused to answer other programmatic questions, particularly around health disparities. Our hope is that the data enhancements, which characterize patterns of mortality for the state, will also enable Colorado to better understand which populations are *not* included in the CO APCD and that this knowledge will be useful in administering programs, reducing disparities and for reporting to federal agencies. It is our intention that the enhanced CO APCD would be available for reuse by other researchers as approved by the State of Colorado.

Second, the proposed work will help the State better understand the epidemiology of the SARS CoV-2 pandemic in Colorado to inform deployment of programs, mitigation of disparities and predict impacts on safety net programs, the medical system, workforce participation, disability claims (e.g. due to long-haul COVID), mortality (all cause) and population health. The pandemic has had significant impacts on state revenues (estimates of up to 16% for 2021)¹⁴ and the ability to predict and mitigate the pandemic should have a substantive impacts on the health status of Colorado residents and in turn, the economic outlook for the state.

Third, the proposed work will help the State better understand the epidemiology of long haul COVID-19 and its potential impact on programs, disparities and impacts on safety net programs, the medical system, workforce participation, disability claims, mortality (all cause) and population health. As long haul COVID-19 impacts a large fraction of COVID-19 victims, many of them of working age, it is likely to translate into large impacts on the medical system, workforce participation and safety net programs.

- Please answer all applicable questions below (Note that your project must meet one or more of the Triple Aim criteria below to generate a benefit for Colorado):
 - If applicable, how will your project support lowering health care costs?

Acute SARS CoV-2 is costly to treat, particularly if the disease results in hospitalization. Our hope is that a better understanding of risk factors for contracting COVID-19 will help the state target prevention efforts more effectively and identify the most effective treatments to prevent hospitalization and, in the event of hospitalization, either lengthy or intensive care hospital stays.

Long haul COVID-19 is likely to come with significant costs to the health care system both as a direct result of symptoms and indirectly due to uncertainty where physicians may attempt a number of therapies to bring symptom relief. Clear guidance on effective therapies should reduce the duration and severity of long haul COVID-19 and use of ineffective therapies.

- If applicable, how will you project help improve the health of Coloradans?
As noted above, COVID-19 and long haul COVID-19 both have significant impacts w/ respect to morbidity and mortality. Better understanding the epidemiology and optimal treatment course of these diseases will both help prevent Colorado residents from becoming ill to begin with, and, it is hoped, effective treatments will lead to better outcomes in the event that they do fall ill.
- If applicable, how will your project improve the quality of care or patient experience?

As described above, although there have been tremendous strides in reducing mortality in COVID-19 patients, a large fraction of patients still experience persistent and debilitating symptoms (long haul COVID-19). It is our hope that a better understanding of these diseases will lead to more effective treatments and better outcomes for patients. Good outcomes generally track with good Press-Ganey (patient satisfaction) scores.

Our hope is that this research leads to better value, effectiveness and well-being of patients as they are given treatments which lead to better health.

- Do you need a claims data set or would you like a custom report generated by CIVHC that addresses the specific questions/problems your project seeks to address?
1. CO APCD Fully Identifiable Data Set: For the linkage work with obituaries, AFC and other outside sources of data with demographic and socioeconomic variables, *we will require identifiers with full linkages to vital records (birth and death) as well as vaccination information*. Although hashed keys could potentially work for the dataset linkages, for the imputation of ethnicity, we will need names, birth information (e.g. birth name, parent name(s) and other information on the birth certificate will enable more accurate triangulation of likely ethnicity) and residence address. As noted above, these enhancement will be returned to the State of Colorado so that they can be used to meet programmatic needs. However, we will mint a DOI for the enhanced, linked set (with co-authorship with collaborators at CIVHC) and would be appreciative if you require citation if the enhanced set is reused for research by third parties through CIVHC.

The Stanford Center for Population Health Sciences has applied for an NIH grant to study long COVID-19 (Post Acute SARS CoV-2 or PASC). In the event this grant entitled COVReD, is awarded, Stanford will extract the COVID-19 cases from the full dataset, de-identify them and deposit them into the NIH Secure repository. This COVReD SARS CoV-2 dataset will also receive a DOI which can be used to cite the dataset.

We will note that although person readable identifiers have been removed, the data are granular (person-level), rich and longitudinal and as such should be afforded protections generally extended to high-risk data. The NIH Secure enclave is approved for PHI and data derived from PHI which retains residual risk.

- Do you need Protected Health Information (PHI)? Yes
 - Do you need patient-specific dates (e.g., dates of service or DOB) or 5 digit zip code. If so, this is a request for a **Limited Data Set**.

Yes, we will need to make this information available to COVID-19 and long haul COVID-19 researchers in order to enable high quality analyses and data overlay for social and environmental exposures, for example, shelter in place policies, mean income or air quality, can be overlaid by time and date variables.

- Do you need direct patient identifiers such as name, address, or city? If so, this is a request for an **Identifiable Data Set** (requires IRB approval).

For the linkage work with obituaries, AFC and other outside sources of data with demographic and socioeconomic variables, *we will require identifiers with full linkages to vital records (birth and death) as well as vaccination information.* Although hashed keys could potentially work for the dataset linkages, for the imputation of ethnicity, we will need names, birth information (eg, birth name, parent name(s) and other information on the birth certificate will enable more accurate triangulation of likely ethnicity). As noted above, these enhancements will be returned to the State of Colorado so that they can be used to meet programmatic needs. However, we will mint a DOI for the enhanced, linked set (with co-authorship with collaborators at CIVHC).

- If you do not require any PHI, please only complete PART ONE of the application.

Please note: your CIVHC representative will work with you to complete **Addendum I – Analyst Supplement** to address data warehouse specific questions.

PART TWO

I. **Type of CO APCD Analytic Data Set Requested (Not applicable for Custom Report Requests)**

Please select the type of data set that you are requesting by checking one of the boxes below (**select only ONE option**). Details on each type of CO APCD data set can be found in *The CO APCD Companion Instruction Guide* (available from your CIVHC representative):

Types of Analytic Data Sets (Please select ONE below)

For users interested in a wide range of data to analyze on their own.

- ☐ De-Identified Data Set
- ☐ Limited Data Set*
- ☒ Identified Data Set *

*These types of data requests include Protected Health Information (PHI). Under HIPAA, PHI may only be released in limited circumstances for public health, health care operations, and research purposes under the terms of a HIPAA compliant data use agreement (DUA).

2. **Requested Data Elements – Limited and Fully Identifiable Data Sets**

The CO APCD is committed to protecting the privacy and security of Colorado's health care claims data. The CO APCD will limit the use of the data to purposes permitted under applicable laws, including APCD Statute/Rule and HIPAA/HITECH, to information reasonably necessary to accomplish the project purpose as described in this Application.

Data Element Selection and Justification

If you have not already done so, please use the Data Element Dictionary (DED) to identify the specific data elements that are required for this project. In keeping with the minimum necessary standard established under HIPAA, CO APCD policy is to release only those data elements that are required to complete your project.

Type of Data	Justification for Elements on the DED
Names	<ul style="list-style-type: none"> Names will be used in conjunction with birth certificate information and geographic data to impute race and ethnicity for both patients and providers where this information is not available from other sources. Names will also be used for person level linkage to AFC. Names will be used for probabilistic linkage to obituaries.
Street Address	<ul style="list-style-type: none"> Address will be used to assign census block and be used to impute race and ethnicity where this information is not available from other sources. Address will be used for person level linkage to AFC. Address (most often city) will be used for probabilistic linkage to obituaries.
City	<ul style="list-style-type: none"> City will be used for probabilistic linkage to obituaries.
Zip Code	<ul style="list-style-type: none"> As noted above, address with city and zip code will be used for linkages to AFC and probabilistic linkage to obituaries.
Health Plan Beneficiary Numbers	<ul style="list-style-type: none"> These will be used to link to AFC insurance records where available.
Dates (including Day and Month detail.) Specify which date fields are needed and why.	<ul style="list-style-type: none"> Dates are important for monitoring the epidemiology of the COVID-19 outbreak, the severity of the disease (e.g., length of stay) and for probabilistic linkages to obituaries.
Provider Identifying Information	<ul style="list-style-type: none"> The American Family Cohort has rich and detailed provider information. These numbers will enable high fidelity matches to AFC.

A. Counts, Totals and other Summary Statistics

The CO APCD seeks to provide aggregated summary data whenever possible. Applicants are encouraged to request counts, totals, rates and other summary values whenever such information can reasonably accomplish the purpose of the project (add rows to the table below if necessary). The CO APCD supports the federal CMS minimum cell size suppression policy that requires any cell in any report or data table, printed or electronic, with less than eleven records or observations to be replaced by “Less than eleven” or similar text. You must also apply complementary cell suppression techniques to ensure that cells with fewer than eleven records cannot be identified by manipulating data in adjacent rows and columns.

Field Number and Name	Requested Count or Sum
	<i>[add rows as needed]</i>

We do not anticipate the need for summary statistics.

B. Linkages to Other Data Sets

The CO APCD seeks to ensure that data cannot be re-identified if it is linked to or combined with information obtained from other sources. If this project requires claims line level detail or includes linkages to other databases, or if CO APCD data will be combined with other information, provide a justification for each proposed linkage. Be sure to describe how this will contribute to achieving the project purpose, including whether the project can be completed without this linkage, and the steps you will take to prevent the identification of individual patients.

Will you link the CO APCD data to another data source?

- ☐ No.
- ☒ Yes. If yes, please answer the following questions.

- Which CO APCD identifying data elements will be used to perform the linkage?

We expect to use first, last and middle name, date of birth, social security number (these may be scrambled using a hash such as choicemaker such that that neither party is privy to the actual numbers), address and, if available other identifiers such as phone number or email.

- Once the linkage is made, what non-CO APCD data elements will appear in the new linked file?

As outlined in the data management plan in Appendix 3, upon completion of linkage, the data are converted to a limited dataset and person readable identifiers removed and kept separate from the main body of data. Despite the removal of person-readable identifiers, the data retain considerable richness and longitudinally, consequently they are treated as PHI.

- Have all necessary approvals been obtained to receive and link with the other data files (e.g., IRB or Privacy Board approval)?
 - ☐ Yes, if so please provide copy
 - ☒ In progress, anticipated approval date: _____
 - ☐ No or N/A, reason: _____

C. Distribution of the Report or Product: **Prior Review by the CO APCD Administrator**

If you are producing a report for publication in any medium (print, electronic, lecture, slides, etc.) the CO APCD Administrator must review the report prior to public release. The CO APCD Administrator will review the report for compliance with CMS cell suppression rules; risk of inferential identification; and consistency with the purpose and methodology described in this Application.

- Please describe your audience and how to you will make your project publicly available?
We plan to publish findings as academic articles in peer reviewed journals. If the state desires, we will be happy to provide either a report, policy brief or presentation on our findings on an annual or semi-annual basis.
- If the report is not to be made publicly available, then briefly describe how the information derived from this data will be used and by whom:
Findings will be made publicly available insofar as abstracts and publications are available. We will strongly favor open access journals and pay requisite fees for this as needed. If this is not an option, preprints of publications will be deposited into BioRxiv or similar.

Other Organizations: Do you intend to engage third parties who will have access to the data requested as part of this project? If so, list the organizations below, describe their role(s); and explain why they will be granted access to the requested data.

Organization/Company Name:	
Contact Person:	
Title:	
Address:	
Telephone Number:	
E-mail Address:	
Role or responsibility in this project	<i>[add rows as needed]</i>

If we are funded for the COVReD grant by the National Institutes of Health, a de-identified subset of the data containing all COVID-19 cases will be deposited into the NIH HIPAA Compliant secure repository.

The degree of access by other organizations hinges entirely on the provision of NIH funding (for example, participation in the long haul COVID-19 (e.g. PASC) consortium would result in consortium members accessing the data). Other collaborators will be direct Stanford collaborators and required to complete all training and security requirements as outlined above. We will implement whatever vetting process is requested by the state including explicit approval for each new user.

Project Schedule:

Proposed Project Start Date:	June 1, 2021
Project End Date:	May 31, 2025
Proposed Publication or Release Date:	We expect 1- 3 papers annually starting in year 1.
End of Date Retention Period:	We will expect to either renew our collaboration at the end of the study period or to return or destroy the data at that time.

D. Frequency

Data in the CO APCD Warehouse is refreshed every other month and data products can be provided on a one time basis or under a subscription model (e.g., quarterly, bi-annually or annually). Please select frequency below.

We are requesting the frequency required by the NIH in the event the COVReD grant is funded which is quarterly.

☐ One Time

OR

Subscription (Please select subscription model below)

- ☒ Quarterly
☐ Bi-annually
☐ Annually

E. Project Reporting

CIVHC highlights projects and data analysis on the public website: This display of CO APCD projects provides future data requesters with ideas of how they can structure their analysis, and allows CIVHC's stakeholders to see how CO APCD data recipients are working to accomplish the Triple Aim for Colorado. Data recipients have the option of choosing whether to be identified or to not be identified.

- ☒ Yes, it is okay for CIVHC to identify my organization
☐ No, I do NOT wish for CIVHC to identify my organization

If you are requesting a Custom Report with analytics to be provided by CIVHC; please stop here and submit the information above to your CIVHC representative.

PART THREE

DATA MANAGEMENT PLAN (Not applicable for Custom Report Requests)

I. Organizational Capacity

As an Attachment, please provide copies of the Data Privacy and Security Policies and Procedures for the Requesting Organization as well as those of any third parties that will have access to the requested CO APCD data. Attached in Appendix 3

- Has the Requesting Organization or any member of the project team ever been involved with a project that experienced a data security incident? If so, describe the incident, the response procedures that were followed and any subsequent changes in procedures, processes or protocols to mitigate the risk of further events.

To date, PHS has not been involved in a data security incident. We have had cases where an incident was suspected, but in both cases, the source of the data was another organization or the user had followed procedures.

In general, PHS policies are stricter than US law or the DUA such that violations of our rules require corrective action (amending drafts of papers to aggregate cells, returning or destroying data cuts downloaded onto encrypted computers) rather than reporting.

To the extent that the Data Privacy and Security Policies and Procedures, provided as an Attachment, do not already do so, please answer or attach answers for the following:

- **Physical Possession and Storage of CO APCD Data Files:**
 - Describe how you will maintain an inventory of CO APCD data files and manage physical access to them for the duration of the project:
 - Describe your personnel/staffing safeguards, including:
 - Confidentiality agreements in place with individuals identified as being assigned to this study. Include, for example, agreements between the Principal Investigator or Data Custodian and others, including research team members, and information technology and administrative staff:
 - Staff training programs you have in place to ensure data protections and stewardship responsibilities are communicated to the research team:
 - Procedures to track the active status and roles of each member of the research team throughout the project and a process for notifying the CO APCD of any changes to the team:
 - Describe your technical and physical safeguards. Examples include:
 - Actions taken to physically secure data files, such as site and office access controls, secured file cabinets and locked offices.
 - Safeguards to limit access to CO APCD data and analytical extracts among the research team (Note: if the distribution of analytical data extracts among the researcher team is part of your data management plan, the extracts remain subject to the terms of your Data Use Agreement).

- Provide a brief description of your policies and procedures for ensuring that CO APCD data are protected when stored on a server.
 - Describe how your organization prevents the copying or transfer of data to local workstations and other hard media devices (CDs, DVDs, hard drives, etc.). Note that Applicants are required to encrypt CO APCD data both in motion and at rest:
- Data Reporting and Publication
 - Your organization must ensure that all analytic extracts, analyses, findings, presentations, reports, and publications based on CO APCD data files adhere to specific requirements of the Data Use Agreement (DUA: refer to sections 6, 7 and 8 in the Data Use Agreement). **Briefly describe your plan for demonstrating that data reporting and publication processes will be consistent with the DUA, including adhering to CO APCD cell suppression policies:**

Please see the detailed data management plan attached in Appendix 3.

2. Completion of Research Tasks and Data Destruction

Your organization must ensure that it has policies and procedures in place to destroy the CO APCD data files upon completion of the project and that you have safeguards to ensure the data are protected when researchers terminate their participation in the research project. Describe your plan for demonstrating that your organization has policies and procedures in place to reliably destroy the data files upon completion of the research:

Please see the detailed data management plan attached in Appendix 3.

3. Request for Privacy Board Approval *(Only Applicable to Identifiable Data Requests)*

Projects that request Identifiable information for a research purpose may require approval from the DRRC acting as a Privacy Board if an IRB is not available.

- The DRRC, acting as a Privacy Board, may approve a waiver of the individual authorization normally required to release PHI under CFR § 164.508 if:
- It would be impracticable for researchers to obtain written authorization from patients that are the subject of the research; and
- The research could not practicably be conducted without access to and use of the PHI.
- The DRRC, acting as a Privacy Board, is required to evaluate certain criteria in considering whether to approve an authorization waiver. If you are requesting Identifiable Information for a research purpose, explain why your proposed use of PHI involves no more than a minimal risk to the privacy of patients that are the subject of the research. Evidence of minimal risk to the privacy of patients that should be addressed in your explanation includes:
 - An adequate plan to protect PHI identifiers from improper use and disclosure;
 - An adequate plan to destroy PHI identifiers at the earliest opportunity; and
 - Adequate written assurances that PHI will not be reused or disclosed.

Appendix I

Certification of Project Completion and Destruction or Retention of Data

(Please Save)

Name:	
Title:	
Organization:	
Address:	
Tel Number:	
Fax Number:	
E-mail Address;	
Project Title:	
Data Sets:	
Years:	
<input type="checkbox"/> Certification of Data Destruction	Date the Data was Destroyed:
<input type="checkbox"/> Request to Retain Data	Date Until Data Will Be Retained:

Instructions: Data must be destroyed so that it cannot be recovered from electronic storage media in accordance with the methods established by the “Guidance to Render Unsecured Protected Health Information Unusable, Unreadable, or Indecipherable to Unauthorized Individuals,” as established by the U.S. Department of Health and Human Services (HHS).

I hereby certify that the project described in the Application is complete as of this date
_____, ___, 20__.

Complete the appropriate section, below:

☐ I/we certify that we have destroyed all Data received from the CO APCD Administrator in connection with this project, in all media that were used during the research project. This includes, but is not limited to data maintained on hard drive(s), diskettes, CDs, etc.

☐ I/we certify that we are retaining the data received in connection with the aforementioned project, pursuant to the following health or research justification (provide detail, use as much additional space as necessary and state how long the data will be retained).

☐ I/we hereby certify that we are retaining the Data received from the APCD Administrator in connection with the aforementioned project, as required by the following law. [Reference the appropriate law and indicate the timeframe].

By signing this Agreement, the Receiving Organization agrees to abide by all provisions set out in this Agreement.

SIGNATURES:

For the CO APCD: CIVIHC

For Receiving Organization: The Board of Trustees of the
Leland Stanford Junior University

Signature:

Signature:

Name: Pete Sheehan

Name: Michala Welch

Title: VP of Client Solutions & State Initiatives

Title: Senior Contract and Grant Officer

Appendices

Appendix I: Bibliography

1. Menni C, Sudre CH, Steves CJ, Ourselin S, Spector TD. Quantifying additional COVID-19 symptoms will save lives. *The Lancet*. 2020;395(10241):e107-e108. doi:10.1016/S0140-6736(20)31281-2
2. Sominsky L, Walker DW, Spencer SJ. One size does not fit all – Patterns of vulnerability and resilience in the COVID-19 pandemic and why heterogeneity of disease matters. *Brain Behav Immun*. 2020;87:1-3. doi:10.1016/j.bbi.2020.03.016
3. Zhang Y, Li Y, Wang L, Li M, Zhou X. Evaluating Transmission Heterogeneity and Super-Spreading Event of COVID-19 in a Metropolis of China. *Int J Environ Res Public Health*. 2020;17(10):3705. doi:10.3390/ijerph17103705
4. Mahase E. COVID-19: What do we know about “long covid”? *BMJ*. Published online July 14, 2020:m2815. doi:10.1136/bmj.m2815
5. Bilinski A, Emanuel EJ. COVID-19 and Excess All-Cause Mortality in the US and 18 Comparison Countries. *JAMA*. 2020;324(20):2100. doi:10.1001/jama.2020.20717
6. Faust JS, Krumholz HM, Du C, et al. All-Cause Excess Mortality and COVID-19–Related Mortality Among US Adults Aged 25-44 Years, March-July 2020. *JAMA*. 2021;325(8):785. doi:10.1001/jama.2020.24243
7. Rivera R, Rosenbaum JE, Quispe W. Excess mortality in the United States during the first three months of the COVID-19 pandemic. *Epidemiol Infect*. 2020;148:e264. doi:10.1017/S0950268820002617
8. Biden JR. Executive Order On Advancing Racial Equity and Support for Underserved Communities Through the Federal Government. Published January 20, 2021. <https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/20/executive-order-advancing-racial-equity-and-support-for-underserved-communities-through-the-federal-government/>
9. Coston A, Guha N, Ouyang D, Lu L, Chouldechova A, Ho DE. Leveraging Administrative Data for Bias Audits: Assessing Disparate Coverage with Mobility Data for COVID-19 Policy. In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. ACM; 2021:173-184. doi:10.1145/3442188.3445881
10. Anderson M, Fienberg SE. Race and Ethnicity and the Controversy Over the US Census. *Curr Sociol*. 2000;48(3):87-110. doi:10.1177/0011392100048003007
11. Labgold K, Hamid S, Shah S, et al. Estimating the Unknown: Greater Racial and Ethnic Disparities in COVID-19 Burden After Accounting for Missing Race and Ethnicity Data. *Epidemiology*. 2021;32(2):157-161. doi:10.1097/EDE.0000000000001314

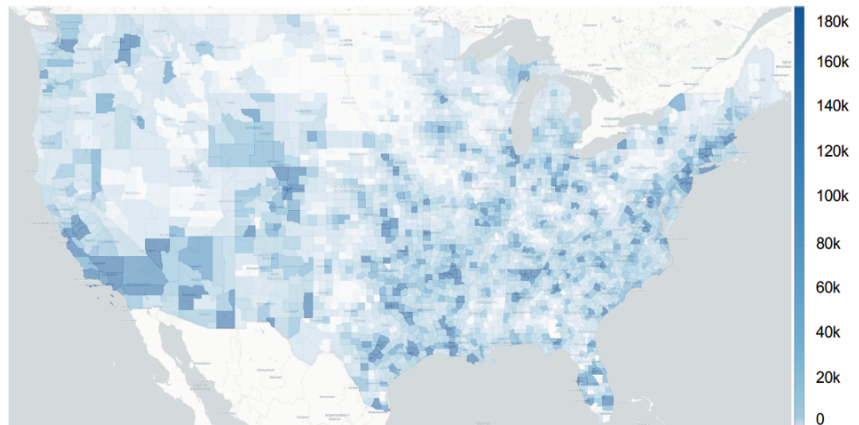
12. Strmic-Pawl HV, Jackson BA, Garner S. Race Counts: Racial and Ethnic Data on the U.S. Census and the Implications for Tracking Inequality. *Sociol Race Ethn*. 2018;4(1):1-13. doi:10.1177/2332649217742869
13. Ahmad FB, Anderson RN. The Leading Causes of Death in the US for 2020. *JAMA*. Published online March 31, 2021. doi:10.1001/jama.2021.5469
14. Center on Budget and Policy Priorities. States Grappling With Hit to Tax Collections. Published November 20, 2020. <https://www.cbpp.org/research/state-budget-and-tax/states-grappling-with-hit-to-tax-collections>
15. Phillips, Robert L, Rehkopf, David H., Vala. Ayin, Basezmore, Andrew, Peterson, Lars, Chu, Isabella. Clinical Registries Could Improve Influenza Like Illness and COVID-19 Surveillance. *Ann Fam Med COVID-19 Collect*. Published online April 29, 2020. <http://hdl.handle.net/2027.42/154853>

Appendix 2: The American Family Cohort and Colorado All Payer Claims linkage

The American Family Cohort. The American Family Cohort (AFC) is a research dataset comprised of electronic medical records from over 6.6 million unique patients throughout the US (Figure 1) with a decade of data and reflects an ongoing

partnership between PHS and the American Board of Family Medicine (ABFM). The AFC is derived from EHR data that have been assembled as part of the ABFM PRIME Registry. The PRIME Registry exists to reduce the burden of quality measurement and reporting and direct attention and resources to reducing inequity and improving patient outcomes. The AFC will enable us to answer the aims of this project as outlined above. The AFC data include:

Figure 1: American Family Cohort (AFC) Patient Distribution by US County



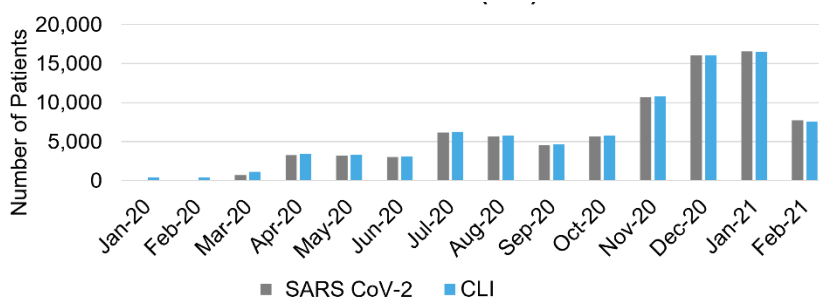
- A large number of individuals from populations that have been disproportionately impacted by SARS-CoV-2 infection from all 50 states, with enrollment roughly proportional to state populations. AFC includes patients on private insurance plans, Medicaid and Medicare, increasing the representation of vulnerable populations, and the generalizability of the sample to the overall US population.
- Racially and ethnically diverse data include 399,000 Black patients, 106,000 Asian patients, 21,000 Native American and Alaska Native patients and 13,000 Native Hawaiian and Pacific Islander patients. The remaining 4 million patients are White, and 502,000 patients have identified as Hispanic or Latino. The diversity is a major strength for addressing all of our key questions which focus on underserved and marginalized populations.
- Approximately 750,000 children whom we are able to link to parent health records, including a large number from minority populations from practices in *rural areas* of the US, allowing for analyses across the life course, and across generations in diverse groups (Q3).
- Individual patient and practice identifiers and a decade of data enables patients to be tracked over time and linked with external datasets. More than 75% of the patients have 10 or more visits enabling longitudinal analyses to study the trajectory of long haul COVID-19. These rich life course data will enable us to answer our Q3.

This rich and comprehensive longitudinal dataset has been used to reproduce patterns of Influenza-like Illness seen in the CDC ILI Network but with much larger sample sizes and ability to detect differences by race, ethnicity, and neighborhood social deprivation.¹⁵ AFC makes us uniquely poised to study disparities in long haul COVID-19 throughout the life course and across socioeconomic, racial/ethnic and geographic domains, and will provide a population platform from which we will recruit underserved patient groups for community-engaged citizen science data collection (see B6 and Appendix for further citizen science information and for ABFM letter of support).

Data from AFC (Figure 2) can be refreshed monthly. AFC RIF data are available on the Stanford Secure Data Ecosystem and, in the event of NIH funding, de-identified SARS CoV-2 data will be deposited to the long haul COVID-19 Data Science Core quarterly.

To address potential gaps in the EHR arising from either out of network care, loss of insurance coverage or clinic closures, and to provide robust endpoint data, we will link AFC records to Medicare, Medicare Advantage, Medicaid and the National Death Index. Through our partnership with CIVHC, we will link the 214,000 patients in AFC who reside in Colorado to the CO APC data including claims, birth, death and vaccination records. Finally, we have an ongoing partnership with the US Census Bureau to link all patients in AFC to individual Census data within the secure environment of the US Census Bureau. The full AFC-Census linked dataset will be updated annually and accessible to consortium researchers with Special Sworn Status (SSS) through the US Census Research Data Center at Stanford. Project PI David Rehkopf currently has SSS for other investigations using these data.

Figure 2: American Family Cohort SARS CoV-2 and COVID-19-Like Illness Patients



Appendix 3: Data Management Plan

Data Management Plan

For PHS Use Only
Privacy Board Approval Date:
Part D Approval Date:

DUA Signatory name and title	Michala Welch, Contract & Grant Officer, Office of Sponsored Research
Requesting Organization	Board of Trustees of the Leland Stanford Junior University (Stanford University)
Type of Organization	Academic
Study PI (if different from DUA User)	David Rehkopf, PhD
Study Title	COVReD: COVID-19 Real World Databases
Date of IRB Approval	Pending
Data Custodian Study Title	Protocol 36332 - Stanford Center for Population Health Sciences Data Repository (Approved July 31, 2021)
Date of IRB Approval	COVEReD approved February 05, 2021

A. DATA MANAGEMENT PLAN

Please reference the [Data Management Plan Guidelines](#), [Data Management Plan Evaluation Guide](#), [Collaborator Checklist](#), and/or the [FAQ document](#) for more information on completing this section. These materials are found under the Executive Summary section of the New Study Requesting Data page of the website.

For research studies involving researchers from another organization that will have access to RIF or non-identifiable files, please refer to the [Collaborator Checklist](#) for guidance and considerations to include in the Data Management Plan.

For collaborating organizations that will be receiving a physical copy of the Data files, a full Data Management Plan should be completed by the collaborating organization.

Definitions:

Containerized: A containerized software package includes everything the software needs to run such as system tools, libraries, and settings. Containers allow software to be isolated from the other software and the operating system itself. This property is often useful in designing secure cloud systems.

Data: Data refers to information at the individual level data as received from the Data Proprietors or generated in the course of research. Data are used to perform analytical work (eg: statistical analyses, machine learning, etc.) and to generate dashboards and other summary findings.

(The) Dataset: *For the purposes of this DUA, the “Dataset” refers to the XXX.* A collection of data tables. Datasets are generally either derived from a particular population, have the same Data Proprietor or some other unifying feature. For the purposes of this document “Dataset”

refers to the data owned by the Data Proprietor. Other “datasets” (small d) are other data which may be combined with the Dataset under consideration.

Data Management Plan (DMP): DMP refers to this document – namely a document drafted to describe in detail the procedures, parameters and technical protections for management, access and use of Data.

Data Use Agreement (DUA): DUA refers to an agreement between two parties prescribing acceptable uses and computational environments and other parameters of a dataset. PHS requires both inter-institutional DUAs and each individual accessing data to sign a DUA.

Data Proprietor: *For the purposes of this DUA, the “Data Proprietor” refers to the the State of Colorado via the Center for Improving Value in Health Care (CIVHC).* Data Proprietor is the institution, entity or individual who originally collected the data and/or owns the data and holds authority over acceptable uses and distribution of the data.

Data Risk Assessment (DRA): A DRA is an assessment of the sensitivity or risk associated with a dataset.

Google Cloud Platform (GCP): Google Cloud Platform or, GCP is the cloud infrastructure managed by Google Inc. GCP is FedRAMP and NIST-800-53 and 800-171 and 800-34 compliant. Stanford has a Business Associates Agreement with Google and the GCP instance as managed by Stanford PHS and SRCC is HIPAA compliant and approved for storage of PHI and other high-risk data. Detailed information on GCP’s security standards can be found here: [Google Compliance Resource Center](#)

High-risk Data: High-risk Data are data which carry significant personal or institutional risk. Data are generally considered high-risk if they contain personally identifiable information about individuals or entities, particularly around beliefs, behaviors or other sensitive topics. Data can also carry some proprietary, institutional, state or national security risk in the event they contain sensitive or proprietary information about an organization or entity, are combined with other data, become widely available or are used inappropriately.

Nero (Wolfe) On Premise Secure Computational Environment for High-risk Data (Nero on-prem): Nero, named after the famous fictional detective, is Stanford’s secure computational environment for high-risk data. The Nero cluster managed by Stanford Research Computing (SRCC).

Nero (Wolfe) Google Cloud Platform Secure Computational Environment for High-risk Data (Nero GCP): Nero GCP is virtually identical to Nero but the servers are managed by Google Inc and administered by Stanford Research Computing (SRCC). Nero GCP is the recommended solution for investigators who require exceptionally resource intensive computational environments.

Outputs: Outputs are graphs, tables, coefficients, formulas and other summaries derived from Data, where outputs or cells derived from coefficients or formulas represent cell sized greater than 10 individuals. Researcher outputs are not considered Data

Redivis: Redivis is a software company based in Palo Alto California. The PHS Data Portal is powered by Redivis software.

Smartsheet: Smartsheet is a web based spreadsheet with advanced functions that allows administrators to create forms and access the spreadsheet from a secure browser.

Stanford Center for Population Health Sciences (PHS): PHS is a center at Stanford University devoted to the social and environmental determinants of health. PHS is the

institutional custodian for the most heavily used high-risk datasets in the Stanford School of Medicine as well as many other high-value, high-risk and proprietary datasets.

Stanford Center for Population Health Sciences Data Core (PHS Data Core): The PHS Data Core is a small team of Data Administrators and Data Managers manage and administer the Data.

Stanford Center for Population Health Sciences Data Portal (PHS Data Portal): The PHS Data Portal is the entry point for discovery, access, exploration and cohort selection of the Data. The portal manages permissions at a tiered, granular level and enables data administrators to track and control data access and use. The PHS Data Portal is based on software developed by Redivis.

Stanford Center for Population Health Sciences Secure Servers (PHS Servers): The PHS Servers are a cluster of four secure servers which are used for analysis using Data. They have adequate computational power for most types of projects.

Stanford Research Computing Center (SRCC): The SRCC is the Stanford team of engineers who manage computational environments for research at Stanford University. SRCC engineers are the architects of the Nero computational environment and manage the PHS server cluster.

Stanford Secure Data Ecosystem (Stanford Secure Data Ecosystem): The Stanford Secure Data Ecosystem refers to the full complement of platforms, tools and personnel who manage high-risk data at Stanford. The Stanford Secure Data Ecosystem consists of both on premise (on-prem) and cloud based tools and computational environments. The Stanford Secure Data Ecosystem is designed to work seamlessly and keep data secure while providing researchers state of the art tools. The Secure Data Ecosystem is described in detail in the data management plan (DMP).

1. PHYSICAL POSSESSION AND STORAGE OF PHS DATA FILES

1.1. Who will have the main responsibility for organizing, storing, and archiving the data? Please provide name(s) and job title(s).

David Rehkopf, Faculty Director, Stanford Center for Population Health Sciences
Ruth Marinshaw, Director of Research Computing, Stanford Research Computing Center
Isabella Chu, Associate Director, Data Core Stanford Center for Population Health Sciences
Ayin Vala, Associate Director, Data Core, Stanford Center for Population Health Sciences
Emma Hallgren, Assistant Director, Stanford Center for Population Health Sciences

The list above represents the management of Stanford's research computing staff and the Stanford Center for Population Health Sciences Data Core. All individuals with custodial responsibility for data management, environment and administration will be either listed above will report to one of these individuals. Access to the data and adherence with regulatory and security requirements will be tracked on the PHS data portal as described below.

1.1. Maintenance of data inventory.

Upon receiving data, the Stanford PHS Data Portal logs a) the database name, b) the date that the data were received by and loaded onto the system, c) the identity and IP address of the analyst who was is responsible for the curation of the data along with all activities and d) the filenames along with e) their file types and sizes, and f) the data curating processes. Once the data have been loaded into the PHS Data Portal, permissions are set so that data users can view the data description, variable list and data documentation, and, assuming all requirements

have been met, the files needed for their analyses. Each of these can have permissions set at a granular level according to the Data Use Agreement.

Application documents are kept in Box or Google Drive. Both platforms are secure shared folder systems with granular controls. Files are linked in the portal for immediate access. Investigators are required to keep updated records of study personnel. These records along and other regulatory documentation are tracked in the PHS Data Portal.

In the event variables are received which are not covered under a DUA, PHS will inform both the Data Proprietor and the Privacy Office within 24 hours of confirmation that excess variables were received. In the event destruction is required, we will record a) Data Destruction date and b) we will include a Certificate of Disposition including the date we submitted the Certification of Disposition to the Data Proprietor.

1.2. Verification of privacy and security trainings prior to use of research data.

Data access is only shared with individuals who: a) have completed required human subjects and data security training b) have had all electronic devices which may potentially be used to access the data or outputs encrypted c) have attested to and provided proof of encryption either via Stanford University School of Medicine's encryption tracking and verification system (amie.stanford.edu) or equivalent institutional verification such as a letter from an institutional privacy or data security officer attesting to verification of encryption. d) have a "need to know" status with regard to the data and cannot practically work on the project without it and; e) have signed a data use agreement with Stanford PHS stating that they will only use the data for the stated research purposes and that they cannot share the data with any third party; f) as applicable - have institutional review board (IRB) approval for their study or are included as personnel on an IRB approved study, g) as applicable - have obtained Dataset Proprietor approval for their project. These requirements are collected, reviewed and approved for all investigators accessing the Dataset. In the case an investigator is from another institution, there may be additional requirements imposed by their home institution.

The primary responsibility for controlling Dataset access, including ensuring that all necessary requirements are correct and complete is the Stanford Center for Population Health Sciences Data Core (PHS Data Core). The PHS Data Core uses the PHS Data Portal to collect all items required for data access as described above. The software used in the PHS Data Portal will automatically suspend access if any requirement falls out of compliance or expires.

Individuals register on the PHS Data Portal using their institutional email and the portal authenticates the user identity through InCommon and, once a SUNet has been registered, SAML, Stanford's identity authentication system. Users are not permitted to share access to the PHS Data Portal. Requirements are either authenticated automatically, for example, institutional affiliation is automatically verified via InCommon and SUNets are authenticated automatically via SAML. Requirements which cannot be automatically verified are reviewed and verified by a PHS Data Administrator. In addition to individual requirements, study teams must register on the portal and a PI and administrative contact for each study team must be designated. Access permission ports automatically to Stanford Secure Data Ecosystem computational environments including PHS Secure Servers, Nero and Nero GCP. Activity logs such as logins, downloads, etc. are available for the PHS Data Portal, Nero and Nero GCP. PHS conducts random audit and regular intervals to monitor for and address unusual and unpermitted activity.

Generally, the initiation of manual or special access termination will come from the PHS Data Core however, both the PHS Data Core or SRCC Nero teams have the ability to revoke access to data instantly. In the event that an individual should only have temporary access to the data, for example if an individual has a temporary affiliation with Stanford or a project using Data, the PHS Data Administrator can set an expiration date on the affiliation of the individual in the PHS Data Portal and data access will terminate automatically upon the ending of the appointment. All requirements in the portal have expiration dates, either entered manually or calculated, and data access is terminated at the point that any requirement becomes expired or invalid.

The signed data use agreement will stipulate that the user will never share access to the data, only output results from analyses can be downloaded from the server – that is, the researcher agrees never to download datasets or subsets of those/analysis files and that all data outputs will conform to Stanford or Data Proprietor cell size policies, whichever is more restrictive.

File transfer activity is audited at regular intervals. In the event of a suspected or reported file download, the researcher in question will be contacted. In the event any PHS data has been downloaded to a computer or other device, the investigator will be required to expunge the files from their computer. Disciplinary action up to and including termination of access to all PHS files will be taken as appropriate. All computers used to access Data must be encrypted and password protected as described above. As stated throughout this document, individuals are not permitted to download Data onto any device, including but not limited to laptop or desktop computers, servers or external storage devices such as flash drives, external hard drives etc. Data, and outputs with small cell sizes must remain in the Stanford Secure Data Ecosystem and Stanford secure computational environment (Nero) and may not be removed.

In the event of termination of Stanford affiliation, either of an employee or student, of their own volition or being terminated prior to the completion of a project, access to the data will be terminated. Students or employees moving to another university or similar organization who wish to complete work on a project may make formal arrangements to continue project affiliation and access. Data are not shared by Stanford with any third parties or with any parties not explicitly named on the project for which the data were received. All personnel working with data must individually sign a PHS-DUA.

1.3. Privacy and Security

Data access is only shared with individuals who: a) have completed required human subjects and data security training b) have had all electronic devices which may potentially be used to access the data or outputs encrypted c) have attested to and provided proof of encryption either via Stanford University School of Medicine's encryption tracking and verification system (amie.stanford.edu) or equivalent institutional verification such as a letter from an institutional privacy or data security officer attesting to verification of encryption. d) have a "need to know" status with regard to the data and cannot practically work on the project without it and; e) have signed a data use agreement with Stanford PHS stating that they will only use the data for the stated research purposes and that they cannot share the data with any third party; f) as applicable - have institutional review board (IRB) approval for their study or are included as personnel on an IRB approved study, g) as applicable - have obtained Dataset Proprietor approval for their project. In the case an investigator is from another institution, there may be additional requirements imposed by their home institution.

The primary responsibility for controlling Data access, including ensuring that all necessary requirements are correct and complete is the Stanford Center for Population Health Sciences Data Core (PHS Data Core). The PHS Data Core uses the PHS Data Portal to collect all items required for Data access as described above. The software used in the PHS Data Portal will automatically suspend access if any requirement falls out of compliance or expires.

Individuals register on the PHS Data Portal using their institutional email and the portal authenticates the user identity through InCommon and, once a SUNet has been registered, SAML, Stanford's identity authentication system. Users are not permitted to share access to the PHS Data Portal. Requirements are either authenticated automatically, for example, institutional affiliation is automatically verified via InCommon and SUNets are authenticated automatically via SAML. Requirements which cannot be automatically verified are reviewed and verified by a PHS Data Administrator. In addition to individual requirements, study teams must register on the portal and a PI and administrative contact for each study team must be designated. Access permission ports automatically to Stanford Secure Data Ecosystem computational environments including PHS Secure Servers, Nero and Nero GCP. Activity logs such as logins, downloads, etc. are available for the PHS Data Portal, Nero and Nero GCP. PHS conducts random audit and regular intervals to monitor for and address unusual and unpermitted activity.

Generally, the initiation of manual or special access termination will come from the PHS Data Core however, both the PHS Data Core or SRCC Nero teams have the ability to revoke access to Data instantly. In the event that an individual should only have temporary access to the data, for example if an individual has a temporary affiliation with Stanford or a project using Data, the PHS Data Administrator can set an expiration date on the affiliation of the individual in the PHS Data Portal and data access will terminate automatically upon the ending of the appointment. All requirements in the portal have expiration dates, either entered manually or calculated, and data access is terminated at the point that any requirement becomes expired or invalid.

The signed DUA will stipulate that the user will never share access to the data, only output results from analyses can be downloaded from the server – that is, the researcher agrees never to download datasets or subsets of those/analysis files and that all data outputs will conform to Stanford or Data Proprietor cell size policies, whichever is more restrictive.

File transfer activity is audited at regular intervals. In the event of a suspected or reported file download, the researcher in question will be contacted. In the event any PHS data has been downloaded to a computer or other device, the investigator will be required to expunge the files from their computer. Disciplinary action up to and including termination of access to all PHS files will be taken as appropriate. All computers used to access Data must be encrypted and password protected as described above. As stated throughout this document, individuals are not permitted to download Data onto any device, including but not limited to laptop or desktop computers, servers or external storage devices such as flash drives, external hard drives etc. Data, and outputs with small cell sizes must remain in the Stanford Secure Data Ecosystem and Stanford secure computational environment (Nero) and may not be removed.

In the event of termination of Stanford affiliation, either of an employee or student, of their own volition or being terminated prior to the completion of a project, access to the data will be terminated. Students or employees moving to another university or similar organization who wish to complete work on a project may make formal arrangements to continue project affiliation and access. Data are not shared by Stanford with any third parties not explicitly named on the

project for which the data were received. All personnel working with data must individually sign a PHS-DUA.

1.4. Notification of project staffing changes.

The Principal Investigator (PI) of any research project using PHS data will promptly inform the Data Core via email of any staffing changes. The PI or their designee must also update the appropriate study team members within the Stanford PHS Data Portal. The Data Core will request an update of team members at least annually. The Data Core Manager will confirm access to data has been terminated for any individuals who should no longer have access. The PHS Data Portal is the preferred method of transmitting this type of information as it provides the notice in writing, is a searchable, indexed format, includes a date and time stamp and allows the sender to confirm receipt. Additionally, the PHS Data Portal tracks permissions on an ongoing basis and individuals who have a requirement drop out of compliance will have data access suspended.

1.5. Training and education programs for high-risk data.

As stated in Section 1.3, analytic datasets are only shared with individuals who have completed the necessary regulatory, security and training requirements. PHS requires data security training as well as the appropriate human subjects trainings for all individuals accessing data with PHI or PII.

Collaborators from outside institutions wishing to gain access to PHS data for the purposes of analyses must adhere to all the same stipulations trainings and conditions as researchers at Stanford including completion and verification of human subjects and data security training.

1.6. Infrastructure (facilities, hardware, software, other) to secure high-risk PHS data.

On-prem infrastructure:

This information is repeated in its entirety, verbatim in section 1.6 and 1.9. Portions of this section are repeated verbatim in section 2.5.

All PHS HOSTED Data is managed and must remain on the Stanford Secure Data Ecosystem. The Stanford Secure Data Ecosystem includes the PHS Data Portal, HIPAA Compliant Google Cloud Platform (GCP) used to store data, Nero on-prem and Nero GCP computational environments. Each component of the Stanford Secure Data Ecosystem is described in detail below.

Data workflows within this ecosystem follow a clearly defined pipeline that ensures data are secured at all steps throughout their lifecycle. Specifically, data are first received on encrypted disks or via secure FTP transfer; authorized personnel (administrators) ingest the initial data into the PHS Data Portal. Data are stored on the PHS Data Portal in a HIPAA and FedRAMP compliant, multi-redundant, AES256 encrypted datastore on GCP (BigQuery). After performing validation checks, administrators configure the various access requirements necessary to view the dataset documentation, metadata, and query the underlying data. Next, authenticated researchers may then apply for access through the PHS Data Portal; upon gaining appropriate access, a researcher may then run queries against the data to create an analytical cohort. Finally, an administrator reviews the selected files to ensure that they are consistent with the request submitted to the State of Colorado. The cohort is then exported to the secure computational environment (i.e. Nero), wherein researchers can leverage additional statistical and computational software to continue their analyses.

The PHS hosted data will be received by the Stanford PHS Data Core via encrypted disk or secure FTP transfer and saved to Google Cloud Platform (GCP) which has been configured to be HIPAA and FedRAMP compliant. The PHS Data Portal, using Redivis software, manages the ability to see the existence of the datasets and access to meta-data and the data themselves at a tiered, granular level.

All data and metadata on the PHS Data Portal are stored in a multi-redundant, AES256 encrypted datastore on GCP. All connections to the data portal must be encrypted using TLS 1.2 or greater protocol. Redivis supports single sign on via eduGAIN, InCommon for overview access and dual authentication for full data access. In addition to HIPAA and FedRAMP compliance, the PHS Data Portal supports HTTP Strict Transport Security and is on the HSTS preload list for all major browsers which prevents users from establishing an unencrypted connection. The PHS Data Portal implements multiple layers of logging and monitoring to quickly identify and remedy anomalous behavior.

On-premise computation will take place on the Nero on-prem computing environment. Nero specifications can be found below:

1. System

The data will be received by the Stanford PHS Data Core via encrypted drive or secure FTP transfer and saved to a secure server that abides by Stanford's Computer and Network Usage Policy and Information Security Policy, as well as Minimum Security Standards for servers (MINSEC) for PHI and other sensitive data. Research analytics will be performed on a system using a standard three-tier architecture. Server resources reside with the Stanford Research Computing Center (SRCC); the server and all compute resources are located behind the Stanford University firewall. There is no incoming web access except as necessary for login, as described below.

2. Hardware/Software (e.g. name of the anti-virus, anti-spyware, firewall, host intrusion, remote access, etc.)

A Kubernetes-based Linux (Ubuntu) containerized Server environment currently running on Dell and Lenovo server hardware.

3. Access Control (e.g. password requirements and safeguards, VPN use, WiFi use, file sharing, data access logs, etc.)

Access is strictly limited. All users of system functions go through a two-factor authentication process, including a valid username (SUNet ID) and password and use of the Duo application. Passwords are evaluated for strength on submission and must adhere to Stanford's password complexity policies. Passwords must be at least 8 characters and include at least one letter, one number and one non alpha-numeric character and they have to be updated on an annual basis. Requirements vary for longer passwords. All passwords are stored in an encrypted format. Administrators within the Stanford Secure Data Ecosystem have unique privileges and interfaces on the Stanford PHS Data Portal and the Nero environment. Administrators also have additional requirements with regard to training and password strength. The Stanford policies and standards for systems administrators are laid out in the Stanford Guide for System Administrators here: <https://uit.stanford.edu/security/sysadmin>.

Remote access is permitted but is protected through secure measures, including use of VPN and two-factor authorization. A log is maintained that indicates when someone has accessed, read or used data.

4. Physical Environment (e.g. monitor position, printer location, screensaver, etc.)

The on-premises server environment is housed in a secure Stanford facility, the Stanford Research Computing Facility (SRCF), maintained by the SRCC group. This facility has 28 cameras that capture motion 24x7. In case of facility power failure, supplemental power is provided by a generator. Testing of the ability to switch between commercial and generator power occurs monthly. The servers have also been outfitted with self-encrypting hard drives.

No physical access to the Nero servers is allowed for non-SRCC group personnel. SRCC provides active monitoring of the server for access/security issues, as well as updating security, anti-virus and firewall software, including maintenance of security patches. Network connections are isolated and network access is restricted as noted above in "Access Control"; server-level access is limited to SRCC and a handful of program staff, requiring both valid username and password and the use of a hardware authentication token. Research outputs can be moved to encrypted laptops using SSL, providing end-to-end encrypted transfer. Both WiFi and Ethernet connections used for access require the use of Stanford VPN to connect; use of the VPN also requires two-factor authentication. Authorized system administrators can physically access the system components when needed for maintenance activities.

5. Data Storage (e.g. removal media storage, hard drive encryption, replicas of the data, etc.)

All data storage and replicas will be controlled by the SRCC-systems group. Replicas are stored at a secure research datacenter facility, the Stanford Research Computing Facility (SRCF). Data are replicated nightly to a replica system in a different secured Stanford data center in Forsythe Hall on the Stanford campus. Both the regular and replica servers adhere to the same security requirements. Access to the SRCF is tightly controlled, with keycard access limited only to authorized individuals. In addition, the server racks for Nero are locked; the key to a given rack is only available to the system administrators of the servers and the SRCF Data Center Manager. The SRCF has 28 cameras that capture motion 24x7. In case of facility power failure, supplemental power is provided by a generator. Testing of the ability to switch between commercial and generator power occurs monthly. Similarly, the racks housing replica data servers in Forsythe hall are locked; access to Forsythe is tightly controlled with keycard access limited to authorized individuals. Nero servers have also been outfitted with self-encrypting hard drives. All statistical software necessary for analyses are available on the Nero servers so that it is never necessary to download data onto local computers or other electronic devices. In addition to the above safeguards, server system administrators are required to attend two days of information security training each year, and they can only access the servers from bastion host systems specifically secured by the Stanford Information Security Office.

6. Encryption (type of encryption used for data storage on hard drives)

All encrypted devices will use AES or better encryption methodology.

Nero requires users to be connected to the Stanford Network via VPN, and to authenticate with two-step authentication (SUNet and password with Duo) for either interactive or ssh access.

Interactive connections require an Internet browser (e.g.: Chrome, Safari or Firefox), while ssh connections require a terminal window.

The Stanford University Network Access Control ([SUNAC](#)) service allows granular and configurable control of data access. The service allows the Manager of the Data Core (or the Local Network Administrator, LNA) to control remote access to departmental resources located behind University IT-managed firewalls. Using Workgroup Manager, access can be granted and customized for any PHS Member by SUNet ID. The PHS Data Core manager is responsible for this permissioning process for the data projects.

Stanford University requires encryption of all devices used to access Stanford resources—whether the computers or devices are owned personally or by the University. All devices used to access Stanford systems must be registered in Stanford's internal central tracking system and proof of this registration and encryption must be attached on the user's account in the PHS Data Portal. Each person accessing Stanford systems must fill out a data attestation form annually or whenever there is a change in either their devices or the types of data they access, whichever is most frequent. Each computer or device used to access Stanford School of Medicine systems PHI must have whole disk encryption. Encryption status of every computer and device used to access high risk data is both tracked and audited continuously per Stanford IRT [Information Privacy and Security Policies](#).

Computer encryption is conducted and verified using the Stanford Whole Disk Encryption (SWDE) service. The SWDE service is for both Windows and Macintosh computers that support native encryption. Once installed, all files are automatically encrypted. The data are protected while the computer is in standby or hibernation mode. This requirement applies to both Stanford and personally owned computers that are used for Stanford activities on the campus network. Users must authenticate with their SUNet ID's to connect to the Stanford network. Additionally, all Stanford systems require two step SUNet authentication with a password. Passwords must comply with [Stanford Password Guidelines](#). Passwords must be a minimum of 8 characters and passwords less than 12 characters require a combination of mixed case letters, numbers and special characters.

Every computer using SWDE automatically checks in with a logging and administrative server every 7 days. In the event of loss or theft of a computer with High Risk Data, Stanford policy requires notification of the Information Security Office (ISO). ISO in turn will use the logs to determine if a lost or stolen computer is a "reportable" event, possibly requiring notification of persons whose data may have been lost or stolen. A record of the encrypted devices owned by each individual requesting data access which might be potentially used by the individual to access the data are tracked in that individuals record in the PHS Data Portal. This encryption is monitored and verified as described above.

In addition to these security policies, PHS does not permit individuals to download any data, including but not limited to, data onto individual's computers or devices. All analyses must be conducted in the Secure Data Ecosystem and researchers are only permitted to download outputs which comply to PHS cell size restrictions, or the cell size restrictions of the owner of the data (Colorado), whichever is more restrictive. Printing in the server environment is disabled. Thus, individuals are unable to print data.

Computer encryption must be conducted and verified using Whole Disk Encryption (WDE). Both Windows and Macintosh computers support native encryption. Once installed, all files must be automatically encrypted. The data must be protected while the computer is in standby or hibernation mode. This requirement applies to both computers owned by the home institution of the researcher and personally owned computers that are used for Stanford activities on the campus network. Computers on campus require SUNet authentication to enter. Personal computers must require a 10 character PIN to open the computer. Additionally, all Stanford systems require two step SUNet authentication with a password.

The drives, discs or media containing the original files, as received, will be stored in a safe in the locked room accessible only to the Data Core management team (Data Custodian, Data Manager and the Center Director at the Stanford Center for Population Health Sciences). The combination of the safe, a SentrySafe SFW205CWB Water-Resistant Combination Safe, 2X-Large, is only known to the PHS Data Core management team (Data Custodian, Data Manager and the Center Director at the Stanford Center for Population Health Sciences). Further, the data will not be physically moved or transmitted in any way from Stanford without written approval from the Data Proprietor.

Google Cloud Platform Infrastructure:

Nero is available in two parallel instances. The first is an on-premise cluster as described above. The second is almost entirely identical from the researcher's perspective and subject to equivalent permissioning controls, but the physical hardware (servers) are managed by Google. These two environments are referred to as Nero on-prem and Nero GCP respectively. Nero GCP is used less frequently, typically in cases where the analyses taking place are particularly resource intensive. Individuals with a Nero GCP account must also provision a Nero on-prem account in order to accommodate data transfer from the PHS Data Portal.

Google Cloud Platform (GCP) and G-Suite are FedRAMP and NIST 800-53 complaint and listed as such on FedRAMP MarketPlace and Google websites. Stanford has a Business Associates Agreement with Google that stipulates that hardware used for Stanford School of Medicine business must be HIPAA compliant, in addition to native encryption and security parameters set by Google.

In addition to the security parameters described above, the PHS Data Portal software (Redivis) manages permissions on an ongoing basis at a tiered, granular level and rescinds data access in the event an individual falls out of compliance with any one requirement. Once an individual has gained access to files, they are permitted to port the necessary cohort and files over to Nero on-prem or Nero GCP. Nero servers abide by the security parameters described above. No downloads or data transfers from the Stanford Secure Data Ecosystem are permitted beyond the porting of cohort files as described above.

1.7. Policies and procedures regarding the physical possession and storage of PHS data files.

All servers will abide by Stanford's Computer and Network Usage Policy and Information Security Policy, as well as Minimum Security Standards for servers for PHI and other sensitive data. In the case of GCP security is maintained in accordance with FedRAMP and NIST-800-53 requirements. The requirements associated with those standards include patching, server hosting in a data center, centralized logging, two-factor authentication, vulnerability scanning

and mitigation, and intrusion detection. In addition, server system administrators are required to attend two days of information security training each year.

The drives, discs or media containing the original Dataset files, as received, will be stored in a safe in the locked room accessible only to the Data Core management team (Data Custodian, Data Manager and the Center Director at the Stanford Center for Population Health Sciences). The combination of the safe, a SentrySafe SFW205CWB Water-Resistant Combination Safe, 2X-Large, is only known to the Data Core management team (Data Custodian, Data Manager and the Center Director at the Stanford Center for Population Health Sciences). Further, the data will not be physically moved or transmitted in any way from Stanford without written approval from Data Proprietor.

1.8. System to track the status and roles of the research team.

The Stanford PHS Data Core, keeps records of study personnel and status with regard to protocols, data access and current human subjects and data security trainings. This information is tracked on the PHS Data Portal and redundant tracking takes place on the Stanford e-protocol system and the Smartsheet which is used for reporting purposes. The official record of data access is the Stanford Data Portal as this system is able to track not only initial access but continuous compliance with access requirements in an automated fashion. In this system roles are delineated and records are updated as protocol or personnel changes occur or on an annual basis, whichever is more frequent. In addition to listing relevant personnel on the PHS Data Portal and IRB protocols, both systems require that required human subjects and security trainings are completed and up to date. In the event research projects are completed, the Stanford Data Manager records the completion of the project and the access to data is terminated. The closure is also noted in the e-protocol system.

For outside collaborators, the same tracking mechanisms are employed and outside collaborators are required to register and authenticate on our portal. The portal collects, the individual's name, ORCID, secondary contact information, affiliation (authenticated), Stanford collaborator and SUNET sponsor, verification of human subjects and data security training, verification of encryption and both IRB and Data Proprietor approval as applicable.

1.9. Physical and technical safeguards used to protect PHS data files (including physical access and logical access to the files).

These safeguards are also outlined in sections 1.6 and 2.5. On-prem infrastructure: This information is repeated in its entirety, verbatim in section 1.6 and 1.9. Portions of this section are repeated verbatim in section 2.5.

All PHS Data is managed and must remain on the Stanford Secure Data Ecosystem. The Stanford Secure Data Ecosystem includes the PHS Data Portal, HIPAA Compliant Google Cloud Platform (GCP) used to store data, the PHS Secure Servers, the Nero on premise and Nero GCP computational environments. Each component of the Stanford Secure Data Ecosystem is described in detail below.

Data workflows within this ecosystem follow a clearly defined pipeline that ensures data are secured at all steps throughout their lifecycle. Specifically, data are first received on encrypted disks or via secure FTP transfer; authorized personnel (administrators) ingest the initial data into the PHS Data Portal. Data are stored on the PHS Data Portal in a HIPAA and FedRAMP

compliant, multi-redundant, AES256 encrypted datastore on GCP (BigQuery). After performing validation checks, these administrators configure the various access requirements necessary to view the dataset documentation, metadata, and query the underlying data. Next, authenticated researchers may then apply for access through the PHS Data Portal; upon gaining appropriate access, a researcher may then run queries against the data to create an analytical cohort. Finally, an administrator reviews the selected files to ensure that they are consistent with the request submitted to PHS and the Data Proprietor as applicable. The cohort is then exported to the secure computational environment, wherein researchers can leverage additional statistical and computational software to continue their analyses.

The Dataset will be received by the Stanford PHS Data Core via encrypted disk or secure FTP transfer and saved to Google Cloud Platform (GCP) which has been configured to be HIPAA and FedRAMP compliant. The PHS Data Portal, using Redivis software, manages the ability to see the existence of the datasets and access to meta-data and the data themselves at a tiered, granular level.

All data and metadata on the PHS Data Portal are stored in a multi-redundant, AES256 encrypted datastore on GCP. All connections to the PHS Data Portal must be over an encrypted, TLS 1.2 or greater protocol. Redivis supports single sign on via eduGAIN, InCommon for overview access and dual authentication for full data access. In addition to HIPAA and FedRAMP compliance, the PHS Data Portal supports HTTP Strict Transport Security and is on the HSTS preload list for all major browsers which prevents users from establishing an unencrypted connection. The PHS Data Portal implements multiple layers of logging and monitoring to quickly identify and remedy anomalous behavior.

On-premise computation will take place on either the PHS servers or the Nero Computing On-Premise Environment. Nero specifications can be found below:

System

The data will be received by the Stanford PHS Data Core via encrypted drive or secure FTP transfer and saved to a secure server that abides by Stanford's Computer and Network Usage Policy and Information Security Policy, as well as Minimum Security Standards for servers (MINSEC) for PHI and other sensitive data. Research analytics will be performed on a system using a standard three-tier architecture. Server resources reside with the Stanford Research Computing Center (SRCC); the server and all compute resources itself will be located behind the Stanford University firewalls and there is no incoming web access except as necessary for login, as described below.

Hardware/Software (e.g. name of the anti-virus, anti-spyware, firewall, host intrusion, remote access, etc.)

A Kubernetes-based Linux (Ubuntu) containerized Server environment currently running on Dell and Lenovo server hardware.

Access Control (e.g. password requirements and safeguards, VPN use, WiFi use, file sharing, data access logs, etc.)

Access is strictly limited. All users of system functions go through a two factor authentication process, including a valid username and password and use of the DUO application. Passwords are evaluated for strength on submission and must adhere to Stanford's password complexity policies. Passwords must be at least 8 characters and include at least one letter, one number

and one non alpha-numeric character. Requirements vary for longer passwords. All passwords are stored in an encrypted format.

Remote access is permitted but is protected through secure measures, including use of VPN and two-factor authorization. A log is maintained that indicates when someone has accessed, read or used data.

Physical Environment (e.g. monitor position, printer location, screensaver, etc.)

The server environment is housed in a secure facility maintained by the SRCC group. This facility has 28 cameras that capture motion 24x7. In case of facility power failure, supplemental power is provided by a generator. Testing of the ability to switch between commercial and generator power occurs monthly. The server has also been outfitted with self-encrypting hard drives.

No physical access to the servers is allowed for non-SRCC group personnel. SRCC provides active monitoring of the server for access/security issues, as well as updating security, anti-virus and firewall software, including maintenance of security patches. Network connections are isolated and network access is restricted as noted above in "Access Control"; server-level access is limited to SRCC and a handful of program staff, requiring both valid username and password and the use of a hardware authentication token. Research outputs can be moved encrypted laptops using SSL, encrypted transfer. Both Wifi and Ethernet connections used for access require the use of Stanford VPN to connect. Authorized system administrators and the PHS Data Center Manager can physically access the system components when needed for maintenance activities.

Data Storage (e.g. removal media storage, hard-drive encryption, replicas of the data, etc.)

All data storage and replicas will be controlled by the SRCC-systems group. Replicas are stored at a secured datacenter facility located at Stanford. Data are replicated nightly to a replica system in a different secured Stanford data center in Forsythe Hall on the Stanford campus. Both the regular and replica servers adhere to all of the same security requirements. The server is housed in a secure research data center, the Stanford Research Computing Facility, with keycard access limited only to authorized individuals. In addition, the server rack itself is locked; the key is only available to the system administrators of the server and the Data Center Manager. The Data Center has 28 cameras that capture motion 24x7. In case of facility power failure, supplemental power is provided by a generator. Testing of the ability to switch between commercial and generator power occurs monthly. The server has also been outfitted with self-encrypting hard drives. All statistical software necessary for analyses live on the server so that it is never necessary to download data onto local computers or other electronic devices. In addition to the above safeguards, server system administrators are required to attend two days of information security training each year and they only access the server from bastion host systems specifically secured by the Stanford Information Security Office.

Encryption (type of encryption used for data storage on hard-drives)

All encrypted devices will use AES or better encryption methodology.

The PHS Windows servers are encrypted with BitLocker. The version of Windows running on our infrastructure (Windows Server 2012 R2 and Windows Server 2016) has a recovery password that is FIPS 140-2 compliant per Microsoft FIPS 140-2 Documentation.

The PHS storage servers on which data will be "permanently" located is encrypted using LUKS, the Linux Unified Key Setup which does block-level encryption on all storage drives. We use FIPS approved AES and 3DES ciphers and encryption algorithms across SSH, NFS, and samba. This is per the FIPS Annex Security Requirements.

The system is accessed by approved users via SSH, X2Go via SSH and RDP via the campus VPN. All authorized users of the data will connect to the Stanford network via VPN and will authenticate with two step authentication (SUNetID and password with Duo or Text Code secondary authentication); they will then also authenticate to the server with SUNetID and password.

The Stanford University Network Access Control (SUNAC) service allows granular and configurable control of data access. The service allows the Manager of the Data Core (or the LNAs Local Network Administrator) to control remote access to departmental resources located behind University IT-managed firewalls. Using Workgroup Manager, access can be granted and customized for any PHS Member by SUNet ID. The PHS Data Core manager is responsible for this permissioning process.

Stanford University requires encryption of all devices used to access Stanford resources—whether the computers or devices are owned personally or by the University. All devices used to access Stanford systems must be registered in Stanford's internal central tracking system and proof of this registration and encryption must be attached on the user's account in the PHS Data Portal. Each person accessing Stanford systems must fill out a data attestation form annually or whenever there is a change in either their devices or the types of data they access, whichever is most frequent. Each computer or device used to access Stanford School of Medicine systems PHI must have whole disk encryption. Encryption status of every computer and device used to access high-risk data is both tracked and audited continuously per Stanford IRT Information Privacy and Security Policies.

Computer encryption is conducted and verified using the Stanford Whole Disk Encryption (SWDE) service. The SWDE service is for both Windows and Macintosh computers that support native encryption. Once installed, all files are automatically encrypted. The data are protected while the computer is in standby or hibernation mode. This requirement applies to both Stanford and personally owned computers that are used for Stanford activities on the campus network. Computers on campus require SUNET authentication to enter. Personal computers require a 10 character PIN to open the computer. Additionally, all Stanford systems require two step SUNET authentication with a password. Passwords must comply with Stanford Password Guidelines. Passwords must be a minimum of 8 characters and passwords less than 12 characters require a combination of mixed case letters, numbers and special characters.

Every computer using SWDE automatically checks in with a logging and administrative server every 7 days. In the event of loss or theft of a computer with High-risk Data, Stanford policy requires notification of the Information Security Office (ISO). ISO in turn will use the logs to determine if a lost or stolen computer is a "reportable" event, possibly requiring notification of persons whose data may have been lost or stolen. A record of the encrypted devices owned by each individual requesting Dataset access which might be potentially used by the individual to access the data are tracked in that individuals record in the PHS Data Portal. This encryption is monitored and verified as described above.

In addition to these security policies, PHS does not permit individuals to download any data, including but not limited to, individual's computers or devices. All analyses must be conducted in the Secure Data Ecosystem and researchers are only permitted to download outputs which comply to PHS cell size restrictions, or the cell size restrictions of the Data Proprietor, whichever is more restrictive. Printing in the server environment is disabled. Thus, individuals are unable to print data.

Computer encryption must be conducted and verified using Whole Disk Encryption (WDE). Both Windows and Macintosh computers support native encryption. Once installed, all files must be automatically encrypted. The data must be protected while the computer is in standby or hibernation mode. This requirement applies to both computers owned by the home institution of the researcher and personally owned computers that are used for Stanford activities on the campus network. Computers on campus require SUNet authentication to enter. Personal computers must require a 10 character PIN to open the computer. Additionally, all Stanford systems require two step SUNet authentication with a password.

The drives, discs or media containing the original Dataset files, as received, will be stored in a safe in the locked room accessible only to the Data Core management team (Data Custodian, Data Manager and the Center Director at the Stanford Center for Population Health Sciences). The combination of the safe, a SentrySafe SFW205CWB Water-Resistant Combination Safe, 2X-Large, is only known to the PHS Data Core management team (Data Custodian, Data Manager and the Center Director at the Stanford Center for Population Health Sciences). Further, the data will not be physically moved or transmitted in any way from Stanford without written approval from Dataset Proprietor.

Google Cloud Platform Infrastructure:

Nero is configured in two parallel versions. The first is an on premise cluster as described above. The second is almost entirely identical from the researcher's perspective and subject to all the same permissioning controls, but the physical hardware (servers) are managed by Google. These are named Nero and Nero GCP respectively. Nero GCP is used less frequently and only in cases where the analyses taking place are particularly resource intensive.

Google Cloud Platform (GCP) and G-Suite are FedRAMP and NIST 800-53 complaint and listed as such on FedRAMP Marketplace and Google websites. Stanford has a Business Associates Agreement with Google that stipulates that hardware used for Stanford School of Medicine business must be HIPAA compliant. In addition to native encryption and security parameters set by Google.

In addition to the security parameters described above, the PHS Data Portal software (Redivis) manages permissions on an ongoing basis at a tiered, granular level and rescinds data access in the event an individual falls out of compliance with any one requirement. Once an individual has gained access to files, they are permitted to port the necessary cohort and variables over to Nero or the PHS on premise servers. Both Nero and the on premise servers abide by the same security parameters as described above. No downloads or data transfers off the Stanford Secure Data Ecosystem are permitted.

2. DATA SHARING, ELECTRONIC TRANSMISSION, DISTRIBUTION

2.1. Policies and procedures regarding the sharing, transmission, and distribution of PHS data files.

Access to Dataset is only granted to individuals, including collaborators, who have completed all required regulatory, security and training requirements. At a minimum, these include all training and security requirements as Stanford investigators and in some cases will also include requirements from their own university. At a minimum these include: a) have completed required human subjects and data security training b) have had all electronic devices which may potentially be used to access the data or outputs encrypted c) have attested to and provided proof of encryption either via Stanford University School of Medicine's encryption tracking and verification system (amie.stanford.edu) or equivalent institutional verification such as a letter from an institutional privacy or data security officer attesting to verification of encryption. d) have a "need to know" status with regard to the data and cannot practically work on the project without it and; e) have signed a data use agreement with Stanford PHS stating that they will only use the data for the stated research purposes and that they cannot share the data with any third party; f) as applicable - have institutional review board (IRB) approval for their study or are included as personnel on an IRB approved study, g) as applicable - have obtained Dataset Proprietor approval for their project. These requirements are collected, reviewed and approved for all investigators accessing the Dataset. In the case an investigator is from another institution, there may be additional requirements imposed by their home institution. These standards are outlined in Stanford's Data Security Policies governing the treatment and distribution of PHI and other high-risk data.

Access to approved files will be managed on the PHS Data Portal which allows a PHS Data Administrator to control access to individual files. Any investigator who has completed all regulatory and security requirements will be granted access to the appropriate files on the PHS Data Portal. The investigator can then cut their analytic cohort which will then be ported over to the Stanford Secure Analytic Environment. All analyses, code and resulting work files related to the project will be kept in the project folder. The investigator is expected to also link the final version of any resultant products of research (in most cases a peer-reviewed publication) in the PHS Data Portal. At the end of the project access to the data will be terminated.

The data use agreement will stipulate that only output results from analyses can be downloaded from the server – that is, the researcher agrees never to download datasets or subsets of those/analysis files and that all data outputs will conform to Stanford or Data Proprietor cell size policies, whichever is more restrictive. It is not permitted to download or physically transported Dataset to a collaborating institution or off of the Stanford Secure Data Ecosystem.

2.2. Data tracking system.

Upon receiving data, the Stanford PHS Data Portal logs a) the database name, b) the date that the data were received by and loaded onto the system, c) the identity and IP address of the analyst who was is responsible for the curation of the data along with all activities and d) the filenames along with e) their file types and sizes, and f) the data curating processes. Once the data have been loaded into the PHS Data Portal, permissions are set so that data users can view the data description, variable list and data documentation, and, assuming all requirements have been met, the files needed for their analyses. Each of these can have permissions set at a granular level according to the Data Use Agreement.

System and access logs associated with PHS servers are collected on each server and forwarded to a remote log server, per the requirements of Stanford's Minimum Security

(MINSEC) standards for servers using Stanford's Splunk service. Splunk is software that allows monitoring, searching and inspection of multiple system logs, across time; it is also a powerful tool for analyzing system logs to identify anomalies and trends. The PHS system administrators and the Manager of the PHS Data Core have access to view and analyze log data associated with PHS servers, as does the university's Information Security Office. Logs are retained for at least 18 months.

Data access will be monitored three ways depending on source and destination. File transfers and access via SFTP will be logged and sent to the Splunk service. File access on the storage system will be logged from the storage system and sent to the Splunk service. File access within the systems will also be monitored via the audited service; however care must be taken to not adversely impact system performance with audited and then sent to Splunk for processing and analysis.

It is not permitted to remove Data from the Stanford Secure Data Ecosystem. No investigator is permitted to download or print Data. Investigators may remove research outputs for publication.

In the event data are destroyed, we will record: a) Data Destruction date and b) we will include a Certificate of Disposition including the date we submitted the Certification of Disposition to the Data Proprietor or vendor of the data. The date of destruction will also be recorded in the PHS Data Portal.

2.3. Policies and procedures for the physical removal, transport and transmission of data files.

The Dataset will be received by the Stanford PHS Data Core via encrypted drives. Stanford will not physically remove, transport or transmit Dataset files except to load the data onto the Stanford Secure Data Ecosystem. These standards are outlined in the Stanford PHS Data Management Policy regarding management of the Dataset. All collaborators must access the data on the Stanford Secure Data Ecosystem and are not permitted to remove data. No data will be transmitted to collaborators.

The Dataset access is only shared with individuals who: a) have completed required human subjects and data security training b) have had all electronic devices which may potentially be used to access the data or outputs encrypted c) have attested to and provided proof of encryption either via Stanford University School of Medicine's encryption tracking and verification system (amie.stanford.edu) or equivalent institutional verification such as a letter from an institutional privacy or data security officer attesting to verification of encryption. d) have a "need to know" status with regard to the data and cannot practically work on the project without it and; e) have signed a data use agreement with Stanford PHS stating that they will only use the data for the stated research purposes and that they cannot share the data with any third party; f) as applicable - have institutional review board (IRB) approval for their study or are included as personnel on an IRB approved study, g) as applicable - have obtained Dataset Proprietor approval for their project. These requirements are collected, reviewed and approved for all investigators accessing the Dataset. In the case an investigator is from another institution, there may be additional requirements imposed by their home institution. These standards are outlined in Stanford's Data Security Policies governing the treatment and distribution of PHI and other high-risk data.

Access to approved files will be managed on the PHS Data Portal which allows a PHS Data Administrator to control access to individual files. Any investigator who has completed all regulatory and security requirements will be granted access to the appropriate files on the PHS Data Portal. The investigator can then cut their analytic cohort which will then be ported over to the Stanford Secure Analytic Environment. All analyses, code and resulting work files related to the project will be kept in the project folder. The investigator is expected to also link the final version of any resultant products of research (in most cases a peer-reviewed publication) in the PHS Data Portal. At the end of the project access to the data will be terminated.

The data use agreement will stipulate that only output results from analyses can be downloaded from the Stanford Secure Data Ecosystem – that is, the researcher agrees never to download datasets or subsets of those/analysis files and that all data outputs will conform to Stanford or the Data Proprietor cell size policies, whichever is more restrictive. It is not permitted to download, electronically or physically transported the Dataset to a collaborating institution. All data will remain on the Stanford Secure Data Ecosystem.

2.4. Policies to tailor and restrict data access privileges based on an individual's role on the research team.

The PHS Data Portal allows the Data Administrator to evaluate and review access to the Dataset. All individuals wishing to gain access to the the Dataset must complete all regulatory, security and training requirements as outlined in section 1.3. Access will be granted on a need to know basis and individuals will only be permitted to use files which are included on the researcher's approved the Data Proprietor DUA if applicable.

Data access privileges are based on a given individual's role in the project. Individual approved researchers will have only read access to the master Dataset files. Approved researchers will create their analysis files from the master Dataset; access to those analysis files is limited to the specific individual researcher and his/her approved research team members. System logs track any file uploads, downloads and transfer to the Stanford Secure Data Ecosystem ; the Stanford Secure Data Ecosystem staff will review those logs on a routine basis to ensure that individuals are complying with security requirements.

2.5. Technical safeguards for data access (including password protocols, log-on/log-off protocols, session time out protocols, and encryption for data in motion and data at rest).

For technical infrastructure, please see Sections 1.6, 1.7 and 1.9.

On-prem infrastructure:

This information is repeated in its entirety, verbatim in section 1.6 and 1.9. Portions of this section are repeated verbatim in section 2.5.

All data curation takes place on the Stanford Secure Data Ecosystem (SDE) and researchers are permitted to remove the data from the Stanford SDE. The Stanford SDE includes the PHS Data Portal, HIPAA Compliant Google Cloud Platform (GCP) used to store data, the PHS Secure Servers, the Nero on premise and Nero GCP computational environments. Each component of the Stanford Secure Data Ecosystem is described in detail below.

Data workflows within this ecosystem follow a clearly defined pipeline that ensures data are secured at all steps throughout their lifecycle. Specifically, data are first received on encrypted

disks or via secure FTP transfer; authorized personnel (administrators) ingest the initial data into the PHS Data Portal. Data are stored on the PHS Data Portal in a HIPAA and FedRAMP compliant, multi-redundant, AES256 encrypted datastore on GCP (BigQuery). After performing validation checks, these administrators configure the various access requirements necessary to view the dataset documentation, metadata, and query the underlying data. Next, authenticated researchers may then apply for access through the PHS Data Portal; upon gaining appropriate access, a researcher may then run queries against the data to create an analytical cohort. Finally, an administrator reviews the selected files to ensure that they are consistent with the request submitted to the Data Proprietor. The cohort is then exported to the secure computational environment, wherein researchers can leverage additional statistical and computational software to continue their analyses.

The Dataset will be received by the Stanford PHS Data Core via encrypted disk or secure FTP transfer and saved to Google Cloud Platform (GCP) which has been configured to be HIPAA and FedRAMP compliant. The PHS Data Portal, using Redivis software, manages the ability to see the existence of the datasets and access to meta-data and the data themselves at a tiered, granular level.

All data and metadata on the PHS Data Portal are stored in a multi-redundant, AES256 encrypted datastore on GCP. All connections to the PHS Data Portal must be over an encrypted, TLS 1.2 or greater protocol. Redivis supports single sign on via eduGAIN, InCommon for overview access and dual authentication for full data access. In addition to HIPAA and FedRAMP compliance, the PHS Data Portal supports HTTP Strict Transport Security and is on the HSTS preload list for all major browsers which prevents users from establishing an unencrypted connection. The PHS Data Portal implements multiple layers of logging and monitoring to quickly identify and remedy anomalous behavior.

On-premise computation will take place on either the PHS servers or the Nero Computing On-Premise Environment. Nero specifications can be found below:

System

The data will be received by the Stanford PHS Data Core via encrypted drive or secure FTP transfer and saved to a secure server that abides by Stanford's Computer and Network Usage Policy and Information Security Policy, as well as Minimum Security Standards for servers (MINSEC) for PHI and other sensitive data. Research analytics will be performed on a system using a standard three-tier architecture. Server resources reside with the Stanford Research Computing Center (SRCC); the server and all compute resources itself will be located behind the Stanford University firewalls and there is no incoming web access except as necessary for login, as described below.

Hardware/Software (e.g. name of the anti-virus, anti-spyware, firewall, host intrusion, remote access, etc.)

A Kubernetes-based Linux (Ubuntu) containerized Server environment currently running on Dell and Lenovo server hardware.

Access Control (e.g. password requirements and safeguards, VPN use, WiFi use, file sharing, data access logs, etc.)

Access is strictly limited. All users of system functions go through a two factor authentication process, including a valid username and password and use of the DUO application. Passwords

are evaluated for strength on submission and must adhere to Stanford's password complexity policies. Passwords must be at least 8 characters and include at least one letter, one number and one non alpha-numeric character. Requirements vary for longer passwords. All passwords are stored in an encrypted format.

Remote access is permitted but is protected through secure measures, including use of VPN and two-factor authorization. A log is maintained that indicates when someone has accessed, read or used data.

Physical Environment (e.g. monitor position, printer location, screensaver, etc.)

The server environment is housed in a secure facility maintained by the SRCC group. This facility has 28 cameras that capture motion 24x7. In case of facility power failure, supplemental power is provided by a generator. Testing of the ability to switch between commercial and generator power occurs monthly. The server has also been outfitted with self-encrypting hard drives.

No physical access to the servers is allowed for non-SRCC group personnel. SRCC provides active monitoring of the server for access/security issues, as well as updating security, anti-virus and firewall software, including maintenance of security patches. Network connections are isolated and network access is restricted as noted above in "Access Control"; server-level access is limited to SRCC and a handful of program staff, requiring both valid username and password and the use of a hardware authentication token. Research outputs can be moved encrypted laptops using SSL, encrypted transfer. Both Wifi and Ethernet connections used for access require the use of Stanford VPN to connect. Authorized system administrators and the PHS Data Center Manager can physically access the system components when needed for maintenance activities.

Data Storage (e.g. removal media storage, hard-drive encryption, replicas of the data, etc.)

All data storage and replicas will be controlled by the SRCC-systems group. Replicas are stored at a secured datacenter facility located at Stanford. Data are replicated nightly to a replica system in a different secured Stanford data center in Forsythe Hall on the Stanford campus. Both the regular and replica servers adhere to all of the same security requirements. The server is housed in a secure research data center, the Stanford Research Computing Facility, with keycard access limited only to authorized individuals. In addition, the server rack itself is locked; the key is only available to the system administrators of the server and the Data Center Manager. The Data Center has 28 cameras that capture motion 24x7. In case of facility power failure, supplemental power is provided by a generator. Testing of the ability to switch between commercial and generator power occurs monthly. The server has also been outfitted with self-encrypting hard drives. All statistical software necessary for analyses live on the server so that it is never necessary to download data onto local computers or other electronic devices. In addition to the above safeguards, server system administrators are required to attend two days of information security training each year and they only access the server from bastion host systems specifically secured by the Stanford Information Security Office.

Encryption (type of encryption used for data storage on hard-drives)

All encrypted devices will use AES or better encryption methodology.

The PHS Windows servers are encrypted with BitLocker. The version of Windows running on our infrastructure (Windows Server 2012 R2 and Windows Server 2016) has a recovery password that is FIPS 140-2 compliant per Microsoft FIPS 140-2 Documentation.

The PHS storage servers on which data will be "permanently" located is encrypted using LUKS, the Linux Unified Key Setup which does block-level encryption on all storage drives. We use FIPS approved AES and 3DES ciphers and encryption algorithms across SSH, NFS, and samba. This is per the FIPS Annex Security Requirements.

The system is accessed by approved users via SSH, X2Go via SSH and RDP via the campus VPN. All authorized users of the data will connect to the Stanford network via VPN and will authenticate with two step authentication (SUNetID and password with Duo or Text Code secondary authentication); they will then also authenticate to the server with SUNetID and password.

The Stanford University Network Access Control (SUNAC) service allows granular and configurable control of data access. The service allows the Manager of the Data Core (or the LNAs Local Network Administrator) to control remote access to departmental resources located behind University IT-managed firewalls. Using Workgroup Manager, access can be granted and customized for any PHS Member by SUNet ID. The PHS Data Core manager is responsible for this permissioning process for the Dataset projects.

Stanford University requires encryption of all devices used to access Stanford resources—whether the computers or devices are owned personally or by the University. All devices used to access Stanford systems must be registered in Stanford's internal central tracking system and proof of this registration and encryption must be attached on the user's account in the PHS Data Portal. Each person accessing Stanford systems must fill out a data attestation form annually or whenever there is a change in either their devices or the types of data they access, whichever is most frequent. Each computer or device used to access Stanford School of Medicine systems PHI must have whole disk encryption. Encryption status of every computer and device used to access high-risk data is both tracked and audited continuously per Stanford IRT Information Privacy and Security Policies.

Computer encryption is conducted and verified using the Stanford Whole Disk Encryption (SWDE) service. The SWDE service is for both Windows and Macintosh computers that support native encryption. Once installed, all files are automatically encrypted. The data are protected while the computer is in standby or hibernation mode. This requirement applies to both Stanford and personally owned computers that are used for Stanford activities on the campus network. Computers on campus require SUNET authentication to enter. Personal computers require a 10 character PIN to open the computer. Additionally, all Stanford systems require two step SUNET authentication with a password. Passwords must comply with Stanford Password Guidelines. Passwords must be a minimum of 8 characters and passwords less than 12 characters require a combination of mixed case letters, numbers and special characters.

Every computer using SWDE automatically checks in with a logging and administrative server every 7 days. In the event of loss or theft of a computer with High-risk Data, Stanford policy requires notification of the Information Security Office (ISO). ISO in turn will use the logs to determine if a lost or stolen computer is a "reportable" event, possibly requiring notification of persons whose data may have been lost or stolen. A record of the encrypted devices owned by

each individual requesting Dataset access data which might be potentially used by the individual to access the data are tracked in that individual's record in the PHS Data Portal. This encryption is monitored and verified as described above.

In addition to these security policies, PHS does not permit individuals to download any data, including but not limited to, the Dataset onto individual's computers or devices. All analyses must be conducted in the Secure Data Ecosystem and researchers are only permitted to download outputs which comply to PHS cell size restrictions, or the cell size restrictions of the Data Proprietor, whichever is more restrictive. Printing in the server environment is disabled. Thus, individuals are unable to print the Dataset.

Computer encryption must be conducted and verified using Whole Disk Encryption (WDE). Both Windows and Macintosh computers support native encryption. Once installed, all files must be automatically encrypted. The data must be protected while the computer is in standby or hibernation mode. This requirement applies to both computers owned by the home institution of the researcher and personally owned computers that are used for Stanford activities on the campus network. Computers on campus require SUNet authentication to enter. Personal computers must require a 10 character PIN to open the computer. Additionally, all Stanford systems require two step SUNet authentication with a password.

The drives, discs or media containing the original Dataset files, as received, will be stored in a safe in the locked room accessible only to the Data Core management team (Data Custodian, Data Manager and the Center Director at the Stanford Center for Population Health Sciences). The combination of the safe, a SentrySafe SFW205CWB Water-Resistant Combination Safe, 2X-Large, is only known to the PHS Data Core management team (Data Custodian, Data Manager and the Center Director at the Stanford Center for Population Health Sciences). Further, the data will not be physically moved or transmitted in any way from Stanford without written approval from the Data Proprietor.

Google Cloud Platform Infrastructure:

Nero is configured in two parallel versions. The first is an on premise cluster as described above. The second is almost entirely identical from the researcher's perspective and subject to all the same permissioning controls, but the physical hardware (servers) are managed by Google. These are named Nero and Nero GCP respectively. Nero GCP is used less frequently and only in cases where the analyses taking place are particularly resource intensive.

Google Cloud Platform (GCP) and G-Suite are FedRAMP and NIST 800-53 compliant and listed as such on FedRAMP Marketplace and Google websites. Stanford has a Business Associates Agreement with Google that stipulates that hardware used for Stanford School of Medicine business must be HIPAA compliant. In addition to native encryption and security parameters set by Google.

In addition to the security parameters described above, the PHS Data Portal software (Redivis) manages permissions on an ongoing basis at a tiered, granular level and rescinds data access in the event an individual falls out of compliance with any one requirement. Once an individual has gained access to files, they are permitted to port the necessary cohort and variables over to Nero or the PHS on premise servers. Both Nero and the on premise servers abide by the same security parameters as described above. No downloads or data transfers off the Stanford Secure Data Ecosystem are permitted.

Password Protocols:

Stanford recognizes that individual passwords often represent the weakest link to system access. Stanford's Password Checking is designed to address password vulnerabilities and verifies the length and complexity of passwords. While Stanford implements a complex set of password rules, detailed at the previous URL, the institution also strongly recommends the use of pass phrases as an alternative. Beyond passwords, Stanford implements additional safeguards for systems generally, and for PHS systems specifically.

System logons will via SSH be authenticated via passwords, publickey and GSSAPI protocol combined with Duo two-factor authentication. Network firewall rules will require users to be on the Stanford VPN and authenticated via SUNAC. The passwords and GSSAPI will be connected to Stanford's Kerberos system and all Stanford SUnet password policies will apply. Every connection to the system will require a new Duo two factor prompt. If web applications are presented, they will be authenticated to via Stanford's SAML2 instance, and Duo will be required.

SSH sessions will time out after 30 minutes of inactivity. To enable graphical interfaces, initially X2Go will be used through SSH for access to graphical environments. X2Go allows users to manager graphical sessions, reconnecting to existing sessions or starting a new session. This will permit longer term work in the graphical environment. To enable a better user experience and security, users will be automatically placed into a screen or tsmux session and reconnected when logging back onto the system. There will be a limit on ssh sessions per user; initially start at 5 sessions per user. The screen or tsmux session will lock the screen after a period of inactivity also.

Data transfers will be performed via SSL encryption, either via SSH or HTTPS.

Security Evaluations: Stanford PHS shall periodically (no less than annually) evaluate its processes and systems to ensure continued compliance with obligations imposed by Stanford policy, law, regulation or contract with respect to data confidentiality, integrity, availability, and security. Results of these evaluations will be documented and any remedial action indicated will be taken in a timely manner.

The Stanford Population Health Sciences Data Management Plan has been reviewed and approved by both the funding agency (Stanford PHS) as well as the Stanford Institutional Review Board (IRB) and the Stanford Information Security Office (ISO). The data management plan will be reviewed no less than annually and updates will be made as indicated either by changes to the protocol or technological requirements for optimal data security.

2.6. Research collaborators

In the event a Stanford investigator has a collaborator from another institution, the names of all collaborators and their institutions will be enumerated explicitly in the Collaborator Checklist, DUA application to the Data Proprietor and the Project Staff List on the PHS Data Portal. Additionally, they will be included on the application to the Data Proprietor for data use or reuse and the IRB at Stanford.

No researcher, whether at Stanford or a collaborating institution is permitted to download data. It is not permitted to download or physically transport the Dataset to a collaborating institution. All

data must remain on the Stanford Secure Data Ecosystem managed by PHS and SRCC. All collaborators must work on Stanford systems for both data exploration and analyses. The Data Use Agreement signed by each collaborator will stipulate that only output results from analyses can be downloaded from the server – that is, the researcher agrees never to download datasets or subsets of analytic files and that all data outputs will conform to Stanford or the Data Proprietor's cell size policies, whichever is more restrictive.

Use of the Dataset by collaborators is governed by the same policies that apply to Stanford investigators. All collaborators must follow the same requirements for access as Stanford investigators. They must access and perform all analyses on the Stanford Secure Data Ecosystem. Access to the Dataset is only granted to individuals who have completed all required security and training requirements as mandated by the Data Proprietor and Stanford as well as the requirements of their own university. the Dataset access is only shared with individuals who: a) have completed required human subjects and data security training b) have had all electronic devices which may potentially be used to access the data or outputs encrypted c) have attested to and provided proof of encryption either via Stanford University School of Medicine's encryption tracking and verification system (amie.stanford.edu) or equivalent institutional verification such as a letter from an institutional privacy or data security officer attesting to verification of encryption. d) have a "need to know" status with regard to the data and cannot practically work on the project without it and; e) have signed a data use agreement with Stanford PHS stating that they will only use the data for the stated research purposes and that they cannot share the data with any third party; f) as applicable - have institutional review board (IRB) approval for their study or are included as personnel on an IRB approved study, g) as applicable - have obtained Dataset Proprietor approval for their project. In the case an investigator is from another institution, there may be additional requirements imposed by the collaborator's home institution.

All analyses, code and resulting work files related to the project will be kept in the project folder associated with the project. The investigator is expected to also link the final version of any resultant products of research (in most cases a peer-reviewed publication) in the PHS Data Portal. At the end of the project or proposed analyses, access to the Dataset will be terminated.

The server will be accessed through a VPN connection using two step authentication, just as it is for Stanford investigators.

2.7. Procedures to house additional copy of the data including data transfer.

All data storage and replicas will be controlled by the SRCC-systems group. Replicas are stored at a secured datacenter facility located at Stanford. Data are replicated nightly to a replica system in a different secured Stanford data center in Forsythe Hall on the Stanford campus. Both the regular and replica servers adhere to all of the same security requirements. The servers are housed in a secure research data center, the Stanford Research Computing Facility, with keycard access limited only to authorized individuals. In addition, the server rack itself is locked; the key is only available to the system administrators of the server and the Data Center Manager. The Data Center has 28 cameras that capture motion 24x7. In case of facility power failure, supplemental power is provided by a generator. Testing of the ability to switch between commercial and generator power occurs monthly. The server has also been outfitted with self-encrypting hard drives. All statistical software necessary for analyses live on the server so that it is never necessary to download data onto local computers or other electronic devices.

In addition to the above safeguards, server system administrators are required to attend two days of information security training each year and they only access the server from bastion host systems specifically secured by the Stanford Information Security Office.

3. DATA REPORTING AND PUBLICATION

3.1. Notification of suspected incidents wherein the security and privacy of the PHS data may have been compromised. Includes policies and procedures for responding to potential breaches in the security and privacy of the PHS data.

In the event of a confirmed privacy incident PHS Data Core shall notify the Data Proprietor within 1 day Stanford PHS becomes aware that a determination of unauthorized access to or disclosure of the Data Proprietor Information and/or the Data Proprietor Information Systems has occurred. A written resolution plan for any such incidents will be provided to the Data Proprietor after any such incident. These plans will be drafted in collaboration with Stanford Research Computing, our Privacy Office and our CTO.

In the event of an incident, a remediation plan will be drafted and executed according to Stanford PHS Data Management Policy regarding management of the Dataset.

3.2. Review and approval of Data Management Plan.

Data Management plans are developed by the Stanford PHS Data Core in partnership with the Director of the Stanford Center for Research Computing and their staff. Data Management plans are then reviewed and by the Stanford IRB and the Stanford University Privacy Office.

Since the Data Management Plan is heavily influenced by data security technology and requirements, The Chief Technology Officer of Stanford Research Computing and the PHS Data Core will review the data management arrangements annually or whenever there is a significant change or improvement in data security technology, whichever is more frequent. Any changes or updates to the Data Management Plan will also be included in the annual IRB renewal or modification, whichever is more frequent. If there is a change in Data Management Plan during the DUA period the Data Proprietor will be informed within 5 business days of the proposed changes. Any changes requested by the Data Proprietor will be incorporated into the DMP and submitted for approval at the time the DMP is submitted to the IRB and Stanford Privacy Office.

3.3. Data Management Plan Updates

Data Management plans are developed by the Stanford PHS Data Core leadership in partnership with the Stanford Center for Research Computing. Data Management plans are then reviewed and by the Stanford IRB and the University Privacy Office.

Since the Data Management Plan is heavily influenced by data security technology and requirements, The Chief Technology Officer of Stanford Research Computing and the Data Core Manager will review the data management arrangements annually or whenever there is a significant change or improvement in data security or computational technology, whichever is more frequent. Any changes or updates to the Data Management Plan will also be included in the annual IRB renewal or modification, whichever is more frequent. If there is a change in Data Management Plan during the DUA period PHS will be informed within 5 business days of the proposed changes. Any changes requested by PHS will be incorporated into the DMP and submitted for approval at the time the DMP is submitted to the IRB and Stanford Privacy Office

3.4. PHS Cell Suppression Policy

Stanford PHS does not disclose direct findings, 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 PHS will not identify or report any identifiable pharmacy, provider, prescriber or health plan in any publication or present any cell or formula that can be used to back out a cell with fewer than 11 individuals. All data are reported in aggregate.

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 fewer than 11. In the event that an investigator is unsure they meet the above criteria, they will submit written products for PHS review with the understanding that PHS agrees to make a determination about approval and to notify the user within 4 weeks after receipt of findings. PHS may withhold approval for publication only if it determines that the format in which data are presented may result in identification of individuals, providers or other cells smaller than 11.

4. COMPLETION OF RESEARCH TASKS AND DATA DESTRUCTION

4.1. Process to complete the Certificate of Disposition form and policies and procedures to dispose of data files upon completion of its research.

If required by Data Proprietors, at the conclusion of the project, the research identifiable data will be purged from Stanford systems using packages like "shred". Any retired data disks will be first "zeroed out" or reformatted. Any CD/DVD ROMs or other physical data storage devices will be destroyed or returned to PHS using PHS's preferred delivery mode at that time. The Manager of Research IT will then complete the Certification of Disposition form and send it to the Data Proprietor/vendor of the data as required. Stanford PHS Data Management policy regarding management of PHS data can be found here: <http://med.stanford.edu/phs/data-center-documents-.html>.

4.2. Policies and procedures used to protect PHS data files when individual staff members of research teams (as well as collaborating organizations) terminate their participation in research projects (which may include staff exit interviews and immediate access termination).

As outlined in section 1.4, the PHS Data Core will inform the Data Proprietor of any changes to personnel. In the event an employee terminates their association with Stanford or any project using the Dataset either of their own volition or prior to the completion of a project, the investigator or individual sponsoring the SUNET ID will terminate the sponsorship according to Stanford's Network Access Control Policy. For individuals who will no longer be employees, this termination is handled centrally in HR. Termination of SUNET sponsorship will prevent the employee or affiliate from having any further access to any Stanford system. HR generally conducts exit interviews with employees who leave unexpectedly. However, in the vast majority of cases, departures are due to either promotion, completion of projects, completion of funding or completion of training. Also, in the case where an employee or trainee moves to a different institution and has completed necessary steps to complete projects and retain Stanford affiliation or continue to collaborate with a Stanford researcher, this will be tracked in the PHS Data Portal and appropriate revisions to data access requirements will be completed.

For non-employees or individuals remaining with Stanford but no longer working on a project which uses the Dataset, the termination will be handled by the Stanford PHS Data Core. The SUNAC allows the Data Core Manager to assign permissions on an individual basis and terminate the ability for individuals to access the Dataset and individual project folders on the Stanford PHS server.

For individuals leaving Stanford, termination of all SUNET access and permissions is handled centrally by HR and occurs on the employees last day of work. The PHS Data Core will verify that individuals leaving Stanford no longer have an active SUNET. For individuals remaining at Stanford but leaving a project which uses the Dataset, access to the data will be terminated the last day of that individual's participation in the project.

As individuals are not permitted to download the Dataset and access to data is restricted to those fields necessary for the analyses the individual was working on, termination of data access will prevent any further contact with the Dataset. These procedures are documented in the Stanford PHS Data Management Policy. The study PI will inform the Data Proprietor of changes in staff or access within 10 business days.

4.3. Policies and procedures used to inform Data Proprietors of project staffing changes, including when individual staff member's participation in research projects is terminated, voluntarily or involuntarily.

Procedures for these cases are identical to those outlined above in Section 4.2.

4.4. Policies and procedures to ensure original data files are not used following the completion of the project.

Upon completion of the project, access to the Dataset will be terminated. All systems with the Stanford Secure Data Ecosystem allow granular and configurable control of data access. The service allows the PHS Data Core (or the LNAs Local Network Administrator) to control remote access to departmental resources located behind University IT-managed firewalls. Both the PHS Data Portal and Workgroup Manager enable administrators to terminate access to the Dataset.

In the event all data use agreements have expired and Stanford is required to destroy or return the data, the Manager of the Data Core will provide the Data Proprietor with a certificate of destruction. All the research identifiable data will be purged from the servers using packages like "shred". Any retired data disks will be first "zeroed out" or reformatted. Any CD/DVD ROMs or other physical data storage devices will be destroyed removed from the safe and destroyed. The Manager of the Data Core will then complete the Certification of Disposition Form and submit to the Data Proprietor per the Stanford PHS Data Management Policy.

5. MANAGEMENT OF DATASETS WITH IDENTIFIERS (PHI AND PII)

5.1. De-risking and de-identification of datasets which contain identifiers, either PHI or PII. In general, we allow investigators access to the minimum data necessary to complete their analyses. However, many research questions require either identifiable data (ie, dates of service, small geographic identifiers or linkage by individual) in order to assemble a complete picture of exposures, behaviors and outcomes. Most person readable identifiers are removed upon receipt (eg, name, social security, email, address) and stored in a separate file. In some

cases, PHI or PII likely to be necessary to conduct analyses such as dates of service are retained in master analytic files. Such de-risked files which retain some identifiers are labeled RIF in order to denote the retention of PHI status.

In the event that data linkage by individual will take place, The Stanford PHS Data Core acts as a Trusted Third Party to perform these linkages. Linkages that included PHI or PII will take place on a separate server or GCP instance and be conducted by a member of the PHS Data Core. Depending on the availability of identifiers in both data sources, direct or probabilistic linkage methods will be applied. To conduct these linkages, an encrypted key is assembled using available identifiers which allows for a high fidelity match without necessitating the retention of person-readable identifiers stored with the dataset. At the behest of or with the permission of the Data Proprietor, we will transmit the Dataset with identifiers to third parties in order to conduct linkages where the third party does not use standard linkage products (eg, Census).

The Stanford PHS Data Core team will have sole access to copies of the raw data and the algorithm used to link individuals will be scrambled and encrypted and identifier that are stored in a both a restricted folder on the secure PHS server sequestered for this purpose and in hard copy in a locked safe as described in Section 1.9. The Stanford Data Manager screens all incoming linked data for identifiers and removes all but those necessary for research purposes, using the Safe Harbor Method or other accepted de-risking methods. The Safe Harbor method was laid out by the U.S Department of Health and Human Services in the context of HIPAA Privacy Rule and consists in removing or blurring data specific to 18 key identifiers . Cleaned, de-identified files are then stored in files accessible for research. In some cases the PHS Data Core will also produce an anonymized version of the data. Data will be anonymized using accepted methods unclustering date jittering, 3 digit zip and similar. Depending on the source of the data, there may still be restrictions and requirements for fully anonymized data. The ability to access raw data which includes identifiers is tightly controlled and monitored by a small team at the PHS Data Core.

In the case where analyses require information which is technically deemed identifiable, identifiers will be removed insofar as is possible. Datasets which contain identifiable information will be tagged RIF on the PHS Data Portal. RIFs or datasets which are rich enough that they are effectively identifiable will generally receive a high-risk designation. Access permissions for these datasets will be set accordingly as described in Section 1.3.

As data are enriched, true de-identification becomes increasingly difficult and there is often an inverse relationship between anonymization and utility. Even if fully de-identified, datasets derived from PII or PHI often retain institutional or proprietary risk. Consequently, datasets derived from datasets containing PHI or PII will be evaluated individually and in most cases, treated as high-risk, even if the risk of reidentifying an individual in the dataset is remote. Datasets which have been truly anonymized are often useful for feasibility or variable identification and for some datasets, we will make an anonymized version available. In all cases, the data Proprietor will retain visibility into linkages and the ability to require additional protections or approvals.

Appendix 4: Annual Approval of the Stanford Center for Population Health Sciences Data Management Plan and Security Attestation Questionnaire

See next page.

STANFORD UNIVERSITY

Stanford, CA 94305 [Mail Code 5579]

Darrell M Wilson, M.D.

(650) 723-2012

CHAIR, PANEL ON MEDICAL HUMAN SUBJECTS

Certification of Human Subjects Approvals

Date: July 31, 2021

To: David Rehkopf, PhD, Med/Primary Care and Population Health

Isabella Chu MPH, Melissa L Bondy PhD, Emma Sofia Thonander Hallgren PhD, Ayin Vala MS, Erin Kathleen Delaney BS, Georgina Armstrong, Ian Barry Mathews MS, Lesley Park, Neal Soderquist, Rebecca Maree Miller, Ruth Marinshaw MS, Sean Francis McIntyre MS, Valerie Carolina Meausoone, William Law

From: Darrell M Wilson, M.D., Administrative Panel on Human Subjects in Medical Research

eProtocol Secure Academic Data Ecosystem: Data Receipt, De-identification, De-risking, Risk Assessment, Storage, Curation and Access Control

eProtocol #: 36332

IRB 5 (Registration 4593)

The IRB approved human subjects involvement in your research project on 07/31/2021. **'Prior to subject recruitment and enrollment, if this is: a Cancer-related study, you must obtain Cancer Center Scientific Review Committee (SRC) approval; a CTRU study, you must obtain CTRU approval; a VA study, you must obtain VA R and D Committee approval; and if a contract is involved, it must be signed.'**


The expiration date of this approval is 07/31/2022 at Midnight. If this research is to continue beyond that date, it is your responsibility to submit a Continuing Review application in eProtocol. Research activities must be reviewed and re-approved on or before midnight of the expiration date. The approval period may be less than one year if so determined by the IRB. Proposed changes to approved research must be reviewed and approved prospectively by the IRB. No changes may be initiated without prior approval by the IRB, except where necessary to eliminate apparent immediate hazards to subjects. (Any such exceptions must be reported to the IRB within 10 working days.) Unanticipated problems involving risks to participants or others and other events or information, as defined and listed in the Report Form, must be submitted promptly to the IRB. (See Events and Information that Require Prompt Reporting to the IRB at <http://humansubjects.stanford.edu>.) Upon completion, you must report to the IRB within 30 days.

Please remember that all data, including all signed consent form documents, must be retained for a minimum of three years past the completion of this research. Additional requirements may be imposed by your funding agency, your department, HIPAA, or other entities. (See Policy 1.9 on Retention of and Access to Research Data at <http://doresearch.stanford.edu/policies/research-policy-handbook>)

This institution is in compliance with requirements for protection of human subjects, including 45 CFR 46, 21 CFR 50 and 56, and 38 CFR 16.

Includes: Protocol, attachments.

Waiver of Individual Authorization under 45 CFR 164.512(i)(2)(ii)(A),(B),(C), pursuant to information provided in the HIPAA section of the protocol application.



Darrell M Wilson, M.D., Chair

Approval Period: 07/31/2021 - 07/31/2022

Review Type: EXPEDITED - CONTINUING REVIEW

Funding: Stanford CTRU (Spectrum)

STANFORD UNIVERSITY

Stanford, CA 94305 [Mail Code 5579]

Darrell M Wilson, M.D.

(650) 723-2012

CHAIR, PANEL ON MEDICAL HUMAN SUBJECTS

Certification of Human Subjects Approvals

Date: July 31, 2021

To: David Rehkopf, PhD, Med/Primary Care and Population Health

Isabella Chu MPH, Melissa L Bondy PhD, Emma Sofia Thonander Hallgren PhD, Ayin Vala MS, Erin Kathleen Delaney BS, Georgina Armstrong, Ian Barry Mathews MS, Lesley Park, Neal Soderquist, Rebecca Maree Miller, Ruth Marinshaw MS, Sean Francis McIntyre MS, Valerie Carolina Meausoone, William Law

From: Darrell M Wilson, M.D., Administrative Panel on Human Subjects in Medical Research

eProtocol Secure Academic Data Ecosystem: Data Receipt, De-identification, De-risking, Risk Assessment, Storage, Curation and Access Control

eProtocol #: 36332

IRB 5 (Registration 4593)

The IRB approved human subjects involvement in your research project on 07/31/2021. **'Prior to subject recruitment and enrollment, if this is: a Cancer-related study, you must obtain Cancer Center Scientific Review Committee (SRC) approval; a CTRU study, you must obtain CTRU approval; a VA study, you must obtain VA R and D Committee approval; and if a contract is involved, it must be signed.'**

The expiration date of this approval is 07/31/2022 at Midnight. If this research is to continue beyond that date, it is your responsibility to submit a Continuing Review application in eProtocol. Research activities must be reviewed and re-approved on or before midnight of the expiration date. The approval period may be less than one year if so determined by the IRB. Proposed changes to approved research must be reviewed and approved prospectively by the IRB. No changes may be initiated without prior approval by the IRB, except where necessary to eliminate apparent immediate hazards to subjects. (Any such exceptions must be reported to the IRB within 10 working days.) Unanticipated problems involving risks to participants or others and other events or information, as defined and listed in the Report Form, must be submitted promptly to the IRB. (See Events and Information that Require Prompt Reporting to the IRB at <http://humansubjects.stanford.edu>.) Upon completion, you must report to the IRB within 30 days.

Please remember that all data, including all signed consent form documents, must be retained for a minimum of three years past the completion of this research. Additional requirements may be imposed by your funding agency, your department, HIPAA, or other entities. (See Policy 1.9 on Retention of and Access to Research Data at <http://doresearch.stanford.edu/policies/research-policy-handbook>)

This institution is in compliance with requirements for protection of human subjects, including 45 CFR 46, 21 CFR 50 and 56, and 38 CFR 16.

Includes: Protocol, attachments.

Waiver of Individual Authorization under 45 CFR 164.512(i)(2)(ii)(A),(B),(C), pursuant to information provided in the HIPAA section of the protocol application.



Darrell M Wilson, M.D., Chair

Approval Period: 07/31/2021 - 07/31/2022

Review Type: EXPEDITED - CONTINUING REVIEW

Funding: Stanford CTRU (Spectrum)

Expedited Under Category: 5

FWA00000935 (SU)

Assurance #:

