Electronic RESEARCH INTEGRITY AND COMPLIANCE THE UNIVERSITY OF UTAH

Date: Friday, June 20, 2025 1:29:18 PM Print Close

IRB_00133374 **Created:** 5/7/2020 11:50 AM

PI: Hilary Coon Ph.D. **Submitted:** 5/19/2020

Title: Identification of subgroups at high risk for suicide using electronic health records data

IRB_00133374
1. Contacts and Title

1. Study Introduction

1. **Responsible Investigator:**
Hilary Coon

Email	Training	CoI Date
hilary.coon@utah.edu	12/3/2024 SMG	6/17/2025

a. **Position of the Investigator:**

- (●) **Faculty or** Non-Academic Equivalent
- ( ) Student
- ( ) Staff
- ( ) Resident/Fellow
- ( ) Other

2. **Contact Persons for the Responsible Investigator:**

Name	Email	Training
Marie Gibson	marie.gibson@hci.utah.edu	11/16/2022 MCG
Emily Sullivan	Emily.Sullivan@hsc.utah.edu	2/26/2024 MG
Jennifer West	jennifer.a.west@utah.edu	4/7/2025 SMCG

3. **Guests of the Responsible Investigator:**

Last Name	First Name	E-Mail

There are no items to display

4. **What type of application is being submitted?**
New Study Application (or Amendment/Continuing Review)

5. **Title Of Study:**
Identification of subgroups at high risk for suicide using electronic health records data

6. **Study Purposes and Objectives:**

OVERVIEW. This is a proposed sub-study of the approved IRB (IRB_00044244) that covers the Utah Suicide Genetic Risk Study (USGRS). This sub-study will specifically focus on additional elements in the Electronic Health Records (EHR) to investigate subgroups of individuals at high risk for suicide death. Specifically, these EHR elements are determination of non-lethal suicide attempt and opioid use/misuse from physician notes, as well as standard published behavior questionnaires routinely given to patients (PHQ-9, PROMIS depression, PROMIS anxiety, BDI, BAI, BSI).

PURPOSE. Suicide is a leading cause of death that continues to increase, with over 47,000 preventable suicide deaths per year in U.S.[1,2] Although the research community has made strides in using electronic health records (EHR) and other factors to predict suicidal ideation and behavior, our ability to reliably predict who actually dies by suicide remains close to zero.[3] We know that the incidence of suicide behaviors is far more common (~4%-5% per year) compared to suicide death (~0.01%-0.02% per year).[1] Essentially, only a small fraction of those who engage in suicidal behaviors will go on to die by suicide. Knowledge of who these highest risk individuals are is critically important in directing prevention efforts and development of future targeted interventions. In addition, well over half of suicide deaths occur among individuals with no prior attempts, even after accounting for lack of documentation of attempts in diagnostic codes.[4-6] These “out of the blue” cases suggest one or more high-risk groups exist that are even more elusive to accurate prediction and prevention. In addition to better prediction using EHR and multiple other data resources, incorporating genetic data from suicide deaths may offer additional improvement; genetic factors account for close to 50% of the risk of suicide behaviors [7,8] and suicide death.[9,10]

Suicide is clearly a complex phenomenon, with likely many biological and environmental risk factors. Understanding high-risk subgroups will be key to better prediction. In this sub-study of the Utah Suicide Genetic Risk Study (USGRS, covered by IRB_00044244), we request use of additional data from physician notes and routine questionnaires in the EHR. These data elements will enable more accurate prediction of high-risk suicide death subgroups incorporating presence/absence of suicide attempt, opioid use, and opioid misuse. These are data elements likely critical for defining high-risk groups, but not well captured in diagnostic codes. Through this sub-study, we also request to compare not only across suicide deaths with and without prior attempts (more accurately defined using physician notes), but also compare to data from accidental and undetermined overdose deaths, attempters who have not died by suicide, and matched population non-suicidal controls.

While we anticipate finding differences among these groups, exact characteristics of these differences are unknown, leading to our inability to predict suicide death. It is therefore the case that a substantial part of this study is exploratory. We will use state-of-the-art machine learning (ML) methods to analyze these complex data. ML methods will allow us to bring together comprehensive data resources to objectively identify such subsets. ML methods offer perhaps the best tools to address heterogeneous, complex data problems; these methods deal well with noisy data and multifarious outcomes. ML methods have already substantially improved prediction for suicide behaviors.[12-16] In addition, in a recent Canadian predictive study of *suicide deaths* using factors from health care data [17], ML methods again outperformed traditional logistic regression. These results are promising for our study, which will benefit from many more diverse and comprehensive sets of predictive data, and will have approximately three times the suicide death sample size of the Canadian study.

Of the ~9,000 Utah suicide deaths with demographics, familial data, and 20 years of longitudinal EHR data, the USGRS also currently has DNA from >7,000, which will increase to ~10,000 by 2024. Genome-wide molecular data is in hand for over 5,000 of these Utah suicides, allowing for tests of association with “genetic phenotypes” represented by polygenic risk scores. We will also use demographics, familial data, and longitudinal EHR data from age/sex-matched Utah population controls, allowing for comparisons between suicide deaths (with/without prior attempts), accidental overdose deaths, non-lethal attempts (opioid related and non-opioid related), and non-suicidal controls.

Our study will also collect blood through the Medical Examiner from unselected non-suicide controls for DNA extraction. This collection will provide a Utah-specific, population-based control sample with genetic data which will improve our risk model development. In addition, we will also collect biosamples from deceased individuals in the Intermountain Healthcare (IH) Biorepository who have evidence of a suicide attempt but died by means other than suicide. This comparison cohort will provide a comparison of risks leading to suicide attempt versus suicide death. Genotyping of these samples from unselected non-suicide deaths and IH Biorepository suicide attempts will provide critical, genetically informative, Utah-specific control and comparison data to enhance our risk modeling. All control/comparison cases will be from deceased individuals. Linking will be done to limited-use data through PPR staff at the UPDB, as approved for the suicide deaths.

Aim 1. Extend and refine data elements needed for development of models to predict suicide. Model development will use data elements already approved in the parent USGRS study (IRB_00044244), also questionnaire responses in the EHR (anxiety, depression scales), and natural language processing (NLP) of physician notes for more accurate identification of suicide attempts, opioid use, and opioid misuse.

Aim 2. Machine learning prediction of suicide death vs. suicide attempt and other high-risk subgroups using EHR, demographics, and familial risks Using demographic and electronic health data spanning the period 2000 – present, study differences in records data between high risk subgroups in ~9,000 Utah suicide deaths with and without prior attempts; compare to data records from ~180,000 Utah suicide attempts and age/sex-matched Utah population controls. Apply machine learning methods to formulate models that can predict: a) differences between suicide deaths vs. non-lethal suicide attempts; b) further discrimination within suicide death between suicides with vs. without prior attempts; c) further discrimination subsetting by opioid use/misuse.

Aim 3. Polygenic Risk Score (PRS) analyses of suicide death and non-lethal attempts. Determine if distinct suicide death subgroups identified in aim 2 show additional differences in polygenic risks of conditions associated with suicidality (genetic data will be available on suicide deaths, on unselected Utah controls, and on Utah individuals who attempted suicide but died by other means).

7. **Is this a multi-site study, where more than one site needs IRB approval?**
   ○ Yes ● **No**

8. **Background and Introduction:**

Prediction of suicidal behavior using traditional methods over five decades of research remains little better than chance, as summarized in a comprehensive meta-analysis [11]. More recent studies have turned to large data resources with electronic health records (EHR) and natural language processing (NLP) of physician notes. These studies, several of which have employed machine learning approaches, have made improvements in classification of individuals with suicidal behaviors, with accuracy > 80% in most studies.[12-17] Psychiatric diagnoses and demographic variables (e.g., sex, age, race/ethnicity, marital status) all unsurprisingly have strong effects. In addition, when studies included comprehensive health information above and beyond demographic and psychiatric factors traditionally thought of as contributing to suicide risk, classification was also improved.[15,16] Additional risk factors include medical diagnoses and chronic pain,[16] suggesting that suicide risk is not a simple extension of psychiatric risk.[18,19] However, though classification has improved, our ability to use these risk factors for prediction of future attempts, and particularly to predict suicide death, remains poor.[20]

Importantly for our proposed study, these previous analyses have focused primarily on suicidal behaviors. The studies that have included a focus on suicide death have had to rely upon small sample sizes, and conclusions have therefore been limited. The literature of suicide death risk factors has also often been done on selected populations. Recent examples of this approach include studies of suicide death among patients with bipolar disorder[21] or schizophrenia,[22] among youth,[23] and in individuals in the military.[24-27] Studies of broader population groups often focus on effects of specific risk factors, such as substance use,[28] stressful early life events,[29] method of death,[30] emergency room visits,[31] chronic pain,[32] or sleep disorders,[33] among others. Several recent studies of a relatively large population-based suicide death cohort (N=2,674) between 2000-2013 extracted from individuals who were members of US healthcare systems have taken this approach, selecting an area of risk factors in each focused study. Examples from this work include findings of increased overall use of healthcare systems,[34] and significant increases in psychiatric diagnoses.[35] A recent summary of the actual predictive power of the largest and most powerful studies from this complex literature has concluded that whilst classification accuracy is generally high for suicidal behaviors (>.80 for most studies), the predictive power, particularly for suicide death, remains near zero.[20] There is much room for improvement.

The complexity of outcomes, lack of prediction, and suggestion of additional evidence from genetic studies outlines remaining broad knowledge gaps. Perhaps most importantly, our knowledge of the extent of overlap in risks between those who attempt and those who go on to die by suicide remains understudied, and genetic differences are unknown. Of equal, urgent importance is our lack of understanding of factors differentiating opioid use, opioid abuse, and opioid-overdose attempts, opioid-related suicide.

As of 2023, we are also including Medicaid data.

it is our hypothesis that amount of medical assistance in specific areas will offer insights to risk and protective factors leading to suicide. In 2019, Medicaid ensured access to healthcare for 11% of Utah’s population and since this time Medicaid has expanded coverage and likely cover a greater share of the population. Studies that have examined suicide in Medicaid populations have found that rates of suicide are higher in the Medicaid versus non-Medicaid populations, but precise reasons for this increase are not yet clear. Inclusion of these data with our request to access the Utah APCD will provide our study with more complete data to investigate similarities and differences of suicide deaths vs. non-suicide attempts vs. controls. We also plan to assess insurance status (including continuity of coverage) and type of coverage as potential additional risk factor for suicide. Again, inclusion of the Medicaid data within the APCD will greatly assist in this analysis. In summary, inclusion of the Medicaid data will provide an essential element in determination of individuals at elevated risk for suicide. Better understanding of specific subgroups of individuals with vulnerability will help direct resources to the development of more effective preventions and interventions. Of note, the Medicaid data does not affect the number of participants; this is merely another data source with which to describe existing already-identified participants.

IRB_00133374 **Created:** 5/7/2020 11:50 AM

PI: Hilary Coon Ph.D. **Submitted:** 5/19/2020

Title: Identification of subgroups at high risk for suicide using electronic health records data

IRB_00133374
2. Study Location and Sponsors

2. Study Location and Sponsors

1. **Add all locations applying for approval of research via the University of Utah IRB or Human Research Protection Program (HRPP).**

   Click the appropriate button(s) below to add locations:

	Site Name	Investigators Name	Covered Entity	Sub Sites
View	University of Utah	Hilary Coon	Yes	

2. **Will a Central IRB (CIRB) or Single IRB (SIRB) model be used for review of this study for the sites listed in this application?**
   ◯ Yes ● **No**

3. **Indicate the source(s) of funding obtained or applied for to support this study.**

	Sponsor	Sponsor Type	Sponsor Contact Information	Prime Sponsor	Prime Sponsor Type	OrgID
View	NIH NATIONAL INSTITUTE OF MENTAL HEALTH	Federal Government				10172

4. **Does this study have functions assigned to a Contract Research Organization (CRO)?**
   ◯ Yes ● **No**

5. **Does this study involve use of the Utah Resource for Genetic and Epidemiologic Research (RGE)?**
   Examples: Utah Population Database (UPDB), Utah Cancer Registry (UCR), All Payers Claims Database (APCD), etc.
   ● **Yes** ◯ No

IRB_00133374 **Created:** 5/7/2020 11:50 AM **IRB_00133374** View/Edit

PI: Hilary Coon Ph.D. **Submitted:** 5/19/2020

Title: Identification of subgroups at high risk for suicide using electronic health records data

Addition of a Site

1. **Site Name:**
   University of Utah

2. **Site Principal Investigator**
   ☑ **Mark if Same as Responsible Investigator (syncs with investigator on the first page)**
   Hilary Coon

Email	Training	CoI Date
hilary.coon@utah.edu	12/3/2024 SMG	6/17/2025

   a. **Position of the Site Principal Investigator**
   Faculty or Non-Academic Equivalent

   b. **Will the Site PI consent participants?** ○ Yes ● **No**

3. **Site Contact Persons, if different from the Site PI:**
   ☑ **Mark if Same as Contacts for Responsible Investigator (syncs with contacts on the first page)**

Name	Email	Training
Marie Gibson	marie.gibson@hci.utah.edu	11/16/2022 MCG
Emily Sullivan	Emily.Sullivan@hsc.utah.edu	2/26/2024 MG
Jennifer West	jennifer.a.west@utah.edu	4/7/2025 SMCG

4. **Site Staff and Sub-Investigators**

Name	Email	Training	Obtaining Consent	CoI Date
Sudhanva Manjunath Athreya	u1529299@utah.edu	11/27/2024 MG	☐	3/4/2025
THEODORE BAENDE	u1188099@umail.utah.edu	3/23/2023 MG	☐	6/5/2024
Lisa Baird	lbaird@genetics.utah.edu	1/6/2025 MG	☐	9/11/2024
Amanda Bakian	amanda.bakian@hsc.utah.edu	4/25/2024 MG	☐	2/6/2025
Taylor Blanding	taylor.blanding@hsc.utah.edu	10/9/2024 MG	☐	12/19/2024
Danli Chen	u6010503@utah.edu	10/20/2022 MCG	☐	2/6/2025
Lawrence Cook	larry.cook@hsc.utah.edu	1/24/2023 MCG	☐	3/20/2025

Name	Email	Training	Obtaining Consent	CoI Date
Heather Cummins	heather.cummins@hsc.utah.edu	3/28/2024 SMG	☐	3/26/2025
EMILY DIBLASI	emily.diblasi@hsc.utah.edu	1/16/2024 MG	☐	6/11/2025
Anna Docherty	docherty.anna@gmail.com	7/27/2022 M	☐	2/12/2025
Elliott Ferris	elliott@neuro.utah.edu	1/2/2024 MG	☐	5/13/2025
Evan Goldstein	evan.goldstein@hsc.utah.edu	MC	☐	9/10/2024
Seonggyun Han	seonggyun.han@utah.edu	3/6/2024 M	☐	2/17/2025
Rachael Jasperson	rachael.jasperson@utah.edu	10/1/2024 SM	☐	4/4/2025
Brooks Keeshin	brooks.keeshin@hsc.utah.edu	3/20/2023 MCG	☐	1/6/2025
Brent Kious	brent.kious@hsc.utah.edu	6/2/2025 MG	☐	4/24/2025
Vincent Koppelmans	vincent.koppelmans@utah.edu	6/5/2023 MCG	☐	5/29/2025
Myke Madsen	Myke.madsen@utah.edu	2/27/2023 MG	☐	5/9/2025
Huong Meeks	huong.meeks@hsc.utah.edu	6/25/2024 MG	☐	3/25/2025
Brian Mickey	brian.mickey@utah.edu	11/19/2024 MCG	☐	3/20/2025
Nicole Mihalopoulos	nicole.mihalopoulos@hsc.utah.edu	4/5/2025 MCG	☐	6/3/2025
Eric Monson	eric.monson@hsc.utah.edu	11/4/2024 MG	☐	4/22/2025
Milena Palacios	u1191642@utah.edu	9/5/2022 M	☐	6/12/2025
Warren Pettine	warren.pettine@hsc.utah.edu	5/22/2023 M	☐	9/10/2024
Andrey Shabalin	andrey.shabalin@utah.edu	11/23/2022 MG	☐	9/17/2024
Michael Staley	mstaley@utah.gov	2/6/2024 MG	☐	2/19/2025
Jon Strohmeyer	hunter.strohmeyer@hsc.utah.edu	5/24/2024 M	☐	5/15/2025
Emily Sullivan	Emily.Sullivan@hsc.utah.edu	2/26/2024 MG	☐	6/4/2025
Stuart Telford	russell.telford@hsc.utah.edu	1/16/2024 MG	☐	3/20/2025
Douglas Tharp	doug.tharp@hsc.utah.edu	3/5/2025 SMG	☐	9/25/2024

Name	Email	Training	Obtaining Consent	CoI Date
Andrea Thomas	andrea.thomas@hsc.utah.edu	3/21/2024 MG	☐	2/20/2025
Rachel Weir	Rachel.Weir@hsc.utah.edu	4/14/2025 M	☐	4/8/2025
Zhe Yu	zhe.yu@hci.utah.edu	5/17/2024 MG	☐	9/24/2024
Mingyang Zhang	Mingyang.Zhang@hsc.utah.edu	10/24/2022 SMG	☐	11/22/2023
Sally Zuspan	sally.zuspan@hsc.utah.edu	9/8/2024 MCG	☐	4/4/2025

5. **Site Guests:**

Name	Email	Training

There are no items to display

6. **Select HIPAA coverage for this study:**
Study procedures will be conducted within a HIPAA Covered Entity at this site (HIPAA Privacy Rule applies)

7. **Select the study procedures that will be conducted at this site:**
Data analysis

8. **Select the University of Utah department responsible for this research:**
PSYCHIATRY

9. **Add any additional sites that are part of this performance group**
There are no items to display

IRB_00133374 **Created:** 5/7/2020 11:50 AM

PI: Hilary Coon Ph.D. **Submitted:** 5/19/2020

Title: Identification of subgroups at high risk for suicide using electronic health records data

IRB_00133374
IRB Smart Form

Sponsor Information

a. Are you receiving award or contract management for the sponsored funds through the University of Utah Office of Sponsored Projects?

● **Yes** ○ No

If yes, select the associated OSP Proposal ID/DSS through eAward to link it to the ERICA system.

You must have a fully approved Proposal ID/DSS number through eProposal which will show up in eAward after OSP has integrated the ID. To access the eAward application, use the instructions on the OSP website.

Link to a Proposal ID/DSS through eAward

Proposal ID/DSS: 10073790
PI: COON,HILARY H
Sponsor: NIH NATIONAL INSTITUTE OF MENTAL HEALTH
Prime Sponsor:
Department:
Short Title: COON_NIMH_R01_RENEWAL_JULY24
Sponsor Award Number:
Type: Federal Government
Award Start Date: 7/1/2025
Award End Date: 6/30/2030
Prime Sponsor Type:

IRB_00133374 **Created:** 5/7/2020 11:50 AM

PI: Hilary Coon Ph.D. **Submitted:** 5/19/2020

Title: Identification of subgroups at high risk for suicide using electronic health records data

IRB_00133374
3. Participants

3. Participants

1. Ages of Participants:

7 to 17 years old	(Parental permission and assent form needed)
18 and older	(Consent form needed)

2. Specific age range of participants (e.g., 7-12 years old, 60+, etc.):
records from ages >10 for suicide deaths, non-lethal attempts, opioid use/misuse, matched controls

3. Indicate any vulnerable participant groups (other than children) included:

None

If "Other", please specify:

If "None" and no children are involved, answer the following question.
Has the participant selection process overprotected potential subjects who are considered vulnerable so that they are denied opportunities to participate in research?
○ Yes ● **No**

4. Number of participants to be included and/or enrolled in this entire study, across all study locations: ~20,000 suicide deaths, comparison non-lethal attempts and opioid groups from 1.5 million with EHR data, and 9 million persons overall to accommodate selection for each of the risk factors that will be analyzed for the study

5. Characteristics of Participants/Inclusion Criteria:

There are about 20,000 suicides identified through death certificate data in the UPDB. Of these, about 7000 have already been identified and DNA collected through our long-term collaboration with the OME. This number of cases grows by 700 per year. There are ~1,500,000 cases with Electronic Heath Records (EHR) data from which we can identify non-lethal attempts and opioid use/misuse (using diagnostic codes supplemented by information extracted from physician notes, as described in the Procedures and Statistical Methods sections). The ~9,000,000 records in the UPDB will be used to screen for broader matched controls within the UPDB for epidemiological studies of risk of demographic characteristics. All proposed studies will be done using limited data sets, and no subject contact is requested. Records from family members related to suicides, non-lethal attempts, opioid groups, and controls will be used to determine pedigree risk and aggregated familial risk of co-occurring conditions (familial subtypes).
We plan to continually add records to the study. DNA for genetic studies of the suicide death subgroup will come from the Utah Office of the Medical Examiner (OME) or from the Intermountain Healthcare Biorepository. We anticipate that new samples may add information regarding familial risk and/or will add to information regarding polygenic risk of other psychiatric and behavioral conditions (aim 3 of this project).

Our study will also collect blood through the Medical Examiner from unselected non-suicide controls for DNA extraction. This collection will provide a Utah-specific, population-based control sample with genetic data which will improve our risk model development. In addition, we will also collect biosamples from deceased individuals in the Intermountain Healthcare (IH) Biorepository who have evidence of a suicide attempt but died by means other than suicide. This comparison cohort will provide a comparison of risks leading to suicide attempt versus suicide death. Genotyping of these samples from unselected non-suicide deaths and IH Biorepository suicide attempts will provide critical, genetically informative, Utah-specific control and comparison data to enhance our risk modeling. All control/comparison cases will be from deceased individuals. Linking will be done to limited-use data through PPR staff at the UPDB, as approved for the suicide deaths. Limited use data from suicide deaths who were transgender and their matched controls will be shared with IRB_00136838 (Mihalopoulos, PI) to add to that study's comprehensive ascertainment of records from transgender individuals. This will be a one-way data sharing; no data will be shared from IRB_00136838 to this study.

As of 2023, we are also including Medicaid data. Of note, the Medicaid data does not affect the number of participants; this is merely another data source with which to describe existing already-identified participants. it is our hypothesis that

amount of medical assistance in specific areas will offer insights to risk and protective factors leading to suicide. In 2019, Medicaid ensured access to healthcare for 11% of Utah's population and since this time Medicaid has expanded coverage and likely cover a greater share of the population. Studies that have examined suicide in Medicaid populations have found that rates of suicide are higher in the Medicaid versus non-Medicaid populations, but precise reasons for this increase are not yet clear. Inclusion of these data with our request to access the Utah APCD will provide our study with more complete data to investigate similarities and differences of suicide deaths vs. non-suicide attempts vs. controls. We also plan to assess insurance status (including continuity of coverage) and type of coverage as potential additional risk factor for suicide. Again, inclusion of the Medicaid data within the APCD will greatly assist in this analysis. In summary, inclusion of the Medicaid data will provide an essential element in determination of individuals at elevated risk for suicide. Better understanding of specific subgroups of individuals with vulnerability will help direct resources to the development of more effective preventions and interventions.

6. Participant Exclusion Criteria:

All data records from study subsets (suicide deaths, non-lethal attempts, opioid groups), and from matched population controls will be included.

7. Is a substantial percentage of the participant population anticipated to be non-English speaking?

○ Yes ● **No**

IRB_00133374 **Created:** 5/7/2020 11:50 AM

PI: Hilary Coon Ph.D. **Submitted:** 5/19/2020

Title: Identification of subgroups at high risk for suicide using electronic health records data

IRB_00133374
- Vulnerable Populations

Vulnerable Populations

Justification Requirements for the Inclusion of Vulnerable Populations

1. * **How does the nature of the research require or justify using the proposed subject population:**
This is a records-only study. All analyses will be conducted on limited data sets. All samples to be studied for the genetic part of this study (aim 3) have been or will be collected from 1) suicide deaths or unselected deceased non-suicide controls through the Utah State Office of the Medical Examiner, as approved in the parent study (IRB_00044244), or 2) from suicides or deceased non-suicides with evidence of suicide attempts in the IH Biorepository.
Records from all study subgroups (suicide deaths, non-lethal attempts, opioid groups, matched population controls) will be included to create an unbiased study of risk factors associated with suicide. No contact will be made with living subjects or family members of subjects.

2. * **Would it be possible to conduct the study with other, less vulnerable subjects?**
○ Yes ● **No**

If yes, justify the inclusion of vulnerable subjects:

3. * **Is this population being included primarily for the convenience of the researcher?**
○ Yes ● **No**

If yes, explain:

IRB_00133374 **Created:** 5/7/2020 11:50 AM

PI: Hilary Coon Ph.D. **Submitted:** 5/19/2020

Title: Identification of subgroups at high risk for suicide using electronic health records data

IRB_00133374
4. Study Information

4. Study Information

1. **Design of Study (select all that apply):**

   ☑ **Non-Experimental and/or Descriptive Research Design:**

   Secondary/Archival Data Analysis or Retrospective Chart Review

   ☐ **Experimental and/or Interventional Research Design:**

   There are no items to display

   ☐ **Development of a research resource (repositories, databases, etc.)**

   There are no items to display

   ☐ **Other**

2. **Does your study involve the use of any placebo?**
   ○ Yes ● **No**

3. **Length of entire study, from initiation through closeout:**
   until 7/1/2026

4. **How will participants be recruited or identified for inclusion in the study?**

   a. **Select all methods that will be used:**

   Other

   Suicide deaths are identified through ongoing collection of DNA and tissue on suicides by the Utah Office of the Medical Examiner (OME), and from suicides and non-suicide controls from the IH Biorepository. Collection of comparison DNA and tissue from unselected Utah non-suicide deaths from the Utah OME. Collection of comparison DNA from deceased individuals in the IH Biorepository who have evidence for suicide attempts but died by meas other than suicide. Records data for other case groups will be ascertained from individuals with EHR data (ICD9 and ICD10 diagnostic codes supplemented with data on non-lethal attempts and opioid use/misuse from physician notes). Records from controls will be matched by age/ sex within the UPDB by PPR staff. Limited use data from suicide deaths who were transgender and their matched controls will be shared with IRB_00136838 (Mihalopoulos, PI) to add to that study's comprehensive ascertainment of records from transgender individuals. This will be a one-way data sharing; no data will be shared from IRB_00136838 to this study.

   b. **Describe the recruitment/participant identification process in detail (e.g. who will review charts or records, who can refer participants to the study, where will flyers be posted, how often will recruitment letters be sent, when will follow-up phone calls be made, etc.):**

   All analyses in this study will use records data and genotyping from DNA samples from suicide deaths and deceased comparison cases (unselected, and also non-suicide deaths with prior suicide attempts).

   For suicide deaths and unselected non-suicide Utah controls, blood samples are collected by Office of the Medical Examiner (OME) staff. Records from suicide deaths and controls are transferred by OME staff to the Utah Population Data Base (UPDB) for linking and creation of a limited data set for analysis. Similarly, the IH Biorepository will send UPDB Distribution IDs, along with anonymous sample IDs, from suicides and deceased non-suicide controls to the UPDB for linking and creation of a limited data set for analysis.

For data from non-lethal attempts, ICD codes for attempt (ICD-9 codes E950-E989; V62.84; corresponding ICD-10 codes) will be screened by Data Sciences Services (DSS) Service Recharge Center (SRC) staff. This definition will be supplemented by determination of non-lethal attempt in the physician notes, as specified in the Procedures and Statistical methods sections. DSS/SRC will transfer case identifiers to the UPDB for linking and creation of a limited data set for analysis.

Similarly for opioid use/misuse, ICD codes will be screened (304.0*, 304.7*, 305.5*, 965.*, E850.0, E935.0, and ICD10 equivalents). This definition will be supplemented by opioid use/misuse information in the physician notes, as specified in the Procedures and Statistical methods sections. DSS/SRC will transfer case identifiers to the UPDB for linking and creation of a limited data set for analysis.

Age/sex-matched controls will be assigned by PPR staff within the UPDB once case groups are determined. A limited data set for analysis will be created by PPR staff. Limited use data from suicide deaths who were transgender and their matched controls will be shared with IRB_00136838 (Mihalopoulos, PI) to add to that study's comprehensive ascertainment of records from transgender individuals. This will be a one-way data sharing; no data will be shared from IRB_00136838 to this study.

As of 2023, we are also including Medicaid data. Of note, the Medicaid data does not affect the number of participants; this is merely another data source with which to describe existing already-identified participants. it is our hypothesis that amount of medical assistance in specific areas will offer insights to risk and protective factors leading to suicide. In 2019, Medicaid ensured access to healthcare for 11% of Utah’s population and since this time Medicaid has expanded coverage and likely cover a greater share of the population. Studies that have examined suicide in Medicaid populations have found that rates of suicide are higher in the Medicaid versus non-Medicaid populations, but precise reasons for this increase are not yet clear. Inclusion of these data with our request to access the Utah APCD will provide our study with more complete data to investigate similarities and differences of suicide deaths vs. non-suicide attempts vs. controls. We also plan to assess insurance status (including continuity of coverage) and type of coverage as potential additional risk factor for suicide. Again, inclusion of the Medicaid data within the APCD will greatly assist in this analysis. In summary, inclusion of the Medicaid data will provide an essential element in determination of individuals at elevated risk for suicide. Better understanding of specific subgroups of individuals with vulnerability will help direct resources to the development of more effective preventions and interventions.

5. **How will consent be obtained?**

   Waiver or Alteration of Informed Consent

6. **Describe all the procedures chronologically, from screening/enrollment through study closeout, which will be completed in the research project.**

   Case sets will be ascertained as described above with the assistance of the Utah State Office of the Medical Examiner (OME), the Data Science Services (DSS) Service Recharge Center (SRC), and the UPDB.

   In addition to data elements approved for use in the USGRS parent study (IRB_00044244), this sub-study requests comparisons to additional groups. Group definition will be facilitated using EHR records data in the physician notes. Groups will include suicide deaths, non-lethal suicide attempts, and groups defined by opioid use and misuse. These cases will be compared to data form age/sex-matched population controls. For this study, all records will be limited use data sets. Processing of physician notes will occur as follows.

   Specifically, we request limited data from comparison groups: non-lethal suicide attempts, opioid use/abuse cases. To better characterize these groups, we request data from physician notes from the University of Utah Health Sciences Center (UUHSC) and Intermountain Healthcare. Current expertise and infrastructure exists to process data within the University Data Science Services (DSS) Service Recharge Center (SRC). Notes will be subjected to an NLP pipeline originally developed by a collaborative group at Columbia University, Mount Sinai University, and the Mayo Clinic. This developed pipeline is designed to replicate the well-validated Columbia Suicide Severity Risk Score (C-SSRS) rating measure. The pipeline contains a lexicon of terms and rules for annotation; these will be tested in our electronic health data, with alterations made if necessary due to potentially unique features of our system. Investigators Vikrant Deshmukh and Subhadeep Nag will provide essential assistance with this part of the study.

   Natural Language Processing (NLP) methods will be applied to extract data for more accurate detection of study opioid use/abuse and suicidal behaviors. The NLP classification will benefit from previous published extensive classification, testing, and validation of NLP to extract information regarding suicide behaviors in published work [12-15, 36]. In particular, we will begin work by applying the pipeline described above for suicide ideation and attempt. Similarly, recently published work has based on expert annotation of notes for opioid use/ misuse/ dependence/ overdose will be used for our NLP lexicon development for opioid use/misuse.[37] NLP will make use of the publicly available CLiX NLP platform (https://clinithink.com/clix-notes/, and other available DSS/SRC service tools, Structured Query Language (SQL), and other techniques. The DSS/SRC already has developed processes and queries in support of related work for psychiatric phenotypes. Dr. Abdelrahman will provide expertise as we tailor the pipeline to our system.

Briefly, each document must be searched using a lexicon of terms relating to suicide attempt (e.g., overdose, self-harm, self-injury, suicidal, in addition to key medication names). The process must also filter negating phrases (e.g., “denies suicidal thoughts” or “denies opioid use”). In addition, we will take care to redesign lexicons to ignore “suicide” or “opioid” or related terms when they appear solely in the titles of note sections, or when they appear in reference to persons other than the subject. While refining of each lexicon will take place iteratively, we will benefit from existing published development and validation noted above. Our analyses will additionally adjust for completeness and historical length of the EHR records to control for potential bias due to differences in data depth.[38]

The pipeline will classify evidence of ideation and/or attempt. This information will be supplied to the UPDB, where it will be linked with the case's anonymous numeric id; the resulting data will then be transmitted to the study team for analysis. We note that Drs. Conradt, Staley, and Mickey, who have recently been added to the team, will join our existing analysis group in using the resulting data, as follows: 1) studies of peri- and post-partum depression (Drs. Conradt); 2) basic epidemiology, youth, and also non-binary gender risk groups (Dr. Staley); and 3) treatment resistant depression and bipolar disorder (Dr. Mickey). Within this study, these risk groups will be characterized (e.g., age at death, co-occurring conditions, change in rate of suicide in the subgroup over time); comparisons will be made with other suicide deaths not in these subgroups and/or with the population controls matched to these cases.

We include a subset of the Data Table from the parent study (IRB_00044244). To this, we have added additional study groups (non-lethal attempts, opioid use/misuse); also explicitly list physician notes, and standard published questionnaires in the EHR. All new data for this sub-study is in bold type.

UPDB Data Source	Data Elements to be Accessed	Scientific Justification for Access
Demographic data	Gender, Race, Ethnicity, Marital Status, Education Level, Industry/Occupation Codes, Year of Birth, Year of Death, Age at Death, Birth Month/Year, Death Month/Year	Demographic data is necessary in comparing case groups to matched controls and in determining familial risk patterns. Average age at death from all causes (excluding suicide cases) within each family will be used as indicator of serious chronic disease.
Family relationship information	High risk extended pedigrees	We are looking for families with excess suicides, **excess suicide attempt and excess opioid use/misuse.**
Birth certificate details	Age of parents, details of birth of child of female suicides (for studies of peri-/post-partum depression)	For studies of early life stressors and health outcomes; exposures associated with place of birth could be important to this work. Stressors will be compared to matched controls. Control cases and their families are chosen from the 9 million records in the UPDB, matched for age at death, gender, and birth year.
Death Certificate Details	Primary and secondary causes of death	Identifying suicides from death certificates is critical. Causes of death also will help define familial risks of co-occurring conditions.

Driver License Division	Height/Weight (BMI)	BMI is a known risk factor for suicide death **and suicide attempt**
UUHSC/ IHC	For suicide deaths, relatives of suicides, **non-lethal attempts and their relatives, opioid use/ misuse cases and their relatives,** controls, and relatives of controls, all codes will be provided. Diagnosis age, Diagnosis month/year. For suicide deaths, diagnoses within 1 week of death will be specifically flagged as these may relate to the actual death (e.g., respiratory or heart failure) rather than being risk factors. **Responses to published questionnaires in the EHR will be provided. Physician notes for NLP processing will be provided.**	These comorbid conditions, **questionnaire responses, and NLP-derived non-lethal attempt and opioid behavior data** will be used to determine risk patterns **and define high-risk subgroups for characterization**. Occasionally diagnoses/ procedures/ trauma happen within the same year; month will allow us to determine longitudinal order of these important risks. Study of all codes on cases and family members will provide opportunity to potentially identify novel co-occurring risk factors. Cluster analysis **and machine learning** techniques will be used to identify and aggregate significant co-occurring conditions.
Health Facility Data (Inpatient, Ambulatory surgery, Emergency Department)	For suicide deaths, relatives of suicides, **non-lethal attempts and their relatives, opioid use/ misuse cases and their relatives,** controls, and relatives of controls, all diagnosis and procedure codes will be provided. Diagnosis/procedure age, Diagnosis/procedure month/year. For suicide deaths, diagnoses/procedure within 1 week of death will be specifically flagged as these may relate to the actual death (e.g., respiratory or heart failure) rather than being risk factors.	These comorbid conditions, will be used to determine risk patterns **and define high-risk subgroups for characterization**. Occasionally diagnoses/ procedures/ trauma happen within the same year; month will allow us to determine longitudinal order of these important risks. Study of all codes on cases and family members will provide opportunity to potentially identify novel co-occurring risk factors. Cluster analysis **and machine learning** techniques will be used to identify and aggregate significant co-occurring conditions.
US Census Data for Utah	Sibship size, SES (as indexed by the maximum achieved occupation code), number of offspring, marital history	For older cases in the cohort, these factors will be approximated using census data
Prescription medication data	Prescriptions in the EHR for psychiatric, behavioral, pain,	Prescription data will provide:

	inflammatory, or immune conditions, hormone medications	1) a second data point to validate of ICD9/ICD10 diagnosis (including hormone medications to help identify transgender individuals); 2) data to study characteristics of individuals who have been prescribed opiates who have gone on to die by suicide**; 3) additional data to define opioid use/misuse case groups.**
All Payer Claims Database (APCD)	For suicide deaths, relatives of suicides, **non-lethal attempts and their relatives, opioid use/ misuse cases and their relatives,** controls, and relatives of controls, all diagnosis and procedure codes will be provided. Diagnosis/procedure age, Diagnosis/procedure month/year. For suicide deaths, diagnoses/procedure within 1 week of death will be specifically flagged as these may relate to the actual death (e.g., respiratory or heart failure) rather than being risk factors. Additional the data from the pharmacy file will identify if medications were prescribed and filled for psychiatric, behavioral, pain, inflammatory, or immune conditions, hormone medications	These comorbid conditions, will be used to determine risk patterns **and define high-risk subgroups for characterization**. Occasionally diagnoses/ procedures/ trauma happen within the same year; month will allow us to determine longitudinal order of these important risks. Study of all codes on cases and family members will provide opportunity to potentially identify novel co-occurring risk factors. Cluster analysis **and machine learning** techniques will be used to identify and aggregate significant co-occurring conditions. Prescription data will provide: 1) a second data point to validate of ICD9/ICD10 diagnosis (including hormone medications to help identify transgender individuals); 2) data to study characteristics of individuals who have been prescribed opiates who have gone on to die by suicide**; 3) additional data to define opioid use/misuse case groups.**

In addition to these overarching aims, IRB_00133374 is seeking to define risk factors within what we assume will be many risk subgroups of individuals ascertained for suicidal behavior. Suicidal cases are ascertained, independent of clinical/ demographic subgroup, and controls matched to the suicidal cases. Co-investigators who are experts in studying these subgroups can help us with these studies that center around suicidal behavior as the defining phenotype. Our design compares suicide deaths to other suicide deaths, and to suicide attempts and those who did not die by suicide. It uses the controls matched to this selection to refine these comparisons. At present, we are investigating several subgroups, including suicides with clinical diagnoses, such as bipolar disorder, psychosis, PTSD, treatment resistant depression, or autism. We are also investigating transgender suicides, and suicides that occur postpartum. For these subgroups, our study, and parent study IRB_00044244, can characterize defining aspects of suicides within the subgroup, comparing to other suicides and to the matched controls.

7. **Are all procedures for research purposes only (non-standard or non-standard of care procedures)?**
   ● **Yes** ○ No

If no, list the procedures that are performed for research purposes only (non-standard or non-standard of care procedures):

8. **Is there a safety monitoring plan for this study?**
   ○ Yes ● **No**

9. **Provide a summary of the statistical methods, data analysis, or data interpretation planned for this study. Factors for determining the proposed sample size (e.g., power) should be stated.**

Resulting designations of non-lethal suicide attempt and opioid use/misuse from ICD-9, ICD-10 codes supplemented by the NLP data described above will define comparison categories for the aims of this project. The sample will be divided into training (70%) and testing (30%) sets. The CLiX annotators will use the risk lexicon to mark words and phrases in the document corpus based on established annotation guidelines (e.g., cTAKES, NCBO, Metamap, BeCAS).[39] Annotator agreement will be measures using kappa and F-score tests; agreement will be assumed if scores are>0.7.[40] The annotators will mark up the body of notes iteratively, and agreement will be periodically automatically assessed. Cases with low agreement will be resolved through consensus. During this process, the lexicon and annotation guidelines will be updated as needed.

In addition to this automated validation, manual validation of a subset of records will also be done by DSS/SRC staff. Excellent validation rates are expected, as the lexicon and annotator agreement will be evaluated and updated throughout the process. The resulting NLP suicide behavior classifications (opioid use/misuse, non-lethal suicide attempt) will create one important part of our modeling.

Statistical power for NLP classification. Prediction of behavior classes using NLP will be evaluated through both automated and manual validation. Prior studies using NLP to classify cases with suicidal behaviors and cases with opioid use/misuse have shown high validity at the document and patient level (>80%-90%). [15,36,37,41,42] We propose to make use of these existing lexicons, and will be applying similar methods to records with similar characteristics to these studies. We therefore expect similar high performance in our data, and aim to train the algorithm until we achieve this >80% validity.

Machine Learning. We will develop algorithms that distinguish suicide death, non-lethal attempt, and opioid groups. Predictive modeling will have the advantage of additional comparison to large data resources of non-suicidal individuals, perhaps particularly important in understanding suicide deaths with no prior attempt. Predictors will include the USGRS parent study (IRB_00044244) elements (demographics, ICD-9/10 billing codes, aggregated familial risks) and NLP-derived designation of non-lethal suicidal behaviors and opioid behaviors. To minimize the false positive rates, we will set a high predictive threshold of >0.95 for defining individuals as a member of each group. Individuals whose predicted probability exceeds the threshold value for subgroup status will be so marked. Individuals who do not meet this threshold cutoff will be assigned unknown status. We will estimate area under the receiver operating curve (AUROC) for all the models developed by comparing their prediction performance against the suicide death status assigned by the Medical Examiner, and clinician reviewed presence/ absence of attempts/behaviors and opioid behaviors.

With the assistance of DSS/SRC staff, classical machine learning methods (“base learners”) will be applied. We will use a regular expression rule-based approach (RB) as our primary base learner for predicting group membership. However, we will explore the performance of several classical machine learning approaches (e.g., decision tree, naïve Bayes, support vector machine, and neural networks).[43-49] The different methods will be used because each has strengths and weaknesses highlighted by as-yet unknown aspects of the predictions to be tackled in this aim. These strengths and weaknesses include aspects of methods regarding more or less automation vs. manual model development, and aspects regarding flexibility of feature selection or parameter tuning.

Our approach will involve testing various combinations of base learners in an approach known as ensemble learning, [50] which has been shown to improve performance. Specifically, we will compare three modeling techniques, as follows: 1) bagging: in this method, the base learners are constructed using random independent bootstrap replicates from a training dataset, and the final results are calculated by majority vote; 2) boosting: in this method, the base learners are constructed using weighted versions of the training set, which are dependent on the results of previous base learners; the final result is calculated by weighted majority vote; and 3) random subspace: in this method, the base learners are constructed using random partitions of the feature space. The base learners will be selected from our primary rule-based approach, but also from the available classical machine learning methods listed above. Algorithms will be implemented using the R statistical programming language.

Models will include familial risks in the Utah suicide, Utah non-lethal attempt, opioid groups, and matched control data. Other high-risk subgroups, such as individuals in a particular diagnostic category or categories or in a particular demographic group, may also be studied.)The study team will only undertake other high-risk subgroups after RGE approval of amendment(s) to this application.) Model development will involve ongoing assessment of the performance of the models using precision, recall, and F1-score for detection of discriminating risk factors. For the initial model development, we will randomly divide the data into three sets (70% for training, 15% for validation, and 15% for testing). In addition, we will perform 10-fold cross validation for the training set to minimize test set variability.

Stochastic gradient descent will be used for training, updating all parameters, including embeddings and transition probabilities, at each gradient step. To compare the performance of the ensemble learners, we will conduct the Friedman test with the corresponding Bonferroni-Dun post-hoc test to rank each learner.[51] Tests of the validity of the models for suicide death will make use of the unambiguous determination of suicide death made by the Utah State Office of the Medical Examiner as well as diagnostic codes in the EHR for opioid groups. We anticipate that model fitting will result in predictive variables for suicide death vs. attempt, and potential different high-risk opioid-related subgroups. We note that when a subgroup is also being studied via an external IRB focusing on that subgroup where ascertainment is more comprehensive, we will share data from our cohort with that IRB, but data will not be shared back to this study.

For within-suicide and/or suicide/control comparisons of defined subgroups (including but not limited to trans individuals, postpartum suicide, treatment resistant depression, other clinically-defined subgroups), comparisons will be done using logistic regression, general linear models, and/or parametric/non-parametric clustering techniques that can allow testing of multivariate clustering of co-occurring phenotypes (e.g., nearest neighbor analysis). Models will include covariates, when appropriate, such as age at death, ancestry, sex.

We additionally will be able to test for familial clusters using We propose to use a newly developed method of assessing risks at the individual level of diagnoses in family members (Hanson et al., 2020). This method is now part of a suite of data organization and statistical tools available for our access through the Utah Population Database. Briefly, for relatives of suicide deaths and matched controls, we can aggregate relevant diagnostic data using the PheWAS hierarchical system (Denny et al., 2010). The method then constructs a familial-enrichment matrix and uses K-mediods clustering (Park et al., 2009) to assess strength of familial enrichment and assign a Familial Multi-phenotype Configuration (FMC). K is derived through iterative analyses of the Hamming distance (Hamming, 1950), a well-established metric used for quantifying similarity for discrete values. The FMC can then be used in our subsequent model fitting. This familial risk method has been applied to successfully identify four distinct types of familial cancer with the purpose of driving gene discovery in more homogeneous subtypes (Hanson et al., 2020).

Expected results. We have chosen a study design that includes different ML methods in an ensemble approach that has been previously demonstrated to improve performance.[52,53] Evaluation will include the F1 accuracy measure, [54] and the area under the receiver operating curve for all models. Our data resources exceed the sample sizes used in previous published studies. Importantly, previous studies have not been able to bring to bear large samples of suicide death together with extensive EHR data and additional early life stress scores, SES scores, and familial risks across multiple diagnostic domains.

Polygenic risk scores (PRS). We will Investigate unique and overlapping polygenic risks associated with suicide death vs. non-lethal attempts using molecular data from the Utah suicide deaths and the non-suicide deaths with evidence of attempts, comparing to unselected Utah controls.

We will investigate unique and overlapping polygenic risks associated with the EHR-derived subgroups. Significant PRS associations have already been found for *suicidal behaviors*, (e.g., depression, neuroticism, schizophrenia, and sleep disorders). In addition, our own Utah suicide death data already shows PRS differences between cases with documented prior suicidal behavior and cases without this attribute. Our preliminary data and data of others[75] suggest that polygenic risks of interest will not be confined to psychiatric traits. The enormous capacity to calculate polygenic risks across the spectrum of psychiatric diagnoses, medical diagnoses, and behavioral traits continues to grow. Summary statistics from hundreds of conditions are now freely available on LD Hub;[76] statistics for 597 new traits analyzed in the UKBiobank have been recently uploaded. We will prioritize traits in LD Hub where discovery statistics were derived from large cohorts (N>10,000) to minimize error due to small discovery GWAS. We will then calculate polygenic risk scores (PRSs) of traits and behaviors as described above in our preliminary data using the PRSice v2.0 software.[52] We will additionally apply recently-proposed PRS tuning and thresholding to improve prediction.[83]

Our work will focus on analyses of multiple PRS, recognizing that clusters of PRS from associated diagnoses and traits will very likely perform better than single PRS in describing genetic differences between the EHR-derived suicide death vs. attempt groups. [77,123] In addition to the Utah suicide deaths, the data resources for this aim for non-lethal suicide attempt, and can be replicated using 1,500 individuals in the Mount Sinai BioMe Biobank, 3,200 from the Vanderbilt biobank,[36] and 2,433 attempt cases from the UK Biobank,[36] for a total of at least 7,300 attempt cases with molecular data from other studies.

We propose to test for differences in multiple PRS (MPS) that will exploit the genetic group differences that are likely associated with clusters of multiple diagnoses and traits. To avoid overfitting, we will use cross-validation.[123] We will first estimate individual prediction of each univariate PRS using linear regression, adjusting for residual ancestry effects, and site differences. We will use elastic net regularized regression to avoid overfitting and upward bias in estimates of variance explained (R^2). Elastic net analyses (R library 'glmnet')[124,125] provide robust methods for k-fold cross-validation. Co-I Dr. Dr. Shabalin is an expert in applying elastic net methods to genomic data.[126,127] Elastic net methods are expected to perform well in the face of correlated PRS, as they incorporate a grouping effect.[128] Repeated 10-fold cross-validation will be employed, and independent validation will be possible using new Utah suicides, controls, and external comparison cases.

Statistical power. Anticipated effects in this aim are extrapolated from results in the largest study of suicidality to date.[36] This of 3,200 genotyped individuals with suicidal behavior, and a meta-analyses across an additional in 2,433 suicidal individuals in the UK Biobank found significant PRS associations surpassing the multiple-testing threshold of 1.07E-04 for eight traits of the 466 tests performed. Our project will include similar case numbers with the more extreme phenotype of suicide death. In addition, our preliminary data already suggests the presence of significant effects within a smaller comparison of suicide deaths with vs. without prior attempts. In addition, we propose to study clusterings of multiple PRS, which may increase effect sizes through aggregation.

IRB_00133374 **Created:** 5/7/2020 11:50 AM

PI: Hilary Coon Ph.D. **Submitted:** 5/19/2020

Title: Identification of subgroups at high risk for suicide using electronic health records data

IRB_00133374
- Request for Waiver of Consent

Request for Waiver or Alteration of Consent

* Requested Waivers

	Date Created	Type of Request	Purpose of Waiver Request
View	5/7/2020	Waiver of Informed Consent	For suicide deaths and deceased controls/comparison attempt non-suicides whose data are part of this study, victims cannot be contacted for consent. Extended family information will be used ONLY to determine familial risk, and co-morbid risk; no actual living family members will be studied, and no identifying information will be used from family members. To identify additional subgroups at high risk of suicide death, this study will compare to records of cases with non-lethal suicide attempts and to opioid use/misuse. These groups will be identified using the ~1,500,000 records in the UPDB with EHR data. A waiver is needed for 2 primary reasons: 1) for accurate interpretation of results, study groups must be unbiased, drawn inclusively from population data. Inability to reach some individuals for consent could be significantly associated with factors that would confound conclusions (e.g., SES, minority status). 2) The scope of the work precludes consent of individuals; we will screen state-wide records and anticipate ~200,000 suicide attempts, and >300,000 with opioid use/misuse. As part of this project, only data records will be studied and limited data sets will be used; there will be no participant contact. Finally, to determine significance of the risk of co-occurring conditions, and demographic characteristics, matched controls must be found from the ~9,000,000 records in the UPDB. Again, the scope of this work makes consent impractical. Matching will only be done within the UPDB by PPR staff. Only limited data sets will be used by the analysis team.

IRB_00133374 **Created:** 5/7/2020 11:50 AM

PI: Hilary Coon Ph.D. **Submitted:** 5/19/2020

Title: Identification of subgroups at high risk for suicide using electronic health records data

IRB_00133374
IRB Smart Form

Request for Waiver or Alteration of Consent

1. **Purpose of the Waiver Request:**
   For suicide deaths and deceased controls/comparison attempt non-suicides whose data are part of this study, victims cannot be contacted for consent. Extended family information will be used ONLY to determine familial risk, and co-morbid risk; no actual living family members will be studied, and no identifying information will be used from family members.
   To identify additional subgroups at high risk of suicide death, this study will compare to records of cases with non-lethal suicide attempts and to opioid use/misuse. These groups will be identified using the ~1,500,000 records in the UPDB with EHR data.
   A waiver is needed for 2 primary reasons: 1) for accurate interpretation of results, study groups must be unbiased, drawn inclusively from population data. Inability to reach some individuals for consent could be significantly associated with factors that would confound conclusions (e.g., SES, minority status). 2) The scope of the work precludes consent of individuals; we will screen state-wide records and anticipate ~200,000 suicide attempts, and >300,000 with opioid use/misuse.
   As part of this project, only data records will be studied and limited data sets will be used; there will be no participant contact.
   Finally, to determine significance of the risk of co-occurring conditions, and demographic characteristics, matched controls must be found from the ~9,000,000 records in the UPDB. Again, the scope of this work makes consent impractical. Matching will only be done within the UPDB by PPR staff. Only limited data sets will be used by the analysis team.

2. **Type of Request:**
   Waiver of Informed Consent

3. **List the identifying information you plan to collect or keep a link to (e.g. names, dates, or identification numbers such as social security numbers or medical record numbers, etc.).**
   Identifying information needed by the UPDB PPR staff to link suicide cases to EHR, demographic, and genealogical data will be transferred directly from one of two sources to PPR staff at the UPDB (NOT to the research team). These sources are: 1) For suicide deaths, unselected deceased non-suicide controls, and deceased non-suicides with prior suicide attempts, the Utah State Office of the Medical Examiner (OME) or the Intermountain Healthcare Biorepository. Information from the OME or IH Biorepository for UPDB linking will include State Death Certificate number or other identifiers for linking. Because there is a lag of up to 2 years in the UPDB's acquisition of State Death Certificate numbers and also to provide a secondary check in linking, name, sex, and birth date will also be transferred from the OME or IH Biorepository to the UPDB for linking. 2) For additional records from non-lethal suicide attempt and opioid comparison groups, cases will be identified directly within the UPDB by PPR staff (by relevant ICD-9/ICD-10 codes) with additional cases supplemented by extraction from physician notes using Natural Language Processing (NLP) by DSS/SRC staff (as described elsewhere in this protocol). DSS/SRC staff will send MRN numbers (no other identifiers) of the additional cases identified through NLP to UPDB staff. Identifiers from the OME and from the DSS/SRC listed above will be used only by the UPDB for linking, and will not be transferred to the research personnel on this project.
   After linking the UPDB will send to the research personnel a limited data set, including: high-risk pedigree structure, numeric DNA CCTS number,

demographic data, and UPDB record information as specified in the Data Table.

4. **Explain why the research could not be practicably conducted without using identifiable information. Examples of such explanation could include the following:**
   Identifiers are needed for the PPR staff at the UPDB to link suicide deaths, non-lethal attempts, opioid use/misuse groups, and age/sex-matched population controls to demographic, diagnostic, and family risk information essential to achieve the risk prediction in this study. Only limited-use data will be analyzed. Identifying information will only be for linking within the secure UPDB server environment.

5. **Explain why the research could not practicably be conducted without the waiver or alteration. For example, complete the following sentence "If I had to obtain consent, the research could not be conducted because...":**
   For suicide deaths and other deceased control/comparison cohorts, consent from victims is not possible.
   For records data form non-lethal attempts, opioid use/misuse groups, and matched controls, the scope of the work precludes consent of individuals; we will screen state-wide records and anticipate ~200,000 suicide attempts, and >300,000 with opioid use/misuse. Records from matched controls are needed for risk modeling.
   Records from cases in the study groups and age/sex matched controls must be population-ascertained to ensure the study is unbiased. In addition, numbers of cases and controls for this study will be impractically large.

6. **Explain why the research and privacy risk of the research are no more than *minimal:***
   There will be no subject contact. Identifying information will ONLY be transferred from the OME or DSS/SRC to the UPDB for linking to UPDB records, and will not be used for analysis. The research team will analyze limited data sets from the UPDB after linking.

7. **Describe the measures you will take to ensure the waiver or alteration will not adversely affect the rights and welfare of the *subjects:***
   Any publication involving diagnosis and pedigree structure will be disguised to protect the identity of subjects. No attempt will be made to discover the identity of family members. All publications and presentations will be reviewed by the RGE staff.

8. **Explain how you will, if applicable and appropriate, provide the subjects with additional pertinent information *after* they have participated in the study, or indicate *"Not applicable":***
   Not applicable

IRB_00133374 **Created:** 5/7/2020 11:50 AM

PI: Hilary Coon Ph.D. **Submitted:** 5/19/2020

Title: Identification of subgroups at high risk for suicide using electronic health records data

IRB_00133374
5. Data Monitoring

5. Data Monitoring Plan

1. **Privacy Protections:** Privacy refers to persons and to their interest in controlling access of others to themselves. Privacy can be defined in terms of having control over the extent, timing and circumstances of sharing oneself (physically, behaviorally, or intellectually) with others. **What precautions will be used to ensure subject privacy is protected?**

Select all that apply:

The research intervention is conducted in a private place

Other or additional details (specify):

Other or additional details (specify):

Research done in this study will use only limited information on subjects or family members. All computer files with high-risk pedigree structure and suicide status will be kept on encrypted servers with password protected access. A secure, encrypted database of numeric ids, demographics, co-occurring conditions will be created and maintained in RedCap with the assistance of the CCTS data professionals. Any data analysis will be done only on computers with password access, behind the university firewall, and with full encryption.
For suicide deaths and deceased comparison cases included for genetic data ONLY: While limited data will be analyzed at Utah, data will be further completely de-identified (only de-identified phenotypes with no dates, no zip codes, also de-identified molecular data) will be deposited into the NIMH Repositories (DNA/tissue) and NDA as specified by the NIH. Similarly, completely de-identified molecular and phenotype data will also be stored at the Psychiatric Genomics Consortium Genetics Cluster Computer for future collaborative studies. All repository data will be strictly de-identified. Access to NIMH and PGC repository data is governed by a strict application/ approval process. Detailed pedigree data will not be put into any repository. Use of pedigree data by those accessing the repository could only occur through explicit additional approval by the RGE. All data sent to any repository containing ages over 89 will be combined into a single category of "age 90 and older."

2. **Confidentiality Precautions:** Confidentiality is an extension of the concept of privacy; it refers to the subject's understanding of, and agreement to, the ways identifiable information will be stored and shared. Identifiable information can be printed information, electronic information or visual information such as photographs. **What precautions will be used to maintain the confidentiality of identifiable information?**

Select all that apply:

Storing research data on password protected computers or in locked cabinets or offices

All data that will be transferred or transported outside of the institution will be encrypted

Other or additional details (specify):

Research done in this study will not use any identifying information on subjects or family members. All computer files with high-risk pedigree structure and suicide status will be kept on encrypted servers with password protected access. A secure, encrypted database of numeric ids, demographics, co-occurring conditions will be created and maintained in RedCap with the assistance of the CCTS data professionals. Any data analysis will be done only on computers with password access, behind the university firewall, and with full encryption.
For suicide deaths and deceased comparison cases included for genetic data ONLY: Completely de-identified data (phenotypes without dates, no zip codes, sequence and genotype data) will be deposited into the NIMH Repositories (DNA/tissue) and NDA as specified by the NIH. Similarly, completely de-identified molecular and phenotype data (no dates, no zip codes) will also be stored at the Psychiatric Genomics Consortium Genetics Cluster Computer for future collaborative studies. All repository data will be strictly de-identified. Access to NIMH and PGC repository data is governed by a strict application/ approval process. Detailed pedigree data will not be put into any repository. Use of pedigree data by those accessing the repository could only occur through explicit additional approval by the RGE. All data sent to any repository containing ages over 89 will be combined into a single category of "age 90 and older."

3. **Will photos, audio recordings, or video recordings, or medical images of participants be made during the study?**

○ Yes ● **No**

If yes, describe the recording/images and what will become of them after creation (e.g., shown at scientific meetings, stored in the medical/research record, transcribed, erased, etc.):

4. **How will study data and documentation be monitored throughout the study?**

Select all that apply:

Periodic review and confirmation of participant eligibility

Periodic review of the transfer/transcription of data from the original source to the research record

Other additional details (specify):

5. **Who will be the primary monitor of the study data and documentation?**

Select all that apply:

Principal Investigator

Other or additional details (specify):

Other or additional details (specify):

Investigators at the OME (Erik Christensen) and the UPDB will also monitor data (these individuals will be the only study personnel with access to identifying data).
The NIH will oversee use of de-identified data that are part of the NIMH repository as a result of this study. The PGC only allows access for data analysis after an approval process similar to that of the NIH.

6. **How often is study data and documentation monitoring planned (e.g., monthly, twice a year, annually, after N participants are enrolled, etc.)?**

Data monitoring will occur each time data are transferred from the OME or from the DSS/SRC to the UPDB, and each time data are obtained from the UPDB (approximately twice per year). Additional oversight by the RGE of representation of pedigree data will be done any time a publication is submitted.
For suicide deaths and deceased comparison cases included for genetic data, results will additionally be checked with each deposition into the NIMH repositories (at the end of the award period). Samples and data will be checked for quality prior to sending to the NIMH repository.

IRB_00133374 **Created:** 5/7/2020 11:50 AM

PI: Hilary Coon Ph.D. **Submitted:** 5/19/2020

Title: Identification of subgroups at high risk for suicide using electronic health records data

IRB_00133374
6. Risks and Benefits

6. Risks and Benefits

1. Describe the reasonable foreseeable risks or discomforts to the participants:

This is a records-only study. The researchers will have access only to limited data for suicide deaths, non-lethal attempts, opioid use/misuse cases, controls, or family members of any group. There will be no procedures performed on subjects. There will be no direct contact of any subject or family member.
The only risk is potential loss of confidentiality. Data will be encrypted before being transferred from OME or DSS/SRC to the UPDB. Data firewalls are firmly in place at each center to safeguard data. Data transfers between the IH Biorepository and UPDB will be handled via the Master Linkage File.
Data transfers to the NIMH repositories or the PGC consortium will include only strictly de-identified data (no dates or zip code data); however, transfers will still occur only using fully encrypted devices. All data sent to any repository containing ages over 89 will be combined into a single category of "age 90 and older."

2. Describe the potential benefits to society AND to participants (do not include compensation):

In this study, we hope that by identifying subgroups at high risk for suicide death, we may begin to understand and predict risk more accurately. Achieving this goal would lead to the development of more effective, personalized prevention. Because de-identified samples and data from suicide deaths in this study will be part of the NIH repositories and the Psychiatric Genomics Consortium collaborative studies, there is a greater chance that high-impact findings will result from this work.

3. Are there any costs to the participants from participation in research?

○ Yes ● **No**

If yes, specify:

4. Is there any compensation to the participants?

○ Yes ● **No**

a. **If yes, answer the following:**
Specify overall amount:

b. **Specify when participants will be paid (e.g. at each visit, at end of study, etc.):**

c. **If applicable, please specify payment by visit or other time interval (e.g. $10 per visit, etc.):**

d. **If applicable, explain plan for prorating payments if participant does not complete the study:**

IRB_00133374 **Created:** 5/7/2020 11:50 AM

PI: Hilary Coon Ph.D. **Submitted:** 5/19/2020

Title: Identification of subgroups at high risk for suicide using electronic health records data

IRB_00133374
7. HIPAA & the Covered Entity

7. HIPAA and the Covered Entity

1. **Does this study involve Protected Health Information (PHI) or de-identified health information?**
   ● **Yes** ○ No

   a. **Select the method(s) of authorization that will be used:**

   Waiver or Alteration of Authorization

   Limited data set

   b. **Will PHI be disclosed outside the Covered Entity?**
   ○ Yes ● **No**

Does this study involve any of the following:

2. **The investigational use of a drug?**
   ○ Yes ● **No**
   Mark yes, for an expanded access application.

3. **The investigational use of a medical device or humanitarian use device?**
   ○ Yes ● **No**
   Mark yes, for an expanded access application.

4. **The investigational use of a dietary supplement, food, or cosmetic?**
   ○ Yes ○ No

5. **Is this an investigator-initiated drug or device trial lead by the Principal Investigator'**
   ○ Yes ● **No**
   All investigator-initiated drug or device trials are required to have a full research protocol attached to the Documents and Attachments page.

6. **Will this study involve the use of an imaging modality from the department of Radiology?**
   ○ Yes ● **No**

7. **Exposure to radioisotopes or ionizing radiation?**
   ○ Yes ● **No**

8. **Genetic testing and/or analysis of genetic data?**
   ● **Yes** ○ No

9. **Creating or sending data and/or samples to a repository to be saved for future research uses?**
   ● **Yes** ○ No

10. **Are you:**
    - **Collecting samples of blood, organs or tissues from participants for research purposes;**
    - **Introducing Recombinant or Synthetic Nucleic Acids (e.g. viral vectors, oligonucleotides) or cells containing recombinant nucleic acids (e.g. CAR-T) into participants; OR**
    - **Introducing other biological materials (e.g. bacteria, viruses) into participants.**

    ○ Yes ● **No**

11. **Does this study involve any of the following?**
    - **Cancer Patients**
    - **Cancer Hypothesis**
    - **Cancer risk reduction**
    - **Cancer prevention**

    ○ Yes ● **No**

12. **Any component of the Clinical and Translational Science Institute (CTSI)?**
    ● **Yes** ○ No

    The Clinical Research Unit (CRU)?
    ○ Yes ● **No**

IRB_00133374 **Created:** 5/7/2020 11:50 AM

IRB_00133374 - Request for Waiver of Authorization

PI: Hilary Coon Ph.D. **Submitted:** 5/19/2020

Title: Identification of subgroups at high risk for suicide using electronic health records data

Request for Waiver or Alteration of Authorization

Request for Waiver of Authorization for Recruitment Only

This option must only be used if you are reviewing PHI in order to identify eligible participants BEFORE approaching them to obtain consent and authorization. All other waiver requests must be entered below.

Other Requests for Waivers of Authorization:

Click "Add" below to add a new waiver request to this application.

Click the waiver name link to edit a waiver that has already been created.

To delete a waiver request, contact the IRB.

	Date Created	Type of Request	Purpose of Waiver Request
View	5/7/2020	Waiver of Authorization	For suicide deaths and deceased controls/comparison attempt non-suicides whose data are part of this study, victims cannot be contacted for consent. Extended family information will be used ONLY to determine familial risk, and co-morbid risk; no actual living family members will be studied, and no identifying information will be used from family members. To identify additional subgroups at high risk of suicide death, this study will compare to records of cases with non-lethal suicide attempts and to opioid use/misuse. These groups will be identified using the ~1,500,000 records in the UPDB with EHR data. A waiver is needed for 2 primary reasons: 1) for accurate interpretation of results, study groups must be unbiased, drawn inclusively from population data. Inability to reach some individuals for consent could be significantly associated with factors that would confound conclusions (e.g., SES, minority status). 2) The scope of the work precludes consent of individuals; we will screen state-wide records and anticipate ~200,000 suicide attempts, and >300,000 with opioid use/misuse. As part of this project, only data records will be studied and limited data sets will be used; there will be no participant contact. Finally, to determine significance of the risk of co-occurring conditions, and demographic characteristics, matched controls must be found from the ~9,000,000 records in the UPDB. Again, the scope of this work makes consent impractical. Matching will only be done within the UPDB by PPR staff. Only limited data sets will be used by the analysis team.

IRB_00133374 **Created:** 5/7/2020 11:50 AM **IRB_00133374**
IRB Smart Form

PI: Hilary Coon Ph.D. **Submitted:** 5/19/2020

Title: Identification of subgroups at high risk for suicide using electronic health records data

Request for Waiver or Alteration of Authorization

1. **Purpose of the Waiver Request:**
For suicide deaths and deceased controls/comparison attempt non-suicides whose data are part of this study, victims cannot be contacted for consent. Extended family information will be used ONLY to determine familial risk, and co-morbid risk; no actual living family members will be studied, and no identifying information will be used from family members.
To identify additional subgroups at high risk of suicide death, this study will compare to records of cases with non-lethal suicide attempts and to opioid use/misuse. These groups will be identified using the ~1,500,000 records in the UPDB with EHR data.
A waiver is needed for 2 primary reasons: 1) for accurate interpretation of results, study groups must be unbiased, drawn inclusively from population data. Inability to reach some individuals for consent could be significantly associated with factors that would confound conclusions (e.g., SES, minority status). 2) The scope of the work precludes consent of individuals; we will screen state-wide records and anticipate ~200,000 suicide attempts, and >300,000 with opioid use/misuse.
As part of this project, only data records will be studied and limited data sets will be used; there will be no participant contact.
Finally, to determine significance of the risk of co-occurring conditions, and demographic characteristics, matched controls must be found from the ~9,000,000 records in the UPDB. Again, the scope of this work makes consent impractical. Matching will only be done within the UPDB by PPR staff. Only limited data sets will be used by the analysis team.

2. **Type of Request:**
Waiver of Authorization

3. **List the identifying information you plan to collect or keep a link to (e.g. names, dates, or identification numbers such as social security numbers or medical record numbers, etc).**
Identifying information needed by the UPDB PPR staff to link suicide cases to EHR, demographic, and genealogical data will be transferred directly from one of two sources to PPR staff at the UPDB (NOT to the research team). These sources are: 1) For suicide deaths, unselected deceased non-suicide controls, and deceased non-suicides with prior suicide attempts, the Utah State Office of the Medical Examiner (OME) or the Intermountain Healthcare Biorepository. Information from the OME or IH Biorepository for UPDB linking will include State Death Certificate number or other identifiers for linking. Because there is a lag of up to 2 years in the UPDB's acquisition of State Death Certificate numbers and also to provide a secondary check in linking, name, sex, and birth date will also be transferred from the OME or IH Biorepository to the UPDB for linking. 2) For additional records from non-lethal suicide attempt and opioid comparison groups, cases will be identified directly within the UPDB by PPR staff (by relevant ICD-9/ICD-10 codes) with additional cases supplemented by extraction from physician notes using Natural Language Processing (NLP) by DSS/SRC staff (as described elsewhere in this protocol). DSS/SRC staff will send MRN numbers (no other identifiers) of the additional cases identified through NLP to UPDB staff.
Identifiers from the OME and from the DSS/SRC listed above will be used only by the UPDB for linking, and will not be transferred to the research personnel on this project.
After linking the UPDB will send to the research personnel a limited data set, including: high-risk pedigree structure, numeric DNA CCTS number, demographic data, and UPDB record information as specified in the Data Table.

4. **Explain why the *PHI* to be used or disclosed is the minimum necessary to accomplish the research objectives:**

Identifiers are needed for the PPR staff at the UPDB to link suicide deaths, non-lethal attempts, opioid use/misuse groups, and age/sex-matched population controls to demographic, diagnostic, and family risk information essential to achieve the risk prediction in this study. Only limited-use data will be analyzed. Identifying information will only be for linking within the secure UPDB server environment.

5. **Explain why the research could not practicably be conducted without the waiver of authorization.** *For example, complete the following sentence: "If I had to obtain authorization, the research could not be conducted because..."*
   For suicide deaths and other deceased control/comparison cohorts, consent from victims is not possible.
   For additional records from non-lethal attempts, opioid use/misuse groups, and matched controls, the scope of the work precludes consent of individuals; we will screen state-wide records and anticipate ~200,000 suicide attempts, and >300,000 with opioid use/misuse. Records from matched controls are needed for risk modeling.
   Records from cases in the study groups and age/sex matched controls must be population-ascertained to ensure the study is unbiased. In addition, numbers of cases and controls for this study will be impractically large.

6. **Describe your plan to protect the identifiers from improper use and disclosure, and indicate where the *PHI* will be stored and who will have access:**
   There will be no subject contact. Identifying information will ONLY be transferred from the OME or DSS/SRC to the UPDB for linking to UPDB records, and will not be used for analysis. The research team will analyze limited data sets from the UPDB after linking.

7. **The identifiers must be destroyed at the earliest opportunity consistent with conduct of the research, unless there is a health or research justification for retaining the identifiers or such retention is otherwise required by law. Describe how and when you will destroy the identifiers, or justify their retention:**
   Identifiers will only continue to exist for purposes other than this study, at the Office of the Medical Examiner (for suicide deaths, as required by law), within the EHR data warehouses (for clinical purposes), at the UPDB (for general research use). No identifiers are retained by this study team. The specific association of identifiers within the UPDB for this study will be eliminated at the end of the study.

8. **Describe the measures you will take to ensure the PHI will not be reused or disclosed to any other person or entity, except as required by law, for authorized oversight of the research study, or for other research approved by the IRB:**
   Because the research personnel will never have the identifiers, they will not be able to disclose them to any other entity.

IRB_00133374 **Created:** 5/7/2020 11:50 AM

IRB_00133374 - Limited Data Set Agreement

PI: Hilary Coon Ph.D. **Submitted:** 5/19/2020

Title: Identification of subgroups at high risk for suicide using electronic health records data

Limited Data Set Statement and Assurance

This assurance applies to the following part(s) of this study (select all that apply):

☐	**All of the information used or disclosed in this study**	
☐	**The information received or collected from these sources:**	
☑	**The information shared with or disclosed to these groups:**	NIMH, Psychiatric Genomics Consortium

Data Use Assurance:

1. As an employee of the University of Utah and the Principal Investigator for the attached study, I understand that I must comply with the requirements below regarding the permitted uses and disclosure of the limited data set I am receiving from the University.
2. I have described the limited data set with specificity in the Protocol submitted with this form.
3. I have described who will be permitted to use or receive the limited data set in the same Protocol.
4. If I am disclosing the limited data set outside the Covered Entity, I am submitting, in addition to this form, a Limited Data Set Statement and Assurance (if the recipient is within the University of Utah) or a Data Use Agreement (if the recipient is outside the University of Utah) signed by an individual able to bind the entity receiving the limited data set.
5. I understand that the use of the limited data set is governed by federal law (45 CFR Parts 160 and 164, particularly 164.514(e)).
6. I agree to the following:
    a. I will only use or disclose the information as described or permitted in the Protocol, or as permitted in writing by the Institutional Review Board;
    b. I will use appropriate safeguards, which I have described in the Protocol, to prevent use or disclosure of the information in any ways outside the Protocol or as permitted in writing by the Institutional Review Board;
    c. I will promptly report to the IRB any use or disclosure of the information not provided for in the Protocol;
    d. I will take all reasonable measures to ensure that any agents, including any subcontractors, to whom I provide the limited data set will follow the same restrictions and conditions regarding that information that I have set forth in the Protocol, and will report any violations to the Institutional Review Board (801 581-3655); and
    e. I will not attempt to identify the information and will not contact the individuals.

Limited Data Set Statement.
I declare that none of the following types of information, regarding subjects or relatives, employers, or household members of subjects, are used in this study:

1. Names;
2. Postal address information (but town or city, State, and zip code may be kept);
3. Telephone numbers;

4. Fax numbers;
5. Electronic mail addresses;
6. Social security numbers;
7. Medical record numbers;
8. Health plan beneficiary numbers;
9. Account numbers;
10. Certificate/license numbers;
11. Vehicle identifiers and serial numbers, including license plate numbers;
12. Device identifiers and serial numbers;
13. Web Universal Resource Locators (URLs);
14. Internet Protocol (IP) address numbers;
15. Biometric identifiers, including finger and voice prints; and
16. Full face photographic images and any comparable images.

IRB_00133374 **Created:** 5/7/2020 11:50 AM **IRB_00133374** - Genetic Research

PI: Hilary Coon Ph.D. **Submitted:** 5/19/2020

Title: Identification of subgroups at high risk for suicide using electronic health records data

Genetic Research

1. **Describe the risks to participants in regard to genetic testing, including applicable risks to privacy and confidentiality, as well as psychological and social risks.**

   This is a records-only study. The researchers will have access only to limited data for suicide deaths, non-lethal attempts, opioid use/misuse cases, controls, or family members of any group. There will be no procedures performed on subjects. There will be no direct contact of any subject or family member.
   The only risk is potential loss of confidentiality. Data will be encrypted before being transferred from OME or DSS/SRC to the UPDB. Data firewalls are firmly in place at each center to safeguard data. Data transfers between the IH Biorepository and UPDB will be handled via the Master Linkage File.
   Data transfers to the NIMH repositories or the PGC consortium will include only strictly de-identified data (no dates or zip code data); however, transfers will still occur only using fully encrypted devices. All data sent to any repository containing ages over 89 will be combined into a single category of "age 90 and older."

2. **Describe the privacy protections in place for participants in regard to genetic testing. This includes how family member privacy will be protected.**
   Research done in this study will use only limited information on subjects or family members. All computer files with high-risk pedigree structure and suicide status will be kept on encrypted servers with password protected access. A secure, encrypted database of numeric ids, demographics, co-occurring conditions will be created and maintained in RedCap with the assistance of the CCTS data professionals. Any data analysis will be done only on computers with password access, behind the university firewall, and with full encryption.
   For suicide deaths and deceased comparison cases included for genetic data ONLY: While limited data will be analyzed at Utah, data will be further completely de-identified (only de-identified phenotypes with no dates, no zip codes, also de-identified molecular data) will be deposited into the NIMH Repositories (DNA/tissue) and NDA as specified by the NIH. Similarly, completely de-identified molecular and phenotype data will also be stored at the Psychiatric Genomics Consortium Genetics Cluster Computer for future collaborative studies. All repository data will be strictly de-identified. Access to NIMH and PGC repository data is governed by a strict application/ approval process. Detailed pedigree data will not be put into any repository. Use of pedigree data by those accessing the repository could only occur through explicit additional approval by the RGE. All data sent to any repository containing ages over 89 will be combined into a single category of "age 90 and older."

3. **Are you performing whole genome or whole exome sequencing?**
   ○ Yes ● **No**

4. **Describe the confidentiality protections in place for participants' genetic information. Discuss if and how data will be shared and protected outside the local study team.**
   Research done in this study will not use any identifying information on subjects or family members. All computer files with high-risk pedigree structure and suicide status will be kept on encrypted servers with password protected access. A secure, encrypted database of numeric ids, demographics, co-occurring conditions will be created and maintained in RedCap with the assistance of the CCTS data professionals. Any data analysis will be done only on computers with password access, behind the university firewall, and with full encryption.
   For suicide deaths and deceased comparison cases included for genetic data ONLY: Completely de-identified data (phenotypes without dates, no zip codes, sequence and genotype data) will be deposited into the NIMH Repositories (DNA/tissue) and NDA as specified by the NIH. Similarly, completely de-identified molecular and phenotype data (no dates, no zip codes) will also be stored at the Psychiatric Genomics Consortium Genetics Cluster Computer for future collaborative studies. All repository data will be strictly de-identified. Access to NIMH and PGC repository data is governed by a strict application/ approval process. Detailed pedigree data will not be put into any repository. Use of pedigree data by those accessing the repository could only occur through explicit additional approval by the RGE. All data sent to any repository containing ages over 89 will be combined into a single category of "age 90 and older."

5. **Will incidental findings relevant to individuals or families be communicated to the participants?**
   ○ Yes ● **No**
   If yes, answer the questions below:

   a. **Describe the process for determining which incidental findings will be returned to the participants. Describe the information and expert consultation that will be used to make this determination.**

b. **Indicate the process that will be used to return information about incidental finding to participants:**
There are no items to display
If Other, describe and justify the process that will be used:

6. **Will genetic information or samples be submitted to a national or international database because of this research?**

- ◯ Yes
- ◯ No

IRB_00133374 **Created:** 5/7/2020 11:50 AM

PI: Hilary Coon Ph.D. **Submitted:** 5/19/2020

Title: Identification of subgroups at high risk for suicide using electronic health records data

IRB_00133374
- Data & Tissue Banking

Data & Tissue Banking

1. **Select the items that will be banked:**
   - [x] **Biological samples**
   - [x] **Data**

   List all the specific, participant information (identifiable and non-identifiable) that will be contributed to the repository (e.g., name, date of birth, phone number, age, gender, diagnosis, treatment status, outcome, date of collection, etc.):
   Data and samples will be kept a the University of Utah Psychiatry Department Genetics Lab (limited data set).
   Only strictly de-identified data/samples from suicide deaths and other deceased comparison/control cohorts will be at repositories: anonymous numeric identifier, sex, suicide status, age at death, aggregated yes/no status from EHR of major depression, anxiety, bipolar disorder, schizophrenia; molecular data.
   Type(s) of samples to be collected:
   DNA from blood; cells in culture from skin biopsy

2. **What type(s) of future research will be allowed on the data/samples?** Research on samples/data at the U of U Psychiatry lab must be IRB-approved.

   For the de-identified samples/data at the NIMH repository, projects must be be submitted by qualified researchers, then reviewed and approved/denied by the NIMH.
   Only de-identified data (not samples) will be kept by the PGC. Projects will be related to psychiatric research. Projects will be submitted by qualified researchers, then reviewed and approved/denied by PGC leadership.

3. **Who manages the repository and where will the data/samples be stored?**
   University of Utah Department of Psychiatry Genetics Laboratory.
   NIMH repository (Rutgers University Cell and DNA Repository, RUCDR).
   Psychiatric Genomics Consortium Data Access Committee; data are stored on the PGC Genetic Cluster Computer (LISA server) managed by the Dutch National Computing and Networking Services.

4. **Indicate whether the data/samples in the repository will be identifiable directly or through a code/link.**

   a. **Select one of the following options:**
   OPTION 2: Some data/samples will be identifiable and some data/samples will be de-identified to one or more individuals who have responsibilities to manage or oversee the repository.

   b. **If you selected OPTION 1 or 2 above, describe the process for managing the identifiable data:**

   ### Who will manage and have access to the identifiable data?
   Utah State Office of the Medical Examiner and IH Biorepository staff and the U Data Science Services Service Recharge Center will have access to identifiers; data from these sources will be given to PPR staff at the UPDB who will create a limited data set for the research analysis team.
   For external repositories (NIH, PGC), only completely de-identified data (no dates) will be transferred.

   ### Where will the data be kept?
   At Utah: CHPC secure servers.

   ### How will the data be kept confidential?
   All limited data will be analyzed on CHPC secure servers; these are password protected and only the research team will have access. All publications/presentations will be additionally reviewed by the RGE before public presentation or publication.

   c. **If you selected OPTION 2 or 3 above, describe the process for de-identifying the data/samples:**

   ### Who will de-identify the data/samples?
   At Utah: A limited data set will be created by PPR staff at the UPDB. Samples will only be associated with an anonymous ID generated by the UPDB, an anonymous ID generated by the Office of the Medical Examiner, and by their anonymous CCTS LABID number. Phenotypes will include month/year and zip codes.

External (NIH, PGC): For external the samples/data, only anonymous ids will be used, and phenotype data will include only de-identified data (summary yes/no diagnostic status with no dates, also no zip codes).

When will the data/sample be de-identified?

A limited use data set will be created for the Utah research team after being submitted by the Office of the Medical Examiner or the IH Biorepository and linked with UPDB genealogical, demographic, and co-morbidity data. Before submission to NIH and PGC repositories, all data/samples will be completely de-identified.

5. **Describe the procedures for participants to withdraw their data/samples from the repository. If participants will not be able to withdraw their samples, please provide an explanation:**

Cases cannot be consented for practical reasons, and therefore cannot contact the Utah research team for removal. For external repositories, all samples and data will be completely de-identified from suicide decedents and cannot be removed.

6. **Will future research results or findings be communicated to the participants?**

○ Yes ● **No**

7. **Describe the procedures for other researchers to obtain data/samples from the repository for use in future research.**

DNA and tissue from these subjects is a limited resource. If available, collaborators will be able to obtain it for future research with approval of University of Utah investigators still involved in the research (if any), and ONLY with the approval of the RGE committee, the University of Utah IRB, the Utah Health Department IRB, and a fully executed material transfer agreement.

Use of samples in the NIMH repositories will be through an application/approval process governed by the NIH. Use of additional detailed pedigree structure data not in the repository by investigators requesting repository access will only occur with the explicit approval of the RGE. Use of completely de-identified data for national consortium analysis will first be approved using a process similar to that used by the NIH. For Psychiatric Genomics Consortium collaborative research studies, data access is by application only, and is controlled by the PGC data access committee. Applications for access follow those used by the NIH data repositories. Those using the data must do so only on the secure LISA server (no data are permitted to be downloaded), and access only occurs if
the project is approved by the data access committee and a PGC scientific working group.

IRB_00133374 **Created:** 5/7/2020 11:50 AM

PI: Hilary Coon Ph.D. **Submitted:** 5/19/2020

Title: Identification of subgroups at high risk for suicide using electronic health records data

IRB_00133374
8. Resources and Responsibilities

8. Resources and Responsibilities

1. * **State and justify the qualifications of the study staff:**

Dr. Coon will take the administrative lead to manage aspects of the project, including sample selection, sample management, transfer of samples to the genotyping core, management of genotype data, direction of data flow and analysis of resulting UPDB information, scheduling of regular meetings with collaborators, and writing of grants and manuscripts.

Drs. Amanda Bakian and Dr. Michael Staley will direct aspects of the analysis related to epidemiology and demography. Michael Staley and Nicole Mihalopoulos will study risks to transgender individuals. Dr. Evan Goldstein will study risk to uninsured and underinsured populations as well as other how other risk factors related to social determinants of health.

Drs. Brooks Keeshin, Eric Monson, Taylor Blanding, Rachel Weir and Rachel Jasperson will provide study design assistance and expertise with clinical interpretation of results. Dr. Staley will address non-binary gender risk groups. Brian Mickey will focus on treatment resistant depression and bipolar disorder.

Dr. Scott Langenecker is now at Ohio State; he will serve as an advisor for study design. We are currently determining the feasibility of future collaborative studies. No data will be shared until there is a MTA agreement in place.

Drs. Andrey Shabalin will provide expertise with machine learning methods, and will assist the U of U DSS/SRC. Dr. Warren Pettine will also be developing and deploying ML methods to determing risk subgroups together with fellow Sudhanva Manjnath Athreya.

Danli Chen and Amanda Bakian will assist with data linking and quality control of phenotype data.

Lisa Baird manages the University of Utah Psychiatry Genetics Lab, and will manage all aspects of the study related to DNA samples. Milena Palacios is a student and Theodore Baende is a volunteer affiliate; they will be assisting Ms. Baird in the lab.

Drs. Anna Docherty, Andrey Shabalin, and Emily DiBlasi will help with the polygenic risk analyses of aim 3. Heather Cummins will assist this team in studying polysubstance misuse in suicide and overdose death.

Elliott Ferris will assist the team in studies of whole genome sequence data from suicide deaths and non-suicide death comparison cases.

Zhe (David) Yu and Myke Madsen are data experts with the Utah Population Database, and will assist with data linking and with familial risk analyses using UPDB relative data. Mingyuan Zhange will be providing data science services specifically related to running the NLP algorithm.

Drs. David Crockett (DIrector of the Intermountain Healthcare Biorepository) is PI of the companion parent study for IH. Ken Pena (IH) will be assisting with data linking and analysis of Intermountain Health EHR Data. An approved IH IRB already exists for use of IH Biorepository samples and EHR data (IH IRB #1024977). This IRB is updated continually to add the elements specific to this study (e.g., linkage and processing of physician notes, comparisons between suicide deaths, attempts, opioid risk groups). IH data will not be used until specific IH IRB permissions are granted.

Dr. Vincent Koppermans is a Research Assistant Professor in the department of Psychiatry. He will assist with study design and apply his expertise in the study of neurodegenerative disorders, and how this could be related to suicide.

Dr. Brent Kious has expertise in NLP processing of physician notes to study psychiatric conditions. He will assist with study design, and will help supervise psychiatry/psychology trainees who will be doing manual validation on a subset of records to verify our NLP algorithm

Dr. Strohmeyer assists with validation for the NLP study of suicidality.

Seonggyun Han is a postdoctoral fellow n the Department of Psychiatry. He has training in analysis of genomic data and in machine learning methods that will be applied to this study. We will assist him in devising projects using our de-identified data, allowing him to gain real-world experience with these methods.

Doug Tharp is a Research Associate in the Department of Psychiatry. He will assist with the acquisition of contextual risk factors, assist with data management and analysis.

Dr. Larry Cook-PhD, Dr. Huong Meeks PhD, Sally Jo Zuspan, Russ-Telford and Andrea Thomas are with the University of Utah Data Coordinating Center (DCC). The DCC team will support the study team by for providing an overall assessment of data structure, data validity, data quality.

Emily Sullivan, MPH, will assist with project management and aspects regarding study data (permissions, transfers, storage, quality, analysis).

2. * **Describe the training that study staff and investigators will receive in order to be informed about the protocol and understand their research-related duties and functions:**

Communication will be ongoing by phone, teleconference, and email. Meetings will occur regarding specific grants and analyses in conjuction with the study. These meetings are project-dependent, but happen weekly with relevant subsets of investigators. Drafts of grants and papers are circulated among all study personnel for review

3. * **Describe the facilities where the research activities will be performed (e.g. hospitals, clinics, laboratories, classrooms/schools, offices, tissue banks, etc.).**

Resources for bio-specimen collection, storage, and handling: Blood from decedents is collected from Utah suicides when possible by the Office of the Medical Examiner (OME) personnel. Samples are transported to the University of Utah Center for Clinical and Translational Science (CCTS) Translational Technologies and Resources core facility for
DNA extraction. This core uses the Qiagen Autopure LS automated DNA extractor, a highly reliable and repeatable DNA extraction method. All samples have been stored in -80 freezers housed in the University of Utah Psychiatry Department lab. Freezers are equipped with automatic alarm systems. This lab has a computerized database indexed by anonymous sample ID number. The database can store freezer location, amount, concentration, and quality of each sample. All samples are bar-coded with an anonymous numeric OME ID and another anonymous sample ID assigned by the CCTS system. Suicide research in Utah builds on over 15 years of collaboration on projects involving University of Utah researchers, the Office of the Medical Examiner, government agencies including the Utah Department of Health, and many community-based organizations. The common goal is to learn as much as we can about the science related to death by suicide. We are fortunate to have an Office of the Medical Examiner that handles autopsies for the entire state, and a Medical Examiner (Dr. Erik Christensen) who is dedicated to suicide research. The OME also houses the Utah Medical Examiner Database (UMED); data from this comprehensive database also exists in the Vital Records database. With IRB approval, this project has obtained specific data that allowed linking of suicide decdents into extended pedigrees. Additional data in Vital Records/UMED that will be available for this project includes age at death, race, gender, manner of death, diagnosis at death reported to the OME, medications/drugs reported at death, OME-generated toxicology data, and zip code location at time of death. Data will be indexed by anonymous numeric IDs, and can be easily linked to DNA sample IDs. Utah Population Data Base (UPDB). The UPDB is one of the premier data bases for familial and genetic research. More than 30 years old, the UPDB is one of the world's richest
sources of in-depth information that supports research on genetics, epidemiology, demography, and public health.

Computer: Each investigator and research assistant has a PC with the appropriate word processing and data analysis programs. The University of Utah Psychiatry Department also has two RAID-5 Linux servers with sufficient space and speed for analysis used by Dr. Coon. These servers are behind the University firewall, which also provides VPN ICSA-certified secure network access, network anti-virus scanning, content filtering, and authentication services. Human Genetics and Bioinformatics researchers have access to a large capacity, high-speed computer systems. Working spaces for all investigators have multiple additional computers with large capacity for data storage as part of the University of Utah's Genetic Epidemiology and Psychology Department computer network systems (respectively). Psychiatry, Human Genetics, and Genetic Epidemiology computers all have ample analysis software packages for familial linkage and sequence analysis. All three research groups are highly collaborative, and have active teams for analysis using existing tools and development of new software tools. The master database will be housed with the University of Utah
Center for Clinical and Translational Science Biomedical Research Informatics Core (BRISC). This core provides Information Technology IT resources for clinical and translational research for the University of Utah Health Sciences. BRISC faculty and staff are experienced in all aspects of investigator initiated studies, including multi-center trials. The Research Electronic Data Capture (REDCap) application will be used by this study to enter and capture existing data. REDCap provides HIPAA compliant data collection, storage and data set creation. BRISC servers are located within the University of Utah Center for High Performance Computing (CHPC). The publicly accessible

REDCap web server has an encrypted connection to the REDCap database server which resides within an secure VM server environment. REDCap can also be used to extract data easily for analysis. Access to this protected area is physically and digitally controlled and continuously monitored. Data backups are performed every two hours. Nightly and weekly backups are part of controlled recovery plan which includes offsite storage. The database servers currently support 16 TB of data space.

Office: Investigators have offices within in the U of U Research Park or at the main Medical school campus, in close proximity to each other, and to the University of Utah. All investigators have arranged computer access behind the university firewall so that data and analyses can be easily shared. There is ample office space available at the OME. Standard office equipment is accessible to all personnel.

Laboratory, DNA extraction, storage, quality control, preparation (includes listing of major available equipment): DNA for this study is made at the University of Utah Clinicaland Translational Science Center (CCTS) DNA facility using standard procedures. Also available in nearby equipment rooms are: ultracentrifuges, refrigerated centrifuges, and tissue culture rooms. DNA is housed in the Molecular Genetics Laboratory in the Department of Psychiatry, University of Utah School of Medicine. In addition, DNA quality assurance, Whole Genome Amplification (WGA), and DNA preparation for assays will be done in this laboratory. The laboratory consists of a total of 1400 square feet of modern lab space in two adjoining rooms. The labs are fully stocked with all necessary equipment and supplies to carry out DNA purification, Whole Genome Amplification, DNA quality control and DNA storage. The DNA is stored in three locked -800 C upright freezers, that are monitored 24/7 by an auto alarm system. The lab is also equipped with a tissue culture room containing a Baker SterilGard Biological hood, two Forma C02 incubators and all necessary supplies needed to grow cells for DNA
extraction. Other equipment in the lab consists of, four centrifuges, three balances, four microscopes, one 40 refrigerator, one -200 freezer, a chemical hood, and an autoclave. The lab has unlimited access to a NanoDrop ND-1000 Spectrohotometer to quantify and qualify all DNA samples. All equipment is on a yearly
preventative maintenance schedule to ensure that is always in good working order and calibrated. The lab has two password protected computer work stations that are connected to the University of Utah secure servers, so that all sample data, sample inventories and information are securely stored and backed up routinely.

Support equipment includes a NanoDrop spectrophotometer, and Agilent Bioanalyzer, six BioRad thermal cyclers, a GeneAmp PCR system 9700, a Bio-Rad CFX Connect qPCR instrument, an Agilent Bravo Liquid Handling system, an Illumina cBot, an Illumina MiSeq, two Illumina HiSeq 2000 instruments, and and Illumina Genome Analyzer IIx. A state-of-the-art
database and LIMS system (GNomeEx) has been designed specifically to interface with the core. This system tracks, stores, visualizes, and analyzes experimental information and data from the sequencing experiments.

The NIMH has been a leader in the NIH for data and tissue sample repositories. DNA and cell aliquots from this study will be deposited in the Rutgers University Cell and DNA Repository (RUCDR) managed by the NIMH. This repository houses thousands of samples from multiple NIH-sponsored studies. Researchers apply for access, and are granted through a formal approval process directed by the NIH.

4. * **Describe the medical or psychological resources available at this site (and other participating sites, if applicable) that participants might require as a consequence of the research. If not applicable, please state.**

   NA

IRB_00133374 **Created:** 5/7/2020 11:50 AM

PI: Hilary Coon Ph.D. **Submitted:** 5/19/2020

Title: Identification of subgroups at high risk for suicide using electronic health records data

IRB_00133374
Documents and Attachments

Documents and Attachments

If any of your documents (such as investigational brochures, sponsor protocols, advertisements, etc.) are not available in an electronic format, please scan and save them as PDF files or contact our office for assistance.

Naming Documents: Please use the title field to clearly indicate the content of each form. The name you enter will be listed on your approval letter. Use names that will differentiate from earlier versions.

Examples:
Consent Document Control Group 04/14/05
Consent Document Treatment Group 4/14/05
Sponsor Protocol 04/14/05 Version 2
Assent Document(Highlighted Changes)

Apple/Macintosh Users:MS Word documents must have a .doc file extension. See ERICA home page for instructions.

Print View: IRB Draft Protocol Summary

eProtocol Summary:

Name	Version	Date Created	Date Modified	Date Approved
There are no items to display				

Consent Documents, Consent Cover Letters, Consent Information Sheets, Consent Scripts, etc.:

Name	Version	Date Created	Date Modified	Date Approved
There are no items to display				

Parental Permission Documents:

Name	Version	Date Created	Date Modified	Date Approved
There are no items to display				

Assent Documents:

Name	Version	Date Created	Date Modified	Date Approved
There are no items to display				

VA Consent Documents:

Name	Version	Date Created	Date Modified	Date Approved
There are no items to display				

Surveys, Questionnaires, Interview Scripts, etc.:

Name	Version	Date Created	Date Modified	Date Approved
There are no items to display				

Full Protocol (company protocol, sponsor protocol, investigator-initiated protocol, etc.):

Name	Version	Date Created	Date Modified	Date Approved
There are no items to display				

Investigational Brochure (IB) for Investigational Drug or Drug/Device Package Insert:

Name	Version	Date Created	Date Modified	Date Approved
There are no items to display				

Grant Application:

The Federal Government is a direct or indirect sponsor of your research. You are required to provide a copy of the grant proposal, grant award, or sub-award.

By submitting to the IRB, you are confirming the grant and the study protocol are consistent (Design, Study Population, Study Objectives and Goals, Test Interventions and Procedures, etc.)

Name	Version	Date Created	Date Modified	Date Approved
There are no items to display				

Literature Cited/References:

Name	Version	Date Created	Date Modified	Date Approved
There are no items to display				

Principal Investigator's Scholarly Record (CV/Resume):

Name	Version	Date Created	Date Modified	Date Approved
Coon Biosketch_2023.docx(0.01)	0.01	4/11/2023 2:59 PM	4/11/2023 2:59 PM	

Faculty Sponsor's Scholarly Record (CV/Resume):

Name	Version	Date Created	Date Modified	Date Approved
There are no items to display				

Other Stamped Documents:

Name	Version	Date Created	Date Modified	Date Approved
There are no items to display				

Recruitment Materials, Advertisements, etc.:

Name	Version	Date Created	Date Modified	Date Approved
There are no items to display				

Other Documents:

Name	Version	Date Created	Date Modified	Date Approved
There are no items to display				

IRB_00133374 **Created:** 5/7/2020 11:50 AM

PI: Hilary Coon Ph.D. **Submitted:** 5/19/2020

Title: Identification of subgroups at high risk for suicide using electronic health records data

IRB_00133374
Ancillary Applications

Ancillary Application

This page should be used for submitting human research applications to the following ancillary committees:

Resource for Genetic and Epidemiologic Research (RGE) for access to the Utah Population Database (UPDB) and Utah Cancer Registry (UCR)
Phone: 801-581-6351
Website: https://rge.utah.edu/

Radiological Drug Research Committee Human Use Subcommittee (RDRC-HUS)
Phone: 801-581-6141
HUS Website: https://rso.utah.edu/committees/hus-rsc.php
RDRC Website: https://rso.utah.edu/committees/rdrc.php

Institutional Biosafety Committee (IBC)
Phone: 801-581-6590
Website: https://ibc.utah.edu

ID	Name	Status
RGE_00004731	Identification of subgroups at high risk for suicide using electronic health records data	RGE Approved

IRB_00133374 **Created:** 5/7/2020 11:50 AM **IRB_00133374** Finish

PI: Hilary Coon Ph.D. **Submitted:** 5/19/2020

Title: Identification of subgroups at high risk for suicide using electronic health records data

Finish Instructions

Finish Instructions

1. **To view errors, select the "Validate" option at the top-left of the page. If you have errors on your application, you won't be able to submit it to the IRB.**
2. **Selecting the Finish button will NOT submit the application to the IRB.**
   You MUST select the "Submit" option on the workspace once you've selected the "Finish" button.
3. **If your study has a faculty sponsor: Once the PI submits the application, it will be sent to the faculty sponsor for final approval. The IRB cannot review the study until the faculty sponsor submits the application to the IRB.**