Electronic RESEARCH INTEGRITY AND COMPLIANCE THE UNIVERSITY OF UTAH

Date: Tuesday, May 20, 2025 10:37:47 AM Print Close

IRB_00190712 **Created:** 4/24/2025 4:28 PM

PI: Mark Yandell Ph.D. **Submitted:** 5/7/2025

Title: Personalized Risk Stratification in Atrial Fibrillation using Portable, Explainable Artificial Intelligence

IRB_00190712
1. Contacts and Title

1. Study Introduction

1. **Responsible Investigator:**
Mark Yandell

Email	Training	CoI Date
myandell@genetics.utah.edu	4/3/2023 MCG	2/5/2025

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
Haddy Bah	u1123596@utah.edu	10/23/2023 MCG
Nodira Codell	nodira.codell@genetics.utah.edu	1/25/2024 MCG
Martin Tristani-Firouzi	Martin.Tristani@utah.edu	3/30/2023 MCG

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:**
Personalized Risk Stratification in Atrial Fibrillation using Portable, Explainable Artificial Intelligence

6. **Study Purposes and Objectives:**

Stroke remains the primary source of morbidity and mortality associated with atrial fibrillation (AF), despite ma-jor advances in prevention – direct acting oral anticoagulants and left atrial appendage occlusion.2-4 While ef-fective stroke-prevention strategies are available for patients appropriately identified to be at significant risk,4, 6 optimal implementation of these treatments is limited by (1) rudimentary stroke risk stratification tools, and (2) disparities in

care and outcomes of AF.1, 7-12, 14-20 Strokes still occur among patients with AF who are misclassi-fied as low-risk or fail to receive appropriate therapies due to healthcare disparities.8, 21 Thus, **there remains a critical need for personalized and equitable stroke risk stratification among patients with AF, in order to optimally implement contemporary stroke-prevention therapies.12, 13**

The overarching goal of this proposal is to build on recent R56 support to develop a portable, equitable, per-sonalized risk-stratification tool to improve AF-related stroke prevention, a major NIH priority.12, 13 Our objec-tives are to: (i) discover new stroke risk factors for patients with AF, incorporating social determinants of health (SDoH) with millions of health record covariates, using an innovative comorbidity discovery framework (Pois-son Binomial Comorbidity [PBC]);5, 7 (ii) combine these new risk factors with established ones using machine-learning (ML), in order to determine their network structure and provide explainability; and (iii) develop, deploy, and test a personalized stroke risk stratification tool for AF patients across different health systems in a dispar-ity-aware fashion. Our central hypothesis is that stroke risk stratification can be improved through methods that: leverage all available data, including SDoH; capture and quantify synergies among known and newly dis-covered risk factors in socioeconomic context; and can be ported to other health systems, adapting to different populations. The rationale for this project is that current AF-related stroke risk management lacks the precision and awareness required to optimally implement treatments because it does not adequately account for (1) pop-ulation diversity, (2) SDoH and disparities, (3) synergistic interactions among risk factors, and (4) novel, emerg-ing risk factors. We will attain our overall objectives through the following specific aims:

Aim 1. *Discover* new clinical and socioeconomic relationships that influence stroke risk in patients with AF. *We hypothesize that the PBC approach will identify previously unrecognized factors and relationships associated with AF-related stroke.* Our PBC approach to co-variate discovery7 is specifically designed for big-data; it automatically controls for confounding on a per-patient basis and explores temporal connections, allow-ing us to conduct feature selection among millions of elements. PBC retains the power to identify relationships between factors and outcomes of interest, contrasting with stratified sampling.

Approach: We will deploy PBC on PCORnet datasets augmented with uniquely available SDoH. This will allow for primary discovery, cross validation, and investigations of the role of cross-site variance in the discovery pro-cess. The identified clinical and SDoH variables will be included in ML-based risk calculation machinery.

Aim 2. ***Develop* a socially conscious, ML-based machinery for calculating personalized stroke risk among patients with AF.** *We hypothesize that SDoH interact synergistically with some clinical variables more than others, and that there exist subgroups of patients for whom certain SDoH factors are critical for accurate risk stratification.* Our preliminary results show that many risk factors for stroke are conditionally-dependent upon one another; that their combined effects are not simply additive, and that impacts can vary widely de-pending on various SDoH. These synergies may explain how SDoH drive disparities in care and outcomes.

Approach. Building on recent work,5, 7, 22 we will use Probabilistic Graphical Models (PGMs) to combine known and novel variables (Aim 1) for interpretable personalized stroke risk stratification**.** PGMs are explainable ML tools that capture and quantify synergistic interactions among conditionally dependent variables (not captured by predictive models). They can provide more comprehensive, personalized, and equitable risk estimates.5

Aim 3. *Benchmark* an ML-based stroke risk stratification across a diverse cohort of health systems within PCORnet and discover biases and drivers of downstream care disparities. *We hypothesize that site-specific differences in healthcare environments and patient SDoH impact (a) performance of stroke-risk stratification tools and (b) disparities in care and outcomes.*

Approach: We will use ML-based discovery methods to calculate risk and compare to the clinical standard (CHA2DS2-VASc) within and across sites. Unlike traditional modeling where replication failures are terminal, the explainability of PGMs enables easy identification of variables driving performance, including SDoH. This will be used to improve model transportability and to better understand biases and downstream disparities.

Successful completion of this project will (1) increase our understanding of how biases and care disparities im-pact stroke risk across institutions, and (2) inaugurate a new era in precision medicine for patients with AF, providing truly personalized risk stratification for stroke that is, (a) portable, (b) disparity-aware, and (c) equita-ble. The resulting web-based electronic decision support tool will be poised for deployment in a future, prag-matic prospective trial of emerging stroke prevention strategies in AF.

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

8. **Background and Introduction:**

Atrial fibrillation (AF) is the most common heart rhythm disorder in the United States, affecting nearly 6 million adults. One of the most serious complications of AF is stroke, which occurs when blood flow to the brain is blocked, often leading to long-term disability or death. Strokes caused by AF tend to be more severe than strokes from other causes.

To prevent strokes in people with AF, doctors commonly prescribe blood thinners such as warfarin or newer direct-acting oral anticoagulants (DOACs). In some cases, patients may also be treated with a device that closes off the left atrial appendage, a small pouch in the heart where clots can form. These therapies are highly effective — but only if patients who truly need them are correctly identified.

Unfortunately, the main tool used to estimate stroke risk — the CHA_2DS_2-VASc score — has significant limitations. It was developed from a relatively small and not very diverse patient population, and it oversimplifies stroke risk. For example, it does not fully account for medical conditions like chronic kidney disease or for differences in patients' access to care, insurance coverage, education, or other social factors. As a result, some high-risk patients may not receive needed stroke prevention, while some low-risk patients may be unnecessarily exposed to the risks of anticoagulation.

Recent research, including work by this study's investigators, has shown that social determinants of health (SDoH) — such as income, education, housing, and insurance — significantly influence health outcomes, including stroke risk. However, no current stroke risk calculators used in AF management incorporate SDoH, leaving a major gap in personalized and equitable care.

To address this gap, the current study will use modern machine learning methods and large-scale electronic health record (EHR) data from more than 300,000 patients across the U.S. It will also integrate detailed, individual-level social information from Acxiom, a commercial consumer data source. Together, these data will allow the research team to identify new risk factors, build a more precise and equitable stroke risk prediction tool, and test its performance across multiple health systems. The goal is to develop a tool that is explainable, personalized, portable across different populations, and able to reduce disparities in stroke prevention for patients with AF.

IRB_00190712 **Created:** 4/24/2025 4:28 PM

PI: Mark Yandell Ph.D. **Submitted:** 5/7/2025

Title: Personalized Risk Stratification in Atrial Fibrillation using Portable, Explainable Artificial Intelligence

IRB_00190712
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	Mark Yandell	Yes	

   a. **Select the lead site. Select N/A if there is no lead site.**
   ☐ **N/A**

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	DHHS NATIONAL INSTITUTES OF HEALTH	Federal Government				10154

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_00190712 **Created:** 4/24/2025 4:28 PM **IRB_00190712** View/Edit

PI: Mark Yandell Ph.D. **Submitted:** 5/7/2025

Title: Personalized Risk Stratification in Atrial Fibrillation using Portable, Explainable Artificial Intelligence

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)**
   Mark Yandell

Email	Training	CoI Date
myandell@genetics.utah.edu	4/3/2023 MCG	2/5/2025

   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
Haddy Bah	u1123596@utah.edu	10/23/2023 MCG
Nodira Codell	nodira.codell@genetics.utah.edu	1/25/2024 MCG
Martin Tristani-Firouzi	Martin.Tristani@utah.edu	3/30/2023 MCG

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

Name	Email	Training	Obtaining Consent	CoI Date
Haddy Bah	u1123596@utah.edu	10/23/2023 MCG	☐	5/15/2025
Edgar Hernandez	edgarh@genetics.utah.edu	6/3/2022 MCG	☐	2/27/2025
Jordan King	jordan.king@hsc.utah.edu	1/25/2023 MCG	☐	5/15/2025
Martin Tristani-Firouzi	Martin.Tristani@utah.edu	3/30/2023 MCG	☐	4/22/2025

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 collection

   Data analysis

8. **Select the University of Utah department responsible for this research:**
   PHS-HEALTH SYS INNO & RSRCH

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

IRB_00190712 **Created:** 4/24/2025 4:28 PM

PI: Mark Yandell Ph.D. **Submitted:** 5/7/2025

Title: Personalized Risk Stratification in Atrial Fibrillation using Portable, Explainable Artificial Intelligence

IRB_00190712
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: 10072156
PI: STEINBERG,BENJAMIN ADAM
Sponsor: DHHS NATIONAL INSTITUTES OF HEALTH
Prime Sponsor:
Department:
Short Title: PERSONALIZED STROKE RISK
Sponsor Award Number:
Type: Federal Government
Award Start Date: 9/1/2024
Award End Date: 8/31/2029
Prime Sponsor Type:

IRB_00190712 **Created:** 4/24/2025 4:28 PM

PI: Mark Yandell Ph.D. **Submitted:** 5/7/2025

Title: Personalized Risk Stratification in Atrial Fibrillation using Portable, Explainable Artificial Intelligence

IRB_00190712
3. Participants

3. Participants

1. Ages of Participants:

18 and older (Consent form needed)

2. Specific age range of participants (e.g., 7-12 years old, 60+, etc.):
18+

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: 350,000

5. Characteristics of Participants/Inclusion Criteria:

Participants will be adults (aged 18 years or older) who have been diagnosed with atrial fibrillation (AF) and who have sufficient data available in their electronic health records (EHRs) and associated claims databases to evaluate stroke risk and outcomes.

Inclusion criteria include:

- Age 18 years or older at the time of diagnosis
- At least one ICD-coded diagnosis of atrial fibrillation within a 12-month period, with at least one of those codes documented in an outpatient setting
- Minimum of one year of observation prior to and after the AF diagnosis date to ensure adequate data for evaluating comorbidities and outcomes
- Sufficient EHR and/or claims data to determine baseline clinical characteristics and outcomes
- Non-valvular AF (patients with valvular AF, such as those with mechanical heart valves or mitral stenosis, are excluded)
- Patients from PCORnet-affiliated health systems whose records are accessible through the Common Data Model

(CDM)

6. Participant Exclusion Criteria:

- Individuals under 18 years of age
- Patients diagnosed with valvular atrial fibrillation (i.e., those with mechanical heart valves or mitral stenosis), as their stroke risk profile is significantly different and not compatible with the CHA_2DS_2-VASc risk tool
- Patients without sufficient electronic health record (EHR) or claims data to determine baseline clinical characteristics or outcomes
- Patients without at least one year of observation before and after the AF diagnosis
- Records with missing key demographic variables or outcome information that would prevent inclusion in the statistical models

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

○ Yes ● **No**

IRB_00190712 **Created:** 4/24/2025 4:28 PM

PI: Mark Yandell Ph.D. **Submitted:** 5/7/2025

Title: Personalized Risk Stratification in Atrial Fibrillation using Portable, Explainable Artificial Intelligence

IRB_00190712
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:**
   5

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

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

   Written or electronic record review

   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.):**

   Participants will be identified through electronic health records (EHR) from PCORnet-participating health systems and linked consumer data from Acxiom. No direct participant contact, active recruitment, advertising, or enrollment will occur.

   Authorized research personnel (e.g., data analysts) at each site will run computable phenotype queries using pre-defined inclusion criteria (e.g., atrial fibrillation diagnosis codes). These queries will identify eligible individuals based solely on existing data. The resulting data will be provided in a de-identified or limited dataset format for analysis.

   This study involves only retrospective data analysis and will be conducted under a waiver of informed consent. Participants will not be contacted, and no recruitment letters, MyChart messages, in-person outreach, clinician referrals, or advertisements will be used.

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.**

Participants will be identified through computable phenotype queries run by authorized research personnel (e.g., data analysts or coordinators) at each participating PCORnet site. These queries will search local electronic health records (EHRs) for patients diagnosed with atrial fibrillation (AF). Eligibility will be determined using predefined inclusion and exclusion criteria, such as age (≥18), non-valvular AF, sufficient data availability, and at least one year of observation before and after diagnosis.

For all eligible patients, retrospective data will be extracted from EHRs and administrative claims databases. This includes demographic information, diagnoses, procedures, medications, comorbidities, and outcomes such as stroke. Participants will not be contacted, and all data use will occur under a waiver of informed consent.

The University of Utah, in collaboration with the Greater Plains Collaborative (GPC) Coordinating Center at the University of Missouri, will oversee linkage of Acxiom consumer data with EHR records. Identifiers such as name and address will be used for linkage, then replaced with masked identifiers (e.g., PATID) to create limited datasets. All transfers will follow secure, approved protocols.

The research team will use the linked data to develop a stroke risk prediction model using machine learning techniques, specifically probabilistic graphical models (PGMs). The model will be trained and validated using internal and external benchmarks and compared to the CHA_2DS_2-VASc score to evaluate performance, equity, and clinical utility.

Model performance will also be analyzed across different PCORnet sites to assess its consistency, transportability, and subgroup performance based on demographic and social determinants of health (SDoH) variables.

After data analysis is complete, findings will be shared with collaborators and stakeholders. Study data will be archived in accordance with institutional policies and sponsor requirements. A prototype decision support tool may be created based on the model, but no clinical implementation will occur during the current study period.

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.**

   This study will use a combination of traditional statistical methods and modern machine learning techniques to develop and validate a new stroke risk prediction model for individuals with atrial fibrillation (AF). The study population will include approximately 350,000 patients with AF from multiple PCORnet-affiliated health systems, providing a large and diverse dataset to support statistically robust and generalizable findings.

   Primary analyses will identify and validate both clinical and social risk factors for stroke. These analyses will employ multivariable logistic regression, Cox proportional hazards models, and probabilistic graphical models (PGMs). Data sources will include structured electronic health record (EHR) and claims data, as well as linked social determinants of health (SDoH) data from Acxiom (e.g., income, education, housing, and access to care).

   Machine learning algorithms, including explainable models such as PGMs, will be trained using an 80/20 split-sample approach, with internal cross-validation and external benchmarking across multiple sites. Model performance will be evaluated using standard metrics, including:

   - Area under the receiver operating characteristic curve (AUC)
   - Calibration plots
   - Brier scores

- Decision curve analysis

The new model will be compared to the current clinical standard—the CHA_2DS_2-VASc score—to evaluate improvements in predictive accuracy, generalizability, and fairness across demographic and socioeconomic subgroups.

The inclusion of over 350,000 patients ensures high statistical power, even for relatively rare outcomes like stroke. The large sample allows for reliable subgroup analysis, precise effect estimation, and reduced risk of overfitting in machine learning models. This dataset is sufficiently powered to support complex modeling techniques and robust validation procedures.

IRB_00190712 **Created:** 4/24/2025 4:28 PM

PI: Mark Yandell Ph.D. **Submitted:** 5/7/2025

Title: Personalized Risk Stratification in Atrial Fibrillation using Portable, Explainable Artificial Intelligence

IRB_00190712
- Request for Waiver of Consent

Request for Waiver or Alteration of Consent

* Requested Waivers

	Date Created	Type of Request	Purpose of Waiver Request
View	4/25/2025	Waiver of Informed Consent	The waiver is requested to permit retrospective review of electronic health records and linked claims and consumer data without requiring informed consent. This includes access to social determinants of health data for the purpose of developing and validating a stroke risk prediction model in patients with atrial fibrillation. No direct participant contact will occur, and all data will be used under appropriate data use agreements and security protocols.

IRB_00190712 **Created:** 4/24/2025 4:28 PM

PI: Mark Yandell Ph.D. **Submitted:** 5/7/2025

Title: Personalized Risk Stratification in Atrial Fibrillation using Portable, Explainable Artificial Intelligence

IRB_00190712
IRB Smart Form

Request for Waiver or Alteration of Consent

1. **Purpose of the Waiver Request:**
   The waiver is requested to permit retrospective review of electronic health records and linked claims and consumer data without requiring informed consent. This includes access to social determinants of health data for the purpose of developing and validating a stroke risk prediction model in patients with atrial fibrillation. No direct participant contact will occur, and all data will be used under appropriate data use agreements and security protocols.

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.).**
   The following identifiers may be accessed solely for the purpose of eligibility screening, data linkage, and creation of limited datasets:

   Dates of service and diagnosis (e.g., admission, discharge, procedure dates)

   Medical record numbers (MRNs)

   Patient ZIP codes or geographic location information

   Names, addresses, and dates of birth (used only for linkage with Acxiom consumer data)

   Health insurance or claims identifiers, where applicable

   Internal patient or study IDs used for record matching and de-identification

   Identifiable information will be accessed only as necessary for data linkage and preparation of limited datasets and will not be retained in analysis files. All analyses will be conducted using de-identified or limited datasets in accordance with HIPAA and institutional data security protocols.

4. **Explain why the research could not be practicably conducted without using identifiable information. Examples of such explanation could include the following:**
   Identifiable information is needed in order to identify eligible participant records through electronic health records (EHR) and claims data using computable phenotype algorithms.

   Identifiable information is also necessary to accurately link health records to Acxiom consumer data, which contains detailed social determinants of health not available in clinical systems.

   Identifiers such as names, addresses, dates of birth, medical record numbers, and dates of service are essential for data linkage, validation, and longitudinal tracking of clinical outcomes such as stroke events. The research could not be practicably conducted without access to this identifiable information, as accurate integration and analysis of these data sources depends on it.

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...":

If I had to obtain consent, the research could not be conducted because the study relies on retrospective data from approximately 350,000 individuals with atrial fibrillation across multiple health systems. Many of these individuals are no longer receiving care at the participating institutions, may have relocated, or may be deceased. Re-contacting such a large and geographically dispersed population would require significant time, staffing, and financial resources, making the effort infeasible. Additionally, contact information for a substantial portion of the cohort is likely outdated or unavailable. Attempting to obtain consent would result in substantial selection bias and loss of generalizability, undermining the scientific validity of the study.

6. **Explain why the research and privacy risk of the research are no more than *minimal:***

   The research and associated privacy risks are no more than minimal because the primary risk is a potential breach of confidentiality, and multiple safeguards are in place to reduce this risk. These include secure data storage, restricted access to identifiable information, encryption, password protection, and compliance with HIPAA and institutional data security policies. Identifiable information will be accessed only as necessary for record linkage and eligibility confirmation; all analyses will be conducted using de-identified or limited datasets. No physical procedures, sensitive interventions, or direct participant contact are involved.

7. **Describe the measures you will take to ensure the waiver or alteration will not adversely affect the rights and welfare of the *subjects:***

   The waiver will not adversely affect the rights and welfare of the subjects because the data being used were originally collected for clinical care and the research will not impact the medical care individuals have received or will receive. The study involves secondary analysis of existing information and does not involve any direct contact with participants, interventions, or disclosure of individual results. The data are not highly sensitive, and a reasonable person in the participants' position would not likely consider this use of their data to be a violation of their rights or welfare.

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":***

   N/A

IRB_00190712 **Created:** 4/24/2025 4:28 PM

PI: Mark Yandell Ph.D. **Submitted:** 5/7/2025

Title: Personalized Risk Stratification in Atrial Fibrillation using Portable, Explainable Artificial Intelligence

IRB_00190712
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 collection of information about participants is limited to the amount necessary to achieve the aims of the research, so that no unneeded information is being collected

   Other or additional details (specify):

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

   Participant identifiers will be stored separately from the coded, participant data

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

   Other or additional details (specify):

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

   Confirmation that all appropriate information has been reported to the sponsor, oversight agencies (such as the FDA), and/or IRB

   Other additional details (specify):

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

   Select all that apply:

   Principal Investigator

   Study Coordinator or Research Nurse

   Other or additional details (specify):

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

Study data and documentation monitoring is planned on a quarterly basis, with additional targeted reviews after major project milestones such as completion of data linkage, model development phases, and prior to interim or final analyses.

Monitoring will focus on verifying the accuracy of data extraction and linkage processes, ensuring secure handling of identifiable information, and confirming that analysis datasets meet de-identification standards. Cross-site comparisons and data integrity checks will be coordinated by the University of Utah in collaboration with the Greater Plains Collaborative Coordinating Center. No monitoring of consent documentation is required, as the study is being conducted under a waiver of informed consent and involves no participant enrollment.

IRB_00190712 **Created:** 4/24/2025 4:28 PM

PI: Mark Yandell Ph.D. **Submitted:** 5/7/2025

Title: Personalized Risk Stratification in Atrial Fibrillation using Portable, Explainable Artificial Intelligence

IRB_00190712
6. Risks and Benefits

6. Risks and Benefits

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

   The only foreseeable risk is a potential breach of confidentiality, which will be minimized through strict data security protocols.

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

   There is no direct benefit to participants. Potential benefits to society include improved stroke risk prediction for patients with atrial fibrillation, which may lead to more equitable and effective prevention strategies in the future

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_00190712 **Created:** 4/24/2025 4:28 PM

PI: Mark Yandell Ph.D. **Submitted:** 5/7/2025

Title: Personalized Risk Stratification in Atrial Fibrillation using Portable, Explainable Artificial Intelligence

IRB_00190712
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_00190712 **Created:** 4/24/2025 4:28 PM

PI: Mark Yandell Ph.D. **Submitted:** 5/7/2025

Title: Personalized Risk Stratification in Atrial Fibrillation using Portable, Explainable Artificial Intelligence

IRB_00190712
- Request for Waiver of Authorization

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	4/25/2025	Waiver of Authorization	The waiver is requested to allow access to and review of electronic health records and claims data without obtaining individual authorization from participants. This includes identifying eligible individuals, collecting retrospective clinical and social data, and performing data linkage for the development and validation of a stroke risk prediction model in patients with atrial fibrillation.

IRB_00190712 **Created:** 4/24/2025 4:28 PM

PI: Mark Yandell Ph.D. **Submitted:** 5/7/2025

Title: Personalized Risk Stratification in Atrial Fibrillation using Portable, Explainable Artificial Intelligence

IRB_00190712
IRB Smart Form

Request for Waiver or Alteration of Authorization

1. **Purpose of the Waiver Request:**
   The waiver is requested to allow access to and review of electronic health records and claims data without obtaining individual authorization from participants. This includes identifying eligible individuals, collecting retrospective clinical and social data, and performing data linkage for the development and validation of a stroke risk prediction model in patients with atrial fibrillation.

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).**
   dentifying information that may be collected or linked includes:

   Names (for data linkage only)

   Dates of service and diagnosis (e.g., admission, discharge, procedure dates)

   Medical record numbers (MRNs)

   Date of birth

   ZIP codes and limited geographic information

   Internal study IDs or participant codes used for linkage

   Health insurance or claims identifiers, where applicable

   Addresses (for linkage with Acxiom social determinants of health data)

   Identifiable information will be accessed only as needed for eligibility confirmation and data linkage. All analyses will be conducted using de-identified or limited datasets in accordance with HIPAA and institutional policies.

4. **Explain why the *PHI* to be used or disclosed is the minimum necessary to accomplish the research objectives:**
   The PHI to be used or disclosed is limited to the minimum necessary to achieve the research objectives, including accurate identification of eligible participants, linkage of clinical and social data, and longitudinal outcome tracking. Only information essential for data linkage, eligibility confirmation, and model development—such as names, dates, addresses, and medical record numbers—will be accessed. All analyses will be performed using de-identified or limited datasets, and no unnecessary identifiers will be collected or retained.

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..."*
   Obtaining authorization is not practicable due to the large, retrospective cohort of approximately 350,000 individuals, many of whom could no longer in care, have relocated, or are deceased. Contacting all individuals would require excessive time and resources and would introduce bias, limiting the validity of the research.

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:**
   Identifiers will be stored on secure, password-protected servers in compliance with HIPAA and institutional data security policies. Access will be restricted to authorized study personnel involved in data linkage and management.

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 be destroyed immediately after data linkage and validation are complete. All analyses will be conducted using de-identified or limited datasets, and no direct identifiers will be retained beyond the linkage phase.

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:**
   PHI will be accessed only by authorized study personnel and used solely for the purposes outlined in the approved protocol. It will not be reused or disclosed to any other person or entity. All data handling will comply with HIPAA and institutional confidentiality policies.

IRB_00190712 **Created:** 4/24/2025 4:28 PM

PI: Mark Yandell Ph.D. **Submitted:** 5/7/2025

Title: Personalized Risk Stratification in Atrial Fibrillation using Portable, Explainable Artificial Intelligence

IRB_00190712
- Limited Data Set Agreement

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:** Collaborating PCORnet research sites using limited datasets under approved agreements.

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_00190712 **Created:** 4/24/2025 4:28 PM

PI: Mark Yandell Ph.D. **Submitted:** 5/7/2025

Title: Personalized Risk Stratification in Atrial Fibrillation using Portable, Explainable Artificial Intelligence

IRB_00190712
8. Resources and Responsibilities

8. Resources and Responsibilities

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

   All study staff have completed required CITI training in human subjects protection and good clinical practices. Staff involved in data extraction, linkage, and analysis have relevant training and experience in clinical data management, biostatistics, and health informatics.

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:**

   All research staff will review the study protocol in detail, attend required sponsor or coordinating center trainings, and participate in team meetings to ensure a clear understanding of their roles and responsibilities. Ongoing communication with the study leadership will support protocol adherence and data quality throughout the project.

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

   Research activities will be conducted remotely or in secure office environments by authorized study staff. All data access and analysis will occur using secure, password-protected institutional systems in compliance with University of Utah data security policies.

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.**

   N/A

IRB_00190712 **Created:** 4/24/2025 4:28 PM

PI: Mark Yandell Ph.D. **Submitted:** 5/7/2025

Title: Personalized Risk Stratification in Atrial Fibrillation using Portable, Explainable Artificial Intelligence

IRB_00190712
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
Mark Yandell CV_2025(0.01)	0.01	5/9/2025 3:14 PM	5/9/2025 3:14 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_00190712 **Created:** 4/24/2025 4:28 PM

PI: Mark Yandell Ph.D. **Submitted:** 5/7/2025

Title: Personalized Risk Stratification in Atrial Fibrillation using Portable, Explainable Artificial Intelligence

IRB_00190712 Finish

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.**