SHIELD CVD

Public Benefit and Privacy Panel for Health and Social Care (HSC-PBPP) Application Form

Please note that this PBPP application form is made available as an example for researchers who wish to apply for access to the Scottish Medical Imaging Dataset. Note that researchers will be required to submit their own PBPP application and additional information will need to be provided within individual applications.

Researchers are welcome to use this example application to support the development of their own application, but we would recommend contacting the SHIELD CVD team for support in navigating the approvals process. The team can be contacted at bhfdsc@hdruk.ac.uk.

Provide a clear and concise ***lay*** outline of the proposal (max. 250 words). This will be published on the HSC-PBPP website.
*This is a **stand-alone** lay summary of the whole proposal, from participants to outputs, to **inform the public** of the use of their confidential health data. This should include why this is required and how the outcomes will benefit them, and should be written in clear and concise language that the public will understand. All abbreviations should be explained.*

This project aims to enhance the use of medical imaging for cardiovascular research. Cardiovascular disease, which affects the heart and blood vessels, is a leading cause of illness and death. Imaging tests are widely used to diagnose heart conditions and the Scottish Medical Imaging dataset (SMI) is the first national collection of routine imaging data. However, researchers face challenges accessing this dataset because they lack information about its contents, such as the types and numbers of imaging tests, and how best to use it.

In this project, supported by the British Heart Foundation Data Science Centre, we aim to make SMI more accessible to for researchers. We will use information about what imaging tests have been performed along with other linked electronic health care records to summarise information needed to plan research studies. We will summarise the types of imaging tests (including technical information about the test), cardiovascular disease risk factors, types of cardiovascular disease which are diagnosed, and impact on treatments. We will include imaging tests done to look for cardiovascular disease (e.g. computed tomography scans of the heart) and also imaging tests done for other reasons (e.g. wrist x-rays), because both contain hidden information about cardiovascular disease. We will not use the images (pictures) for this project. We will make our analysis code available for other researchers to replicate what we have done.

This will provide important information about how imaging tests for cardiovascular disease are used in Scotland, and help researchers plan research using the SMI dataset.

Provide the specific aims and objectives of the proposal outlined in this application.
This should be the bullet points of the goals of the proposal.

This project aims to improve the identification, access, and use of imaging data in the Scottish Medical Imaging (SMI) dataset for cardiovascular research which benefits the patients and public. The name of the project is “ScottisH medical Imaging dataset with Evaluation of Linked Data for CardioVascular Disease” with the acronym (SHIELD-CVD).

We will work with the SMI team to deliver the following objectives:

1. *Complete application and governance paperwork to access the SMI dataset and linked healthcare information and make this paperwork available to other researchers.*

   Completing this paperwork has been identified as a significant barrier for researchers wishing to access the SMI dataset. We will make our completed and approved paperwork available to other researchers to help reduce the time that it takes them to access the SMI dataset.

2. *Use imaging DICOM meta-data, imaging reports and linked electronic healthcare records to assess the number and type of cardiovascular imaging tests, available imaging sequences/reconstructions, demographic characteristics and comorbidities of patients with different types of cardiovascular disease in the SMI dataset.*

This will enable other researchers to assess the feasibility and planning of research and grant applications. It will also provide a template for other imaging datasets throughout the country to provide comparative information for researchers. This will help researchers identify the most suitable dataset to answer their research question. We will publish summaries of this information on the BHF Data Science Centre website and summarise this information in an article to be submitted to a peer reviewed journal (with all outputs disclosure controlled). We will work with the SMI team to make this information and research ready dataset available to other researchers through the electronic Data Research and Innovation Service (eDRIS). We will produce researcher generated outputs which can be used to further improve the SMI dataset and can be made available to other researchers, after appropriate additional PBPP applications.

3. *Develop analysis code for data curation which can be re-used by researchers speed up the time taken to prepare data for analysis.*

   Data curation is an important step of any research project, but it can be time consuming, requires knowledge of the datasets involved, and is often repeated by researchers who are performing different studies. Using imaging data provides additional challenges over and above linked healthcare records. By developing and sharing analysis code for imaging data curation we will prevent unnecessary duplication of effort, speed up imaging research and help new researchers get started.

4. *Perform more detailed meta-data exploration for coronary artery disease imaging*

   Coronary artery disease remains one of the leading causes of death in Scotland and there is demand from researchers for access to coronary artery disease imaging. We will therefore explore in detail the DICOM meta-data available for imaging performed to assess coronary artery disease. We will explore how the use of imaging for coronary artery disease has changed over time, in response to various guidelines, across regions, in patients with different demographic characteristics, and its association with medication use, management, and cardiovascular outcomes. This will provide useful information for researchers planning to use this dataset and a model for similar research projects for other cardiovascular conditions.

Provide a description of the envisaged **benefits** of this specific proposal to the public and / or patients.
This section must outline why the proposal and its access to data is necessary, and to demonstrate a clear connection between this work, its expected outcomes and the benefit to patients or the wider public which will result from it. The benefit to patients and public of the use of NHS Scotland data must be clear.

The proposed project aims to leverage the population-wide data held in SMI to accelerate cardiovascular research. This will not have a direct benefit for patients, but there is potential future benefit through the research that could be done because of it.

Potential benefits:

- **Accelerated research:** Improving the information that is available to researchers about the SMI dataset could significantly expedite research into heart, circulatory disease and other disease areas. This might accelerate understanding a number of diseases, potentially leading to new treatments or management strategies.
- **Improved lives of patients in Scotland:** By accelerating research and the utility of this dataset, there is increased potential for researchers identifying mechanisms for

preventing and treating disease across the population. The insights gained could be used to enhance the quality of life for thousands living with a number of diseases in Scotland, their families and carers.

- **Early detection and intervention:** The ability to identify high-risk individuals through analysis of electronic health data is a critical benefit. Early detection could include the identification of markers of cardiovascular disease on imaging tests done for non-cardiovascular reasons. Early detection of cardiovascular disease can allow for mechanisms to slow the progression of disease, improve quality of life, and reduce the burden on caregivers and healthcare systems.
- **Enhanced prediction:** Similar to the management of cardiovascular diseases, accurate prediction and early detection could lead to improved management of risk factors. This proactive approach in healthcare can lead to more effective treatment strategies, tailored to individual needs.
- **Tailored healthcare solutions for Scotland:** By analysing data specific to the Scottish population, the research results will be highly relevant and applicable to the local context, particularly in terms of socioeconomic and geographic (remote/rural) contexts. This specificity ensures that the findings are directly applicable to addressing Scotland's unique healthcare challenges.
- **Reducing the carbon footprint of algorithm development:** We plan to extend the researcher generated output (RGO) work to reduce the data-science carbon footprint. During the process of a project many repeated statistical analyses are performed. By curating the dataset for researchers and sharing analysis code and best practices we will help to reduce the carbon footprint of cardiovascular imaging research.

In summary, the project promises to deliver significant benefits in terms of accelerating research and providing tailored solutions to meet the specific health challenges faced by the Scottish population. Our data-driven approach to addressing major public health issues has the potential for far-reaching positive impacts on public and patient health.

Provide concise details of the proposal: background and reason for requesting data, sample size, inclusion and exclusion criteria, time period; data collection; data processing or other means required to achieve the aims of your proposal. Please justify the use of all the datasets requested.

This should describe why and how you will carry out this work, for the whole proposal, from patient to outcomes. The prompt questions below have been provided for the relevant information required by the reviewers. Please ensure all relevant questions are covered. Please do not include academic literature references in the application form. A separate protocol can be provided as a supporting document.

Please be as clear and concise as possible as this will help the review process. Please use language that will be understood by reviewers who will not have the same background or extensive knowledge of your area of work.

- Why is this proposal needed?

This proposal aims to enhance cardiovascular research by making the Scottish Medical Imaging (SMI) dataset more accessible and useful. The dataset contains valuable imaging data that can improve understanding and treatment of cardiovascular disease but it is underused in current research.

In terms of the specific objectives (these are listed chronologically rather than in terms of importance):

1. *Complete application and governance paperwork to access the SMI dataset and linked healthcare information and make this paperwork available to other researchers.*

Gaining approvals to access routinely collected data is an important but time-consuming task. We (the BHF Data Science Centre) ran a workshop, including a large number of cardiovascular imaging researchers from across the United Kingdom, where challenges in gaining regulatory and ethical approvals was highlighted as an important barrier to cardiovascular imaging research (https://zenodo.org/records/6908183). During the discussions at the workshop the challenges of completing paperwork, such as the PBPP form and associated Project Specification Document, were highlighted as barriers which we could help to address. When new datasets become available there is a learning exercise that must be done by researchers in order to best understand and use the data. The SMI dataset has been used by a small number of researchers to date, and we want to help more researchers to use this for cardiovascular disease research. This involves helping researchers understand what the dataset contains in terms of the images and the information about the images (called DICOM meta-data, DICOM stands for Digital Imaging and Communications in Medicine and is a standard for the saving of medical imaging data https://www.dicomstandard.org) and about the population this represents. DICOM meta-data includes information on how the image was made which can be important for selecting the best tools to analyse the images and what research it is possible to do with these images. Summarising the DICOM meta-data contents and completeness will help researchers to better understand this information so that they can plan their research appropriately. We will make appropriate sections of our application paperwork available to other researchers, including explanations of what the different variables in the Project Specification Document mean. We will do this working alongside the eDRIS coordinators who have helped us with this application. All information from the datasets will be disclosure checked, as is normal practice, and no identifiable information will be shared. This will not replace any of the activities required for PBPP approval and researchers planning future studies will still be required to engage fully in the application process.

2. *Use imaging meta-data, imaging reports and linked electronic healthcare records to assess the number and type of cardiovascular imaging tests, available imaging sequences/reconstructions, demographic characteristics and comorbidities of patients with different types of cardiovascular disease in the SMI dataset.*

We want to describe the characteristics of the imaging tests and patients that are available so that researchers can use this information to plan their research. After discussion with imaging researchers at workshops and webinars organised by the BHF Data Science Centre we would like to structure this in three groups.

Firstly, we would like to describe the imaging tests that have been done to look at the heart and blood vessels (e.g. computed tomography scans of the heart, magnetic resonance scans of the blood vessels). This will include both patients who have had cardiovascular diseases diagnosed based on these imaging tests and those where no disease was diagnosed.

Secondly, we would like to describe all other imaging tests in the SMI dataset, as important information on cardiovascular disease or cardiovascular risk can be obtained from these imaging tests even though they were not done to specifically look at the heart or blood vessels. This includes imaging tests done for non-cardiovascular reasons where the heart and blood vessels are seen (e.g. a computed tomography scan of the abdomen includes information on the aorta and major blood vessels). This also includes imaging tests that do not visualise the heart and blood vessels (e.g. magnetic resonance imaging of the spine, x-ray of the hand, ultrasound scan of the kidneys) because information on the composition of other organs such as the muscles, bones, kidneys etc. can provide important information on the presence of hidden cardiovascular disease and cardiovascular risk. There are numerous

examples where information from an imaging test can unexpectedly be used to identify cardiovascular disease or cardiovascular risk. For example, the size and quality of muscles is associated with the risk of heart attacks (https://pubmed.ncbi.nlm.nih.gov/39827905/), the density of the bones is associated with risk of heart disease (https://pubmed.ncbi.nlm.nih.gov/19819456/) and calcification on mammographs is associated with the presence of heart disease (https://pubmed.ncbi.nlm.nih.gov/30803814/).

Thirdly, we will identify patients with cardiovascular diseases or cardiovascular risk factors who have undergone one or more imaging test in the SMI dataset. We will provide summarised information on the types of imaging tests patients with different types of cardiovascular disease or cardiovascular risk factors have undergone.

For each of these groups we will provide summaries of the imaging, demographic and clinical characteristics as described below. This information will be used by researchers to help plan research projects, including knowing whether the specific type of imaging they require is in the dataset, what number of patients is available of the type they require and helping with power calculations. In workshops and webinars, we (BHF Data Science Centre) have been working with cardiovascular imaging researchers to develop a list of the sort of information that researchers need and the questions that will help them plan their research. This extensive list includes questions such “How many patients with diabetes mellitus have had a coronary computed tomography angiogram including contrast enhanced and non-contrast phases”, “For patients who have undergone cardiac magnetic resonance imaging, how many have myocardial perfusion sequences which could be used for quantitative analysis”, “How many patients with x-rays of the ankle have a diagnosis of peripheral vascular disease and what is their average age”, “How many patients with myocardial infarction have undergone a previous computed tomography scan of their chest and for how many was coronary artery calcification identified in the text report”.

3. *Develop analysis code for data curation which can be re-used by researchers speed up the time taken to prepare data for analysis.*

We will create analysis code (in R and Python) to perform the assessments in Objective 2. We will create analysis code to clarify missing data, overlaps between datasets and inconsistencies between datasets (also called data curation). We will create analysis code to summarise the demographic, clinical and imaging characteristics of the patient groups described in Objective 2. This will be done for the information about the images (the DICOM meta-data which includes text reports), cardiovascular diseases, associated co-morbidities, and cardiovascular risk factors. We will not be analysing the images themselves but this information will facilitate other researchers to perform analysis of the images.

- What is the background, design and methodology of your proposal?

Cardiovascular disease, affecting the heart and blood vessels, remains a leading cause of morbidity and mortality globally with 7.6 million people living with heart and circulatory disease in the United Kingdom, including 730,000 people in Scotland. Medical imaging plays a crucial role in diagnosing and managing cardiovascular disease. This includes imaging throughout the body which assesses the blood vessels both directly and indirectly, imaging tests that show the indirect effects of cardiovascular disease (e.g. the size of the kidneys on ultrasound) and imaging tests that show hidden information on cardiovascular disease and cardiovascular risk (e.g. the size and quality of the muscles on an magnetic resonance imaging scan of the spine, and the density of the bones on an ankle x-ray, both of which are associated with cardiovascular risk). Thousands of patients undergo medical imaging as part of their routine clinical care every

day in the United Kingdom. However, this routinely performed imaging is underutilised in cardiovascular research because of difficulties in identifying it, obtaining access to it, and linking it to other routinely collected healthcare information. The Scottish Medical Imaging (SMI) dataset represents a pioneering effort, as the first national compilation of routinely collected imaging data in the United Kingdom. Managed by Public Health Scotland, it includes comprehensive imaging data from the Scottish National Picture Archiving and Communication System (PACS) covering 57.3 million examinations between 2010 and 2018.

This project aims to enhance the SMI dataset's accessibility and applicability for cardiovascular research. By providing clear information on the contents of the dataset in terms of both imaging parameters and disease representation, streamlining governance paperwork, and developing analysis code, we will facilitate research efforts that ultimately benefit patients and the public.

From the SMI dataset we will request access to the imaging DICOM meta-data. This includes information such as what type of imaging test was done, what part of the body was looked at, what types of images were made, whether contrast was used, and the text report if available. We are not requesting access to the images themselves. This will be used to summarise information on the types of imaging test that are available, including tests done to look at the heart and blood vessels and tests done to look at other parts of the body (where hidden information on cardiovascular disease and cardiovascular risk can be identified). We will summarise information on how the images have been made including the technical parameters which are required by researchers to plan future research.

We request access to the other linked electronic healthcare datasets in order to define cardiovascular diseases and cardiovascular risk factors. Cardiovascular disease is a lifelong condition, so we request access to each dataset from their start date. This will enable us to comprehensively assess cardiovascular disease and cardiovascular risk factors, which may start in early life and initially be unappreciated. We understand that some of the datasets are less compete at their start and will incorporate this important fact into our analysis. Specifically, we will use linked data from other datasets available from Public Health Scotland (SMR00, SMR01, SMR02, SMR06, NRS Deaths, UCD A&E, UCD GP out of hours (GPOOH), Prescribing Information System) to understand the demographic and clinical characteristics of patients. Scottish Index of Multiple Deprivation will be derived from postcodes so that socioeconomic factors can be incorporated. SMR00 will provide information on outpatient attendances including for cardiovascular disease and comorbidities. SMR01 will provide information on inpatient and day case admissions including information on cardiovascular disease, cardiovascular procedures and co-morbidities. SMR06 will provide information on cancer diagnoses which is an important risk factor for cardiovascular disease. Cancer and cardiovascular disease share risk factors, cancer treatments increase the risk of cardiovascular diseases, and patients with cardiovascular disease are at increased risk of certain cancers, for reasons that are still uncertain (https://www.nature.com/articles/s41569-024-01017-x). We therefore request access to detailed information in this dataset including the type and stage of cancer and the treatments a patient receives. SMR02 will provide information on pregnancy associated cardiovascular risk factors. Pregnancy, pregnancy outcomes, abortions, miscarriages, still births, complications during pregnancy, treatments during pregnancy, caesarean section, birthweight and neonatal deaths are all important cardiovascular risk factors (https://www.ahajournals.org/doi/10.1161/CIRCRESAHA.121.319895) which have often been ignored in previous research (as highlighted in the Women's Health Plan https://www.gov.scot/news/womens-health-plan/). We therefore request access to this dataset so that we can summarise this important information for researchers alongside the imaging information. This dataset also contains information on drug use, alcohol use and smoking during

pregnancy. These are important cardiovascular risk factors which we would like to summarise so that they can be taken into account in analyses. NRS Deaths will provide information on all-cause mortality, cardiovascular mortality, and mortality associated with comorbidities. UCD A&E and UCD GPOOH will provide information on cardiovascular disease, cardiovascular risk factors and comorbidities. The Prescribing Information System will provide information on medications used for cardiovascular diseases and medications associated with increased risk of cardiovascular diseases. We therefore request access to all of the medications in the Prescribing Information System rather than just those prescribed for cardiovascular disease. Importantly the Prescribing Information Dataset will contain information on medication use for cardiovascular diseases which do not lead to hospital attendances and using this information will allow us to create the most comprehensive description of the presence of cardiovascular diseases within the SMI dataset.

We request access to these datasets from their start so that we can fully describe information on cardiovascular disease and cardiovascular risk factors which are lifelong conditions. Many cardiovascular diseases start at a young age with subtle changes which may only be identified on imaging and patients accumulate risk factors throughout their life. We know that some of these datasets are incomplete, particularly at their start, and will take this into account in our analyses.

We will summarise the information in three ways as described above: groups of patients who have undergone cardiovascular imaging tests, groups of patients who have undergone other imaging tests, groups of patients with cardiovascular diseases, groups of patients with cardiovascular risk factors (including traditional, “non-traditional” like pregnancy complications and cancer, comorbidities, and medications). We will summarise the numbers, demographic characteristics (e.g. age, sex, socioeconomic deprivation), cardiovascular diseases, cardiovascular risk factors, comorbidities and outcomes for each group. We will summarise the technical imaging information contained in the DICOM meta-data. We will use the text reports in the DICOM meta-data to cross reference diagnoses. We will provide more detailed information on imaging test use for coronary artery disease as it is the leading cardiovascular disease, so that we can demonstrate how the dataset can be used to analyse trends and outcomes and to document limitations and challenges of using the dataset

We plan to produce abstracts and research papers to summarise our results. All outputs will be disclosure checked before release. We will share our statistical analysis code on GitHub to facilitate research. We will create derived data representing “ground truth” diagnoses. By “ground truth” we mean we will use information from across the different datasets to identify patients with one or more cardiovascular diagnosis based on ICD 10 codes. These datasets combined will be used to define the diagnoses, rather than in one dataset alone. This has now been clarified in the application. For example, from the combined datasets we will identify the group of patients who have diabetes mellitus and have had a computed tomography scan of their heart. This information and/or the analysis code used to generate this information will be retained by eDRIS as Researcher Generated Outputs (RGOs) for use by other researchers (after appropriate PBPP and governance paperwork has been completed by those wishing to access it). For example, RGOs could include the type of imaging performed classified based on the combination of information in meta-data and text reports, as no single parameter provides enough information on its own.

- How will the datasets and variables requested be able to answer the questions posed in your proposal?

We will use the DICOM meta-data and structured reports from all imaging tests within the SMI dataset, excluding computed tomography and magnetic resonance imaging of the brain as other researchers are already exploring brain imaging. Medical imaging plays a crucial role in diagnosing and managing cardiovascular disease. This includes imaging throughout the body which assesses the blood vessels both directly and indirectly. The meta-data will provide information on the type of imaging, the body part covered, and the parameters used to acquire the imaging (e.g. type of scanner, use of contrast, sequences or reconstructions). The imaging reports will provide information on the type of imaging performed, the body part covered, and the presence of direct or indirect markers of cardiovascular disease.

We will use other linked datasets to characterise the demographic and clinical characteristics of patients, the presence of cardiovascular disease or cardiovascular procedures, the impact of medication use, the impact of covariates and comorbidities on the diagnosis and outcomes of cardiovascular disease, and the impact of socioeconomic status and urban/rural location.

The CHI database will be used for processing and to obtain demographic characteristics. SMR00 (outpatient attendance) will provide information on diagnosis, confounders, covariates, outcomes, why imaging tests are used and regional variations on imaging test use. SMR01 (general/acute inpatient and day case) will provide information on diagnosis, confounders, covariates, outcomes, why imaging tests are used and regional variations on imaging test use. SMR02 (Maternity inpatient and day case) will be used to assess pregnancy specific risk factors for cardiovascular disease. SMR06 (cancer registry) will be used to assess cancer as a covariate in analysis and the impact of cancer treatment on cardiovascular risk. NRS Deaths will be used to classify diagnoses, confounders, covariates and outcomes. UCD A&E and UCD GPOOH will provide information on diagnosis, confounders, covariates, outcomes, why imaging tests are used and regional variations on imaging test use. The Prescribing Information System will provide information on the impact of imaging on medication use.

- How many individuals will be required for this proposal (approximation)? Why is this number required?

We will use information from all adult patients (18 years or older) who are represented within the SMI dataset (approximately 4.2 million people). We will exclude patients who have only undergone brain computed tomography or brain magnetic resonance imaging, as these modalities are being explored by another research group. This information is required to ensure that we provide comprehensive insights into cardiovascular imaging use which can impact all parts of the body, directly and indirectly.

- What criteria will be used to define your cohort or population of interest?

Inclusion criteria are all adults with at least one imaging test within the SMI dataset, not including computed tomography or ~~MRI~~ magnetic resonance imaging of the brain. Exclusion criteria are patients without a CHI number for data linkage.

- Are there any datasets that will only be used for the cohort creation and or linkage and therefore needs to be identified in the project but won't be released to the researcher?

No.

- Will you contact the individuals for this work?

No contact with individuals will be required.

- Please define and justify the time-period of the data required?

Data will be required from the start of each dataset to provide a comprehensive view of trends and patterns in cardiovascular imaging and cardiovascular health. Cardiovascular disease develops slowly over the entire length of a patient's life, and risk factors which occur early in life may predict the development of cardiovascular disease and cardiovascular events in later life. We therefore request access to the available data from the start of each dataset.

- How will the data be obtained and processed?

Data will be accessed via the National Safe Haven, a secure trusted research environment. The eDRIS team will perform necessary linkage. Researchers will use pseudonymous versions of the dataset for analysis.

- Will you require any data linkage to take place? If so, who will carry out the linkage?

Yes, linkage will be conducted by the eDRIS team to combine imaging data with other healthcare datasets.

- Will you be linking datasets from different sources?

No all datasets are held by Public Health Scotland.

- Do you require matched controls for your subjects?

No.

Provide a clear and concise outline of any statistical methods that will be used in the proposal. Is there a formal statistical plan in place?
This should be a brief and non-technical description of the statistical analysis, for people who may not have a background in statistics.

In our proposal, we will use statistical methods to analyse frequency and trends in cardiovascular imaging data. We will create frequency tables to summarise the imaging, demographic and clinical characteristics of patients grouped by imaging test, cardiovascular disease, and cardiovascular risk factors. We will summarise the use of different imaging tests over time. We will summarise the number of imaging tests of different types, including the technical information from the DICOM meta-data. For example, for patients who have undergone a computed tomography scan of the chest we will summarise (mean and standard deviation, median and interquartile range, number and percentage, as appropriate) age, sex, socioeconomic deprivation, cardiovascular risk factors, cardiovascular diagnoses, comorbidities, number of patients taking cardiovascular and non-cardiovascular medications, cardiovascular outcomes and mortality. We will assess these parameters in different subgroups, including based on age, sex and socioeconomic deprivation. We will assess differences in imaging use across regions and at different time points. We will use regression analyses to explore relationships between imaging types, diagnoses and health outcomes.

Provide a diagram to illustrate the data flow or data linkage process envisaged. *This data flow diagram should show the data sources where the data is accessed and stored at each point in the process from patient to outcomes, and by whom, so that roles and responsibilities are clear for data controller and / or processors and for transfers of data.* *If the data flow diagram is in a supporting document, please state where it can be found.*
A data flow diagram is attached as a supporting document (SD3 - 2025 SHIELD CVD Data Flow Diagram V1.0).
Does the proposal focus on or include information from people who might be considered vulnerable? *Definitions of vulnerable people are given in Table 5 of Appendix A of the Guidance for Applicants.*
Might include but not focus
If vulnerable people are the focus of, or included in, your proposal, please give details.
Vulnerable people are not the focus of this application. However, it is possible that vulnerable people may be included with in the inclusion criteria for this application as cardiovascular disease may affect anyone in the population.
Does the proposal seek access to data that could be considered to be highly sensitive or request other (non-health) special category data in addition to health data? *Under GDPR, all health data is classed as special category data. However, some variables are considered highly sensitive health data. In addition, some commonly requested variables are also special category data but not health data (e.g. ethnicity). Classes of special category data and highly sensitive data are given in section 6 of Appendix A of the Guidance for Applicants.*
Yes
If highly sensitive data or non-health special category data are requested, please give details of the variables and why they are required.
We request information on ethnicity as this can impact cardiovascular disease development and outcomes. We will use information on ethnicity held within the PHS datasets, and acknowledge that this has limitations in accuracy and coverage compared to census data. We request access to highly sensitive information including pregnancy outcomes, abortion and drug and alcohol missuse. We request access to this information as they are important cardiovascular risk factors.
Does the proposal seek to use information <u>exclusively</u> about deceased persons? Please give details. *Please note that while deceased people are not subject to data protection law, they are still subject to the Common Law Duty of Confidentiality and legislation governing access to their health records.*
No
Describe how you have included input from the public / lay representatives / patient groups in the design or any other aspect of your proposal.
Our proposal was developed with input from the BHF Data Science Centre's Patient and Public Involvement and Engagement (PPIE) group. They emphasised the importance of using routinely collected data in research and were surprised at how underutilised this information is. Their insights shaped our focus on making the Scottish Medical Imaging dataset more accessible, ensuring our research aligns with public interest and expectations for improving cardiovascular healthcare outcomes.

How did the public / lay / patient input change your proposal?

The feedback from the BHF Data Science Centre's PPIE group highlighted the value of leveraging routinely collected data, motivating us to prioritise accessibility and usability in our proposal. Their surprise at the underuse of such data reinforced our commitment to developing tools and documentation that facilitate broader research use, ensuring practical outcomes that align with public needs and expectations.

How will you keep these patients and the public informed about the ongoing use of their health data for this application and its outcomes?

We will keep patients and the public informed through regular updates via the BHF Data Science Centre's PPIE group emails and meetings. These updates will cover the progress of the project and results. We will produce lay summaries of all our findings to share on the BHF Data Science Centre website. Additionally, we'll include PPIE members in our workshops to ensure continuous engagement and feedback throughout the project.

Describe any scientific peer review undertaken, with details (e.g. formal external scientific review by a peer organisation or funding body, informal internal review, or review by a third party). If no formal external review has been carried out, please explain why not.

The project has been reviewed by the applicants, by the BHF Data Science Centre and HDRUK. We obtained feedback from researchers about the barriers to accessing data at our workshop on “How can we use imaging data to better understand cardiovascular disease” https://zenodo.org/records/6908183 . Obtaining regulatory and ethical approvals was highlighted as a significant barrier to performing research.

The Information Commissioner’s Office (ICO) recommends that a Data Protection Impact Assessment (DPIA) should be carried out at the beginning of any proposal to assess the privacy risks raised by processing people’s personal and special category (e.g. health) data. It is also good practice.
*The ICO has information and screening questions as to whether a DPIA is legally required here (https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-impact-assessments/. If any of these screening questions are answered with a **Yes**, then a DPIA is mandatory as a **legal requirement** by the ICO and a DPIA **must** be provided, which should be signed off by a suitable senior person.*
Some organisations provide their own screening questions and / or require a DPIA anyway. If your organisation does not sign off DPIAs, please provide evidence that your organisation has seen and accepts the risks associated with this processing of personal data. ***Please read the guidance for 3.1.17.***
Has a Data Protection Impact Assessment (DPIA) been carried out for this proposal and the risks accepted by your organisation?

Yes

If Yes, please provide the DPIA as a supporting document and go to Q 3.1.18.
If No, a DPIA has not been done, have the ICO screening questions been answered and agreed by your organisation?

Choose an item.
If Yes, please provide the screening questions and your reasoning for the answers as a supporting document and go to Q 3.1.18. If neither a DPIA nor the ICO screening questions have been carried out, please justify your reasoning and explain how your proposal has undergone a suitable privacy risk assessment.
Is there *any* commercial aspect or commercial dimension to the proposal or its outcomes? *This could include involvement of a commercial organisation, commercialisation of the product or outcome for which the data is required, commercial access to data, outsourced services provided by a commercial company. This needs to be explained carefully. If the commercial organisation is based outside the European Economic Area (EEA), then special consideration has to be made as GDPR does not allow personal data to be transferred outside the EEA.*
No
If no, please go to section 3.2. If yes, please justify the requirement for the involvement of this commercial aspect, how it is necessary for the success of the proposal and what the company will gain from their involvement in this proposal. *Please read 3.1.18 of the Guidance for Applicants.*
Please list the partners involved in the commercialisation of this application, and particularly those from NHSScotland. How will NHSS directly benefit from such use of NHSS data? *Please provide the formal agreement between these partners so the panel can be assured that suitable arrangements are in place for the commercialisation of outcomes from the use of NHSS data.*
How will the commercialisation of any product or outcome and its associated intellectual property be handled, and by whom? Please give details.

Statutory and Regulatory Context ***Please read section 3.2 of the Guidance for Applicants.***
Does your proposal have a statutory or regulatory justification? i.e. is the proposal responding to a statutory or regulatory instruction, duty or order? *This should relate to **specific** statutory or regulatory obligations that are detailed in specific legislation.*
No
If No, please go to Q 3.2.02 If yes, please give details and citation of the specific statutory or regulatory basis involved.
Will both personal and special category data be processed (either by you or on your behalf) as part of this proposal? *Definitions of personal and special category data are given in section 3.2.of the Guidance for Applicants.*

Both personal and special category data

Please tick which legal basis you will use to process **personal data**, under Article 6(1) of GDPR. The most appropriate and commonly ones used for health and social care data are listed below.
Please indicate the lawful basis under current data protection law for processing personal data. If you are unsure which lawful basis is applicable to your proposal, then you may wish to consult your organisation's Information Governance team or Data Protection officer or lead for advice.
Please read the information on legal bases provided in 3.2.02 of the Guidance for Applicants, including the issues concerning using consent as a legal basis for processing data.

☐ 6(1)(c) processing is necessary for compliance with a legal obligation to which the controller is subject.
Please cite the specific legislation that applies:

☐ 6(1)(e) processing is necessary for the performance of a task carried out in the public interest.

☒ Other: if using another legal basis under article 6(1) please cite specific basis:

6(1)(f) processing is necessary for the purposes of the legitimate interests pursued by the controller

Please tick which legal basis you will use to process **special category data**, under Article 9(2) of GDPR. The most commonly used appropriate bases for health and social care data are listed.
A further condition from the Data Protection Act (DPA) 2018 Schedule 1 Part 1 is also required for some legal bases and must be provided.
Please see the table 5 in Appendix B of the Guidance for Applicants for details, the link below, or get advice from your local data protection team.
https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/special-category-data/what-are-the-conditions-for-processing/

☐ 9(2)(h) processing is necessary for the purposes of preventive or occupational medicine, for the assessment of the working capacity of the employee, medical diagnosis, the provision of health or social care or treatment or the management of health or social care systems and services.

Please cite the appropriate condition from the DPA 2018 Schedule 1 Part 1 Paragraph 2

☐ 9(2)(i) processing is necessary for reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health or ensuring high standards of quality and safety of health care, and of medicinal products or medical devices.

Please cite the appropriate condition from the DPA 2018 Schedule 1 Part 1 Paragraph 3

☒ 9(2)(j) processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1).

Please cite the appropriate condition from the DPA 2018 Schedule 1 Part 1 paragraph 4

a) is necessary for archiving purposes, scientific or historical research purposes or statistical purposes, b) is carried out in accordance with Article 89(1) of the GDPR (as supplemented by section 19), and c) is in the public interest.
If you are using another legal basis under GDPR article 9.2, please cite the specific basis and additional DPIA Schedule 1 part 1 conditions, if required.
☐ Other:
Schedule 1 part 1 condition (if required):
Please specify who will process the personal and / or special category data? e.g. eDRIS, trusted third party (CHILi /NRS), local analysts, you, research team, other (please specify)?
In the creation of the datasets, extracting and pseudonymising will be carried out by eDRIS. Data will be processed by researchers in the course of their analysis.
Are there any existing information sharing agreements or contracts in place which support your proposal? *Please give details and provide as supporting documents* *This would include any contracts or agreements with other parties involved in your proposal, which can inform the panel about the bases for access, sharing and / or transfer of data or information, and reassure of the controls in place to reduce any privacy risks arising from these processes.*
No
Are other regulatory approvals pending or received, from within or outside Scotland? *Please give details and provide as supporting documents.* *This would include approvals from other regulatory bodies e.g. Confidentiality Advisory Group (CAG) of the Health Research Authority (HRA).*
No

Research Ethics Governance *If you answered No to Q 2.4, please go to Q 3.4.* *Please read section 3.3 of the Guidance for Applicants and consult your Research Sponsor.* *All research projects potentially need an ethical review, whether by NHS REC or by another ethics body. It is the responsibility of the applicant and research sponsor to ensure that suitable ethical review has taken place.*	
Has your proposal sought NHS or university research / ethics approval?	Choose an item.
If yes, provide committee details, status of approval (i.e. pending, approved) and reference number, as supporting documents and go to Q 3.4	

If no, is your application covered by the National Safe Haven generic ethical approval? *This only applies for applications that will use the National Safe Haven, if the specific conditions outlined in the pre-submission checklist are met.*
Choose an item.
If no, explain why NHS or university research ethics approval is not sought

Safe Havens *Please read section 3.4 of the Guidance for Applicants.*
Do you intend to access the data requested exclusively through any Scottish Government-accredited safe haven? *The Scottish Safe Havens are listed in Table 3 of Appendix A of the Guidance for Applicants.*
Yes
If yes, please go to Q 3.4.03. If No, please answer this question and then go to section 4. If you are applying to use national data from Public Health Scotland (PHS) or NHS National Services Scotland (NHS NSS) and you do not intend to access these data through the National Safe Haven, please explain why. *If you are not obtaining national data, then that should be stated.*
Is this the National Safe Haven or a regional safe haven? *If you are using the National Safe Haven you do not need to complete sections 5.1 or 5.2.*
National Safe Haven
If you are using a Regional Safe Haven, please specify which one. *If you are using a regional Safe Haven you do not need to complete sections 5.1 or 5.2., **unless** you wish to include NHSCR data. Please see section 3.4 of the guidance.*
If you are applying to use national data from Public Health Scotland (PHS) or NHS National Services Scotland (NHS NSS) and you do not intend to access this through the National Safe Haven, please explain why. *If you are not obtaining national data, then that should be stated.*
How and from what location will you access the safe haven specified above? *E.g. remotely from on a university-provided laptop from a university office.* *E.g. using a safe setting from… (specify location)*
Remotely using HDRUK, University of Edinburgh, or King's College London machines on an approved VPN in accordance with local IT policies. The HDR UK Organisation Policy, provided as a supporting document (SD4), provides detailed information on the IT Security arrangements for HDR UK.

Will the safe haven be accessed by anyone working from home?
Yes
If no, please go to section 4. If yes, please provide your organisation's home working policy and / or outline any mitigation measures in place to ensure that the access to the safe haven will be secure.
The HDR UK Organisation Policy, provided as a supporting document (SD4) outlines the working from home policy for HDR UK.

Section 4: Safe Data, Data Subjects and Methodology

4.2	**All Other Existing Datasets or Sources** *Please use a separate line for each dataset.* *Please read section 4.2 of the Guidance for Applicants.* *Contact should be established as early in the process as possible with NHS Scotland boards / data providers to discuss data provisioning requirements for any of the applicable sources listed below.*
Dataset or source Name	Data Controller (Organisation) **For existing dataset/sources for which the data controller is not an NHSScotland board, please append evidence of the data controllers permission to use the data**
SMI database	**PHS**
SMR00	**PHS**
SMR01	**PHS**
SMR02	**PHS**
SMR06	**PHS**
NRS Death Records	**National Records of Scotland via PHS**
UCD A&E	**PHS**
UCD GPOOH	**PHS**
PIS	**PHS**
CHI	**PHS**

Add rows as required.

How were individuals originally informed of the use of their data? Please ensure that you include an appropriate explanation for each of the data sources which you have listed above. *Please see Guidance for Applicants on the use of privacy notices relevant to each dataset, which should be transparent about how people's data will be used and comply with current data protection legislation.*
It is not practical to inform every individual about the use of their routinely collected healthcare data. In line with the other healthcare data stored in the National Safe Haven by Public Health Scotland and the Scottish Government Charter for Safe Havens. Safe havens: charter - gov.scot Information about how routinely collected data involved in this project are managed by PHS (Your rights - Our privacy notice - Public Health Scotland, https://publichealthscotland.scot/our-privacy-notice/your-rights/) and NRS (Privacy, National Records of Scotland (https://www.nrscotland.gov.uk/privacy/). In addition we have published a privacy notice on our website which can be accessed here: https://bhfdatasciencecentre.org/wp-content/uploads/2025/07/SHIELD-CVD-Privacy-Notice-V1.0-FINAL.pdf
Please explain and justify how the principle of data minimisation has been applied to this application, and what measures have been followed to comply with it?

*Data protection law requires that the use of potentially identifiable data is **minimised** to those variables, people and time-frame which are necessary and sufficient to achieve the stated purpose. This is known as the 'data minimisation' principle.(GDPR Article 5)*
The measures we have taken to comply with the principle of data minimisation are that we have only requested data that we need either for the linkage of data sets or that are essential for our analyses (including being able to adjust for covariates). On the level of individual projects requesting access to our RGOs via PBPP, it is reasonably expected that the research team will further refine the study cohort by excluding people whose data is found to not meet thresholds required for their research.

Section 5: Safe Data Processing and Security

Transfer *Please read section 5.3 of the Guidance for Applicants.*	
Please provide details of the security policies and procedures to ensure that data will be transferred in such a way that it is protected from inappropriate or unauthorised access (e.g. email encryption, secure file transfer protocols SFTP, device encryption, physical controls.) *Please provide details and append supporting documents, referencing appropriate sections.* *This should reflect what is in the data flow diagram for Q 3.1.11 and describe the transfer processes in the data flow from the patient to its final destination, including any intermediary stages.*	
No data transfer outside of the National Save Haven is planned. Data will be accessed by academic researchers within the National Safe Haven only, it will be put there by eDRIS according to their security policies and procedures to ensure that data will be transferred in such a way that it is protected from inappropriate or unauthorised access.	
At what intervals/ trigger points will data transfer take place? E.g. one off transfer, monthly intervals.	
Transfer of data will occur at the start of the project and then at yearly intervals.	
Will any personal (identifiable, pseudonymised or potentially identifiable) data be shared with or transferred to any organisation within or outside of the UK?	No
If no, please go to Q 5.3.04 If yes, please specify the organisation and country of destination, and provide details of the method of transfer, the proposed location and method of storage at the destination, and details of the purpose of the data sharing and how the data will be handled and kept secure.	
Other than initial transfers from source systems, is there any copying of data required within the proposal? If no, please go to section 6 If yes, please give details.	No

Section 6: Safe Outputs and Review

Outputs and Dissemination *Please read section 6.1 of the Guidance for Applicants.*	
What procedures will be used for disclosure control for the outcomes of the proposal? *Please outline or attach the policy that will be used.* *This is to ensure that tables and information from the findings does not include outputs from which any person could potentially be identified, e.g. through small numbers in specific groups.*	
Egress of any research outputs, such as statistical results, predictors or associated analysis code, will involve eDRIS checking all outputs and assessing disclosure risk.	
Will proposal outcomes be published or disseminated beyond those listed in Section 1? *If 'No', please go to Section 6.2* *If Yes, please answer questions below*	Yes
How will outcomes from the proposal be published or disseminated, to what audience and in what format, including to patients and the general public? *Please give details.* *How the outcomes from the use of their health data will be fed back to the patients and public needs to be described, as they do not read scientific literature nor attend conferences.*	
Meetings with the BHF Data Science Centre PPIE group are planned for the duration and end of the project, during which the outcomes of this proposal will be shared with the lay public. It is also our expectation that publications arising from the use of the data may garner public interest when disseminated by the respective research teams via national media and social media platforms, through which patients and the public may be informed of the study findings. We will use HDRUK communication channels to disseminate outputs to patients and the public. We will put lay summaries of all our findings on the BHF Data Science Centre website.	
What steps will be taken to ensure that persons cannot be identified in any outputs? *Please give details.*	
Researchers will not have access to any identifiable data. Any outputs describing the data used in the proposed research will be summary descriptions of large groups making identification of individuals not possible. All data will be accessed within the National Safe Haven and outputs can only be released by the eDRIS Research Coordinator who will apply PHS's Statistical Disclosure Policy to any requested outputs.	
Are there any circumstances where a living or dead individual would be cited? (E.g. where a person consented to their data being used as a case study)? *Please give details.*	
No	
Were any permissions to publish data required or sought (e.g. from data controllers)? *Please provide details*	
No	

Retention and Disposal of Data

Please read section 6.2 of the Guidance for Applicants. *Under data protection law, potentially identifiable, identifiable or pseudonymised data should only be retained for a limited time. Once it is no longer needed it should be fully anonymised or securely destroyed. This is known as the principle of storage limitation (GDPR Article 5).*
Which information / data / records retention policy will you apply to the data obtained and used in this proposal? *Please provide details and append supporting documents, referencing appropriate sections.*
We will work with existing Public Health Scotland derived data projects to define data retention policy.
For how long do you intend to retain identifiable or potentially identifiable data after the conclusion of the proposal (including archive/backup copies)?
All data will be retained for 5 years, with the exception of research generated outputs (RGOs). For RGOs we will work with the existing Public Health Scotland derived data projects to define the appropriate data retention policy.
Who will retain the data and where?
Public Health Scotland will retain the RGOs data within the National Safe Haven for future use for cardiovascular imaging research, such projects would require additional PBPP applications.
What is the purpose for retaining the data for the specified time?
Public Health Scotland will retain the RGOs data within the National Safe Haven for future approved research. This will include research into cardiovascular disease where the RGOs can help with cohort curation or disease phenotyping.
What method of disposal or destruction will be used when this period has expired (including archive and backup copies)?
Upon expiry, all files will be disposed of or destroyed according to standard Public Health Scotland policies.
What evidence will be obtained that destruction has occurred (e.g. IT supplier certificate of destruction)?
Evidence of the disposal of data will be generated according to standard Public Health Scotland policies.

Review *Please read section 6.3 of the Guidance for Applicants.*
Describe how the mechanisms which safeguard data security will be audited and reviewed at regular intervals to ensure their continued efficacy.
Access to the safe haven is logged. Logs are monitored for unusual activity. Each person who accesses the data in the National Safe Haven will have signed the eDRIS User Agreement which details acceptable use and penalties for misuse.

Describe any resource implications to any of the proposed measures for the protection of physical or technical security of information which are unresolved at the time of this application (e.g. encryption of devices is an intention not yet fulfilled, IT training is not yet undertaken etc.)
Describe the breach reporting mechanisms to be invoked in the event of any inappropriate access to data or other information security incident
As per the eDRIS User Agreement (section 2.1-2.5), researchers will inform the National Safe Haven research co-ordinator of any breaches.