

CVD-COVID-UK / COVID-IMPACT Ways of Working

Table of Contents

1	Introduction	3
1.1	How to contact us	4
2	Becoming a consortium member	4
3	Communication.....	4
3.1	Meetings	4
3.2	Collaboration.....	5
4	Patient and Public Involvement and Engagement (PPIE)	5
5	Project proposals, approval process and amendments.....	5
5.1	Proposing a project.....	5
5.2	Funding data access.....	6
5.3	Submitting a project proposal and initial triage.....	6
5.4	Approvals & Oversight Board.....	7
5.5	Reviewing / approving projects.....	7
5.6	Contributing to projects.....	8
5.7	Amendments and proposal versioning	8
6	Project lifecycle	9
6.1	Stage 1: Start-up.....	9
6.1.1	Establishing your project’s presence	9
6.1.2	Forming your project team	11
6.1.3	TRE access and safe researcher/information governance training	11
6.1.4	Confirming the number of planned outputs/subprojects	13
6.1.5	Developing and version controlling protocols / SAPs	13
6.1.6	Creating your phenotypes and analysis code	15
6.2	Stage 2: Analysis.....	15
6.2.1	TRE responsibilities.....	16
6.2.2	TRE best practice.....	16
6.2.3	TRE technical and data wrangler support.....	16
6.2.4	Requesting new R/Python packages	17
6.2.5	TRE outputs / statistical disclosure control	17
6.3	Stage 3: Manuscript/report writing	18
6.3.1	Patient/public input and plain English summary of results	18
6.3.2	Manuscript/report review	19
6.3.3	Communications.....	19
6.4	Stage 4: Sharing.....	20
6.4.1	GitHub.....	20
6.4.2	Website and Gateway.....	20

6.5	Stage 5: Completion.....	21
7	Project planning and status reporting	21
7.1	Project planning.....	21
7.2	Status reporting	21
8	Acknowledgements.....	21
	Appendix A: CVD-COVID-UK / COVID-IMPACT Consortium: Principles for Participation	22
	Appendix B: Joint data controller responsibilities framework for COVID research in the NHS Digital Trusted Research Environment for England	23

1 Introduction

CVD-COVID-UK

CVD-COVID-UK, co-ordinated by the British Heart Foundation (BHF) Data Science Centre (DSC), is one of the [NIHR-BHF Cardiovascular Partnership's National Flagship Projects](#).

CVD-COVID-UK aims to understand the relationship between COVID-19 and cardiovascular diseases such as heart attack, heart failure, stroke, and blood clots in the lungs through analyses of de-identified, pseudonymised, linked, nationally collated healthcare data sources in trusted research environments (TREs) across the four nations of the UK¹. Our ethical and regulatory approvals to access data currently run to June 2023, but we plan to seek an extension.

Linkable datasets include those from primary and secondary care, COVID lab tests and vaccinations, deaths, critical care, prescribing/dispensing and cardiovascular, stroke and vascular surgical audits. A dashboard of datasets provisioned in each TRE can be found on the [programme webpage](#).

We have established a consortium of over 330 members across more than 50 academic institutions and NHS trusts including data scientists, clinicians, data custodians and public contributors, all of whom have signed up to an agreed set of [principles](#) which embrace our ethos of being inclusive and transparent, and always working in a collaborative way.

Expansion to COVID-IMPACT

Building on the success of CVD-COVID-UK, the BHF Data Science Centre has now gained ethical and regulatory approval to broaden the scope of the programme to all COVID-related research ([currently using data in NHS Digital's TRE for England only](#)). This is known as COVID-IMPACT and it will help to support research projects from the wider community, including for the [Data & Connectivity National Core Study](#).

Research questions

CVD-COVID-UK / COVID-IMPACT aims to answer the following broad research questions:

1. What are the effects of prior health conditions, their risk factors and medications on susceptibility to and outcomes from COVID-19 disease?
2. What is the direct impact of COVID-19 disease (and interventions used for its prevention and treatment) on acute complications as well as on medium- and longer-term health conditions?
3. What is the indirect impact of the COVID-19 pandemic and the government and NHS response to it on the presentation, diagnosis, management and outcomes of health conditions?

Further detail can be found on the [programme webpage](#), which provides links to the main protocol, the consortium membership (updated periodically) and the consortium's principles.

¹ Currently in England (NHS Digital's TRE for England), Scotland (Scottish National Safe Haven) and Wales (SAIL Databank), with Northern Ireland (Honest Broker Service) to follow.

1.1 How to contact us

Please email us at bhfdsc@hdruk.ac.uk with any consortium related questions, comments, suggestions or requests and we will respond as soon as possible.

2 Becoming a consortium member

[Contact us](#) if you would like to become a member of the consortium and contribute to one of our active projects, or to lead your own.

You will be asked to sign up to an agreed set of principles ([see Appendix A](#)), which will apply at all times during your membership of the consortium, and to opt into receiving email communication from the centre. This will then enable the BHF Data Science Centre (BHF DSC) team to invite you to consortium meetings, provide you with access to programme folders in Box, add you to our dedicated channel on the HDR UK Slack, and invite you to access our private repositories in our GitHub organisation.

We will hold your personal data on our systems for the purpose of carrying out consortium-related activities only. Your data will not be passed to any third parties, except for those members requiring access to TREs where limited detail will be passed to the national TRE providers for the purpose of setting up user accounts. You can unsubscribe from receiving communications/cease your consortium membership at any time by emailing the centre.

3 Communication

3.1 Meetings

The main consortium meetings are currently as follows (note that these are subject to change and updates will be communicated to all consortium members, where necessary):

- **Consortium Update:** this takes place monthly and comprises:
 - A **consortium-wide programme overview** chaired by Cathie Sudlow, Director of the BHF Data Science Centre, including updates on new and existing projects, and dataset provisioning in the respective TREs. As a consortium member, you'll be invited to this 30-minute update, but we appreciate that you might not always be available to attend.
 - The second half-hour is devoted to the **Methods Group**, chaired by Angela Wood, Professor in Health Data Science, University of Cambridge, with a focus on data management and agreeing a programme-wide, methodological approach to performing analyses across all projects. You can choose to opt out if this isn't your area of interest.
 - Minutes of these meetings are circulated to all consortium members.
- **Projects:** each project lead is responsible for setting up meetings as required, supported by the BHF DSC core team. Further detail on contributing to the projects or leading your own can be found below.

- **BHF Data Science Centre Monthly Webinar:** this takes place monthly (currently on the third Wednesday of each month) for all interested parties. It is an opportunity to present brief updates on the work in the centre across any of our thematic areas including presentations on emerging research outputs from the work of the consortium. These presentations are recorded and hosted on the [BHF Data Science Centre YouTube channel](#).

3.2 Collaboration

We have various Slack channels where our trusted research environment (TRE) users and project leads come together to discuss ways of working within the TRE(s) and how best to perform their respective analyses. Contact us to ask for a list of these channels and to be added.

Our consortium documentation and project-related files are held in shared folders in Box which can be accessed by all consortium members.

4 Patient and Public Involvement and Engagement (PPIE)

The BHF DSC works with patients and the public to ensure transparency, and to build trust in the use of health data for research. We actively seek to record and describe tangible examples of the impact of this involvement in changing, shaping and improving research throughout a project.

When proposing a project, you will be asked to indicate if you have involved patients/public contributors in the development of the proposal and how you plan to meaningfully do this throughout the project. Our team of public contributors will review and provide constructive feedback on these plans as part of the [project approval process](#).

You will also be offered the opportunity to [present your planned outcomes back to our public contributors](#). At this meeting, they can advise on routes for effective dissemination with patient groups and the wider general public, and provide suggestions for engagement activities to raise public awareness. You can also work with the public contributors to produce a plain English summary of the results and any other public facing messages.

If you require help with PPIE when drafting a project proposal or as part of an active project, please contact our PPIE Manager, Samaira Khan (samaira.khan@hdruk.ac.uk). We also encourage you to approach existing PPIE networks or contacts at your host institution, or locally, for support and guidance.

5 Project proposals, approval process and amendments

5.1 Proposing a project

If you would like to lead your own project, the latest [project proposal form](#) can be found in Box. The sections to be completed include a plain English summary, project background, the research questions that will be answered, plans for patient/public contributor involvement, an overview of the methods to be used (a detailed plan isn't required at this stage), names of analysts that will work in the respective TREs and required data sources in each TRE.

Projects may be proposed that will utilise the datasets available in the respective TREs in just one nation (England or Scotland or Wales), any combination of two nations or all three nations,

remembering that non-CVD related proposals may only access data through the BHF Data Science Centre in NHS Digital's TRE for England (under the COVID-IMPACT extension).

Before drafting or submitting a proposal, you should check whether a project has already been approved which either covers your planned area of research or overlaps with it by:

- reviewing the plain English summaries of approved projects on our [webpage](#)
- reading the proposals of interest in their respective [project folders](#)
- checking the list of [proposed projects](#) undergoing review

Your proposal may overlap with and/or can be completed as a sub-project of an existing project without the need for additional approvals, or with just a minor amendment ([see below](#)). Where this is the case, the BHF DSC team will connect you with the project lead or convene a meeting to discuss how to collaborate and/or create a sub-project.

Given the size of our consortium and breadth of expertise amongst our membership across the four nations and multiple institutions, we encourage you to take a **team-science** approach when formulating your proposal and to seek contributions/input from as many relevant members as possible. The BHF DSC team will assist with this.

5.2 Funding data access

The BHF DSC's position on funding the cost of analysts accessing data in the respective TREs (including software licenses, compute, desktop upgrades) is as follows:

- For **CVD-COVID-UK** projects, i.e. those seeking to answer COVID-19 **and** cardiovascular specific questions, the BHF DSC will consider funding data access if the project lead has no other funding in place to contribute towards the cost in full or part.

However, named analysts should:

- have capacity and commit significant time to actively work in the TRE(s);
 - commit to write code and run analyses in an optimally efficient way to avoid excessive compute costs.
- For **COVID-IMPACT** projects, i.e. those seeking to answer non-cardiovascular specific COVID-19 questions, you will need to have funding in place to cover data access costs. Please contact us before submitting your proposal to discuss the expected level of contribution per analyst, per month.

Note: the Data & Connectivity National Core Study may agree to fund data access costs for COVID-IMPACT projects until 31 March 2023 only. This will be approved on a case-by-case basis and the BHF DSC team will apply for funding on your behalf.

- Funding for access to data must be appropriately [acknowledged](#) in any related outputs.

5.3 Submitting a project proposal and initial triage

Once you have drafted your project proposal and considered how data access will be funded, completed forms should be sent to bhfdsc@hdruk.ac.uk

On receipt, your proposal will undergo an initial triage by the BHF DSC team to check that the:

- Project falls within the scope of the approvals in place for CVD-COVID-UK/COVID-IMPACT and is feasible using the available datasets/environments;
- Aims and objectives are clear;
- Project doesn't unnecessarily duplicate existing projects;
- Requested data sources are either currently available in the TRE(s) or will be in the future (including whether there is sufficient coverage within specific datasets);
- Data source requests align with the needs of the project;
- Language used within the plain English summary is clear;
- Adequate plans for PPIE are in place.

The initial triage will result in one of three outcomes:

- **Standalone project:** the proposal will be considered as a new project in its own right. Feedback will be provided, and you may be asked to make amendments to your proposal and submit an updated version.
- **Withdrawal:** the proposal may need to be withdrawn for a variety of reasons including the lack of data availability and/or coverage. Suggestions may be made on adapting the study design or data used.
- **Overlap:** if overlap with an existing project is identified, the BHF DSC team will connect you with the relevant project lead and/or convene a meeting to discuss collaboration, including whether an amendment request to broaden the scope of the existing project is required.

5.4 Approvals & Oversight Board

Following the initial triage, new, standalone project proposals are submitted to the Approvals & Oversight Board for their consideration.

Board membership comprises the BHF Data Science Centre Director, national data custodian representatives, consortium members representing academic institutions party to the joint data controller agreement with NHS Digital or with analysts working in other national TREs (Welsh SAIL Databank or Scottish National Data Safe Haven), as well as public contributors from our panel.

The Board is coordinated by the BHF DSC core team and chaired by the BHF DSC Operations Director. Our Senior Health Data Scientists attend meeting as observers to provide background information and feedback or further information on study design and/or data requested.

The current period from submission of a proposal to outcome notification is - on average - one month and the BHF DSC team is looking at ways to reduce this timeframe.

5.5 Reviewing / approving projects

Project leads attend a meeting with two public contributor Board members and provide a brief (up to 5 minutes) overview of their project (with or without slides). The public contributors then have the opportunity to discuss the proposal in more detail as well as to raise questions and provide constructive feedback on:

- Whether the project is relevant to patients and the public;
- The clarity of the aims and objectives;

- The clarity of the plain English summary;
- Whether adequate PPIE plans are in place.

In parallel, the proposal is shared with the other Board members for review against the [same criteria as per the initial triage](#).

The Board then decides whether each project is approved, approved with conditions, or not approved and notifies the project lead of the outcome.

If approved (with or without conditions), the BHF DSC team lists the project reference and title on our webpage, together with the plain English summary (modified where necessary for clarity).

5.6 Contributing to projects

Approved project proposals will be emailed to all consortium members requesting them to contact the BHF DSC team or project lead to register their interest in contributing to the project(s), either from a scientific, clinical and/or analytical perspective.

New consortium members will also be asked to register their interest in contributing to projects during the onboarding process and will be connected with the relevant project leads.

5.7 Amendments and proposal versioning

You'll need to submit an amendment request to the BHF Data Science Centre for review by the Approvals & Oversight Board where a:

- **substantial** change is made to the scope of your planned analyses beyond what has already been approved. This includes where the scope of the approvals in place needs to be broadened to allow a team proposing a similar, but not completely overlapping, project to collaborate with you.
- **non-substantial** change is made which doesn't change the scope of the approved proposal, e.g. requesting access to additional datasets, change in start/end dates, minor amendments to the wording of the proposal for clarification purposes only etc.

Notify the BHF DSC of your intention to submit an amendment request and you will be emailed your project proposal document with any sections and/or datasets added, amended or removed (as per the standard form) since your project was initially approved or last amended. You should then:

- Update the sections in the proposal document, as necessary, **using tracked changes**;
- Complete the [amendment request form](#) located in Box to briefly describe why you need to make an amendment, to state whether the amendment changes the scope of the original proposal and to highlight which sections of the proposal have been updated;
- Return the updated proposal document **and** completed amendment request form.

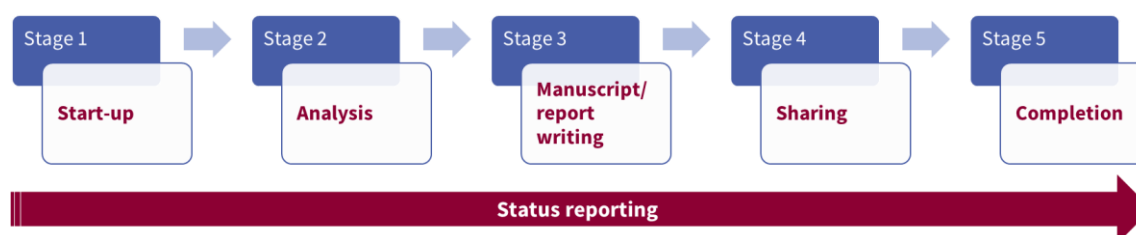
The BHF DSC team will submit a substantial amendment request to the Approvals & Oversight Board for **review and approval** (which may involve a further meeting with the public contributors), whereas non-substantial amendments will be submitted **for information only**. Once approved, or notification has taken place, the BHF DSC team will save a clean version of the proposal in Box and update the webpage, Innovation Gateway and GitHub as necessary.

Project proposals will be numbered/versioned as follows:

Version	Description
0.1, 0.2, 0.X	Submitted proposal and iterations before/during review by the Board
1.0	The version approved by the Board
1.1	Approved version, updated in line with conditions/comments from the Board
2.0, 3.0, X.0	Subsequent amendment requests (version number increases by one)

6 Project lifecycle

Once a project has been approved by our Approvals & Oversight Board, we track the progress of the project through the following stages:



Note that some projects might have multiple outputs/subprojects and/or perform analyses in more than one TRE which run at different rates.

6.1 Stage 1: Start-up

Your project proposal has been approved and it's now time to get your project underway. The Start-up stage is where you:

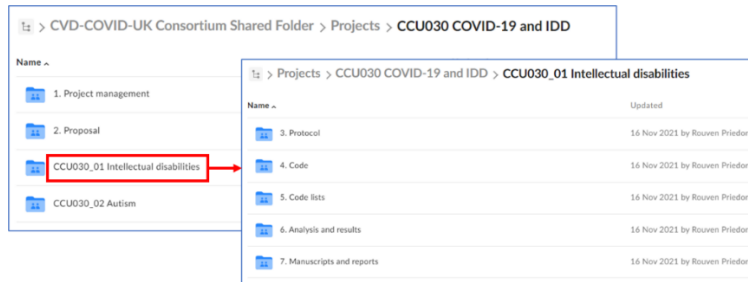
- [establish your project's presence in Box, GitHub, webpages and the Innovation Gateway;](#)
- [form your project team;](#)
- [request TRE access;](#)
- [confirm the number of planned outputs/subprojects and their working titles;](#)
- [develop and version control your protocol/statistical analysis plan \(SAP\);](#)
- [create your phenotypes and analysis code](#), drawing on what already exists.

6.1.1 Establishing your project's presence

The following tasks will be carried out by the BHF DSC team:

6.1.1.1 Box

- A project folder will be created [here](#) in Box, named with the project reference and an abbreviated name, e.g. CCU030 COVID-19 and IDD.
- The project folder will contain a folder to store your approved project proposal (and any subsequent, approved amendments), with additional folders for each planned output/subproject (see below) to store the related protocol and analysis plans, code, phenotype code lists, files relating to your analysis and results, and manuscripts and reports. These folders will be subdivided by nation, where necessary.



- There will also be a separate folder for project management including meeting minutes, project reporting and project completion activities.

6.1.1.2 GitHub

- A repository will be added to the [BHF Data Science Centre GitHub organisation](#) named with the project reference, e.g. CCU030, containing the approved plain English summary and links to repositories for the related subprojects/outputs.
- Additional repositories will be created for each planned subproject/output containing three standard folders, for each nation in which you are performing analyses, to store the [protocol](#), analysis code and [phenotype code lists](#). Repositories will be named with the subproject reference, e.g. CCU030_01.
- Repository visibility will initially be set to **private**, enabling it to be viewed by invited BHF DSC GitHub organisation members only.
- Project leads can elect to change their project repository visibility to **public** immediately, in the knowledge that the full contents of the repository will be accessible to all (including any interim code etc.).
- Alternately, if project leads wish to make just their protocol/SAP public from the outset, these can be placed in the [open-access Protocols](#) repository (see [Developing and version controlling protocols / SAPs](#)).
- In all cases, repository visibility will be changed to **public** in [Stage 4](#) when you submit your manuscript/letter/report to a preprint server and/or for publication.
- Having a repository per subproject/output allows for protocols/code etc. to be made public for projects with multiple planned outputs, but with differing reporting timelines.
- All code in GitHub remains the intellectual property of the contributing authors and is made available under the [Apache version 2.0 license](#).

6.1.1.3 Website

- The project details and plain English summary will be added to the [programme webpage](#)

6.1.1.4 Innovation Gateway and Phenotype Library

- The project will be added to the HDR UK Innovation Gateway and linked to the consortium's [collection](#) (reflected as a data use)
- Phenotyping algorithms will be made available within the BHF Data Science Centre's consortium collection in the [HDR UK CALIBER Phenotype Library](#) (also [see below](#))

6.1.2 Forming your project team

- Everyone working on a consortium project **must be onboarded as a consortium member** by the BHF DSC team and have signed up to the [consortium's principles](#).
- As above, we encourage you to **take a collaborative, team-science approach** in forming your team and to work across institutions with any consortium member who has expressed an interest in contributing to your project. This includes using analyst resource with existing TRE domain/data experience, wherever possible, as this will be the most efficient way of performing your analyses, both from a time and cost perspective.
- The BHF DSC team is also able to assist with providing example job descriptions, advertising open roles or with related funding applications, as appropriate.

6.1.3 TRE access and safe researcher/information governance training

New TRE user access

- The project proposal form specifies analysts needing TRE access per nation. The BHF DSC team will discuss these requirements with the project lead to ensure that TRE access is granted in a manner which both maximises efficiencies and is cost-effective. This can be achieved by:
 - Naming analysts to work in one TRE only for any given project, ideally based on existing domain or dataset experience (an updated list of analysts per TRE/project is maintained by the BHF DSC team).
 - Performing analyses in the respective TREs in parallel, which may include working with analysts from within the consortium who already have access to and are familiar with a TRE and its data, and might be able to assist either on a part-time or full-time basis.
- To be provided with access to data in any of the TREs, named analysts must:
 - Be **accredited**, i.e. they need to provide evidence of completion of an appropriate safe researcher/information governance (IG) training course² within the last 3 years (and will be asked to refresh it as necessary);
 - Have **sufficient analytical capability** to directly contribute to the planned research, with appropriate supervisory and oversight mechanisms in place to support this. **TREs are not to be used as training environments.**
- Once the list of TRE users is agreed for your project, the BHF DSC team will request their access to the relevant TREs, if not already set up:
 - **England (NHS Digital):** the analyst's institution may need to be added as a party to the joint data sharing/controller agreement (if this isn't already the case) or an

² Accepted IG training courses include the MRC's '[Research, GDPR and confidentiality – what you really need to know](#)' course or the ONS Safe Researcher Training course (run by the ONS via the Research Accreditation Service, the UK Data Service, and the Scottish Centre for Administrative Data Research).

honorary contract put in place with an institution that is already party to the agreement³, both of which can take several weeks.

Analysts will need to provide evidence of completion of IG training, their job title, country of residence and mobile number (for two-factor authentication). These details will be shared with NHS Digital who will then invite you to a mandatory TRE induction covering navigation, basic usage and output guidance. If you don't hear from NHSD, email tredatasupport@nhs.net to request an induction session.

Access requests are normally completed within a week of the induction and you'll receive an email with the subject "Trusted Research Environment (TRE) for England – DAE account activation" with details on how to access/activate your account (note: this email may land in your junk mail folder).

- **Scotland (Safe Haven):** analysts will need to provide evidence of completion of IG training (the MRC course) and sign the eDRIS user agreement (this will also need to be countersigned by someone within the analyst's institution). A mobile number is also required for two-factor authentication. Once everything is in place, access is normally granted within a week.
- **Wales (SAIL):** the analyst's institution will need to be added as an organisation with the IGRP to show the list of organisations involved in the project. Analysts will be added to the list of approved users and must then attend an induction session (arranged and led by Ashley Akbari) and provide evidence of completion of IG training⁴ before their access to SAIL and related pages in Confluence are set up. The SAIL new user application process will be completed and the user agreement will need to be signed and countersigned by someone within the analyst's institution.

Removing TRE user access

- Following discussion with users, as appropriate, access may be revoked either on a temporary or permanent basis where analysts no longer need access, have been inactive for over a month and/or are no longer contributing to the analyses for any projects.
- Project leads must **immediately** notify the centre if, for whatever reason, one or more of the named analysts working on their project no longer needs to access data in the relevant TRE(s). This ensures that data access and the related costs are appropriately controlled.

³ Imperial College London, King's College London, Swansea University, University College London, University of Bristol, University of Cambridge, University of Glasgow, University of Leicester, University of Liverpool, University of Manchester, University of Oxford, University of Southampton

⁴ Analysts requiring access to the Census dataset in SAIL must have completed the ONS Safe Researcher Training course and be able to provide their AR (Accredited Researcher) number

6.1.4 Confirming the number of planned outputs/subprojects

Once your project has been approved, you will be asked to confirm the number of planned outputs/subprojects and their working titles. It is accepted that these details may change over time, but they enable the BHF DSC team to structure the folders in Box and set up repositories in GitHub, as appropriate.

6.1.5 Developing and version controlling protocols / SAPs

6.1.5.1 Developing protocols/SAPs

A protocol/SAP should be developed for each planned output/subproject. This provides clarity to the reader and also removes the risk of methods for other planned outputs being made public whilst still undergoing analyses (unless this is agreed – [see above](#)).

As a guide, your protocol/SAP should contain the following sections:

- Version control
- Title (this may differ to the project title if there is more than one planned output)
- Lay summary and/or background
- Research hypothesis, aims and questions
- List of datasets and reasons required
- Study design with detailed definitions of study population, exposures, covariates and outcomes
- Statistical methods
- References

You should liaise with the following people, as necessary, to understand what curated data sources are available in the respective TREs as well as any known issues with data coverage, quality and missingness, as this will determine which of your desired analyses you are able to perform (plans may need to be adjusted accordingly):

- BHF DSC Senior Health Data Scientists:
 - John Nolan: john.nolan@hdruk.ac.uk
 - Thomas Bolton: thomas.bolton@hdruk.ac.uk
- NHSD/TRE for England data wrangler/science team:
 - tredatasupport@nhs.net
- SAIL Databank data wrangler/science team:
 - Ashley Akbari: a.akbari@swansea.ac.uk
 - Fatemeh Torabi: fatemeh.torabi@swansea.ac.uk
 - Hoda Abbasizanjani: hoda.abbasizanjani@swansea.ac.uk
- Methods Group:
 - Lead - Prof Angela Wood: amw79@medschl.cam.ac.uk

If your project proposal states that you will be accessing data from more than one nation, this should be reflected and taken into consideration when developing your protocol(s) and SAP(s).

Where possible, we encourage you to:

- Perform analyses in the relevant TRE environments concurrently, working with others as appropriate.
- Develop your SAP, and data/variables to be used, to enable outputs and results to be comparable across the nations or to articulate what the differences in approach are, the reasons why and what impact this may have on being able to compare/pool the results.
- Publish results for all nations included as part of the same report/letter/paper.

Once you have drafted your SAP(s) and defined your output tables, the BHF DSC team will assist you with sharing any documents with the wider consortium for review and feedback and to determine whether other consortium members may like to actively contribute to the (sub)project. Please allow a minimum of one week for this review cycle to take place.

6.1.5.2 Version controlling protocols/SAPs

Version controlling protocols/SAPs is a simple, but vital, step that must take place **before commencing** any analyses (regardless of whether or not you elect to make your protocol publicly accessible before results are published). You should:

- Add a 'version control' section at the start or end of your document where you can track the version number, date, who made the changes and what the changes are (see below).
- Save the protocol and tables, incorporating any feedback from consortium members, in Box with the filename format **CCUXXX_XX_protocol_vX.X_YYMMDD**
- Upload a pdf version of the file to your project's GitHub repository and/or to the open-access protocols GitHub repository (where the document will be time and date stamped).

It is accepted that protocols/SAPs are likely to evolve whilst performing your analyses. Where this proves to be the case, you should carry out the following steps **before continuing** your analyses:

- Create a new, updated version of the relevant document in Box with the filename format **CCUXXX_XX_protocol_vX.X_YYMMDD**
- Update the version control table within the document to summarise the changes made.
- Upload a pdf version of the file to your project's GitHub repository and/or to the open-access protocols GitHub repository (where the document will be time and date stamped).

Example version control table

Version	Date	Author(s)	Comment
0.1	02/07/2020	Will Whiteley	Initial draft.
0.2	06/08/2020	Will Whiteley	Incorporated feedback from consortium members in sections XXX.
1.0	17/08/2020	Will Whiteley	Approved version (uploaded to GitHub).
2.0	14/11/2020	Will Whiteley	Revised version reflecting changes in variables used in section XXX (uploaded to GitHub).

6.1.6 Creating your phenotypes and analysis code

6.1.6.1 Phenotype code lists

The process of developing phenotypes can be found [here](#).

Many phenotypes have already been curated for the programme, details of which can be found [here](#) or [here](#).

These phenotype code lists have been:

- Uploaded to the NHSD TRE for England in the **bhf_cvd_covid_uk_byod** database
- Imported into the SAIL Databank
- Indexed in the HDR UK Innovation Gateway
- Indexed in the [HDR UK Phenotype Library](#)

If additional algorithms are needed, the requirements should be sent to bhfdsc_hds@hdruk.ac.uk to curate.

Once you have completed your analyses and **before** you publish a manuscript as a preprint and/or submit it to a journal, the phenotyping team will assist with submitting the final versions of the algorithms (codes, metadata and logic) used for your project [to your project's GitHub repository](#) and coordinate submission to the HDR UK Phenotype Library for further dissemination. Information required for each phenotype definition can be found [here](#).

6.1.6.2 Analysis code

See the section below for analysis (and curation) code best practice.

6.2 Stage 2: Analysis

You've formed your project team, got access to the relevant TREs, specified your SAP(s) and defined your code lists and phenotypes. You can now start the analysis stage of your project.

You should always:

- Speak with the Senior Health Data Scientists to identify, re-use or modify research-ready code, if available (Note: remember to acknowledge this in any outputs and to take a copy of any code existing **before** modifying – you should never modify or delete existing code in other projects' folders);
- Make your code available (share in the collaborative space in the TRE);
- Ensure that your code is fully annotated;
- Advise other analysts (via Slack) if you identify a bug in shared code;
- Have your code reviewed, ideally by an analyst not directly involved in your project;
- Arrange for code list to be added to/updated in the phenotype library.

If, for whatever reason, you need to alter the methods used in your analyses, you must reflect any changes in an updated version of your protocol/SAP and upload the new version to GitHub **before continuing** your analyses ([see above](#)). Please also refer to the [amendment process](#) for substantial and non-substantial changes to the scope of your project.

6.2.1 TRE responsibilities

- In line with our [principles](#), analysts are responsible for only conducting analyses that comply with legal, ethical and regulatory approvals and that fall within the remit of the project(s) they are working on.
- Partner institutions vouch for the analysts they employ to enable access to data within the TRE(s). For access to the NHS Digital TRE, this duty is specified within the Data Sharing Agreement that participating institutions sign as joint data controllers (also refer to the joint data controller responsibilities framework in [Appendix B](#)). For access to the Safe Haven, this is specified in sections 2.6 and 2.7 of the eDRIS User Agreement signed by both the analyst and a representative from their institution. For SAIL this is specified in the SAIL Data Access Agreement signed by both the analyst and a representative from their institution.
- Approved projects will have an assigned lead responsible for overseeing the analyses for their project in the TRE(s). The Methods workstream is responsible for coordinating the data management, processing, checking and methodological research carried out to benefit all projects.

6.2.2 TRE best practice

Best practice guidance for working in the respective TREs, including folder/table naming and data management, can be found in the locations below:

- England (NHS Digital): in Box [here](#)
- Scotland (Safe Haven): in Box [here](#)
- Wales (SAIL): see the SAIL Wiki (via desktop link in SAIL environment)

This guidance is a work in progress and will continue to be improved and updated over time.

6.2.3 TRE technical and data wrangler support

If you experience any technical issues with the respective TREs when performing your analyses, you should collate as much detail about the issue as possible, including screenshots, and then:

- **England (NHS Digital):** raise a ticket with the National Service Desk (NSD) by emailing ssd.nationalservicedesk@nhs.net. Please also post a message on the **bhf-dsc-cvd-covid-uk-analysis** Slack channel to see if any other analysts have the same issue or know of any interim workarounds.

Suggested detail to include in the body of the email to assist the NSD in directing your ticket to the right group is as follows:

System: DAE

Tool: databricks (or R-studio)

Cluster: dars_nic_391419_j3w9t_cluster1 (or none if you're not using one in R-studio)

Agreement: dars_nic_391419_j3w9t_v0.2

Action: e.g. trying to query the

dars_nic_391419_j3w9t.primary_care_meds_dars_nic_391419_j3w9t table

Queries on datasets and curation should be emailed to tredatasupport@nhs.net for a response by one of the members of the NHS Digital data wrangler team.

The NHSD data wrangler team also runs **drop-in clinics** on Teams every Monday, Wednesday, and Friday from 10-11am to answer any queries: [link to the data wrangler clinic](#).

- **Scotland (Safe Haven):** email details of the issue, either technical or data, to our Research Coordinator, Diane Rennie, at diane.rennie2@phs.scot who will then liaise with EPCC and eDRIS as necessary. Please also post a message on the **cvd-covid-uk_scotland** Slack channel to see if any other analysts working in the Safe Haven have the same issue or know of any interim workarounds.
- **Wales (SAIL):** the preferred option for technical or data issues is to raise a ticket for the Helpdesk (<https://jira.hiru.swan.ac.uk/servicedesk/customer/portal/6>) to respond to (note that a helpdesk group is set up for our consortium so that tickets can be viewed by other users to avoid multiple tickets for the same issue being raised). Alternatively, you can email helpdesk@chi.swan.ac.uk. You can also post a message on the **bhf-dsc-tre_discussions-wales** Slack channel for Ashley and the SAIL team to respond to.

In all cases, posting a message on the relevant Slack channel will alert other analysts of the issue (who should follow the process above if it is also impacting them).

Our senior health data scientists, john.nolan@hdruk.ac.uk and thomas.bolton@hdruk.ac.uk, are also available to offer support and guidance across all TREs.

6.2.4 Requesting new R/Python packages

- **England (NHS Digital):** email the request to ssd.nationalservicedesk@nhs.net putting “DAE TOOLS REQUEST: ” at the start of the Subject of your email (requests take 2-10 days to be processed for R, often longer for Python).
- **Scotland (Safe Haven):** R is connected to a mirror version of the CRAN, so users can install any available packages. Issues should be raised with our Research Coordinator, as above.
- **Wales (SAIL):** users can install any R package if it is on CRAN. Other packages not on CRAN may be imported into SAIL by users themselves, and installed by users themselves, but users must ensure that any files coming into SAIL are from a trusted source and do not breach terms laid out in the SAIL Data Access Agreement.

6.2.5 TRE outputs / statistical disclosure control

As a general rule, outputs requested from the TREs should:

- Not contain any personal identifiable information or be requested in such a way to enable re-identification (e.g. differencing from previously released outputs);
- Be made at an aggregate/summary level, i.e. not person level;
- Avoid low counts, i.e. <5 and in some cases <10 (e.g. when the output contains sensitive information);
- Be clearly presented (e.g. in the case of code, annotated with comments), so that the output checkers can accurately review;
- Be submitted to allow the output checkers sufficient time to review and assess outputs

Please refer to the TRE-specific guidance below for more detailed information on outputting tables, charts, and code.

Note: interim/draft outputs can be shared with members of the project team or wider consortium only. They must not be onward shared or used in publications.

Outputs can be requested in the respective TREs as follows:

- **England (NHS Digital):** Please see the “Exporting and downloading results” sections of the [Databricks](#) and [RStudio](#) User Guidance documents.
- **Scotland (Safe Haven):** Please see the “Researcher guide: requesting outputs from Safe Haven and disclosure control” document which contains a link to the National Safe Haven Output Request Form, both saved in Box [here](#).
- **Wales (SAIL):** Outputs are requested via the “Data Transfer In-Out” link on the desktop of the SAIL environment. If approved, the user will receive an email with a link to the SAIL portal where the approved output can be downloaded.

6.3 Stage 3: Manuscript/report writing

You’re now ready to write your manuscript and publish your results. Note that:

- Manuscripts must be published with **open access**.
- Publications must include ‘**on behalf of the CVD-COVID-UK/COVID-IMPACT Consortium**’ at the end of the author list, with all relevant individual contributions (coordination, writing, analysis, interpretation etc.) listed and the coordinating role of the BHF DSC acknowledged. Please see the [Acknowledgements](#) section below.

6.3.1 Patient/public input and plain English summary of results

- The results of your research should be presented to and discussed with the public contributors who have been involved in your project from existing PPIE networks or contacts at your host institution, where applicable. You will also be offered the opportunity to meet with the public contributor representatives from our Approvals & Oversight Board. This will help to shape your paper, by highlighting the outcomes of interest to the patients/public, as well as ensuring that the outcomes are communicated clearly and effectively, and that plans to do so have been properly considered.
- If you choose to meet with our public contributor representatives, the BHF DSC PPIE Manager, Samaira Khan (samaira.khan@hdruk.ac.uk), together with the core team, will convene a 30-minute meeting on your behalf, where you will be asked to provide a brief (up to 5-minutes) overview of your outcomes (with a short slide deck), followed by a discussion.
- Public contributors can also assist with preparing/reviewing a plain English summary of results for each output, which will inform communications for your paper at the point of journal publication, such as web stories, social media posts or press releases.
- In the first instance, project leads should complete the [Research Outputs Reporting form](#), which contains sections for a plain English summary of results and communications contacts at your host institution, and email it to both the BHF DSC Communications Manager, Kate McAllister (kate.mcallister@hdruk.ac.uk), and PPIE Manager, Samaira Khan (samaira.khan@hdruk.ac.uk).

6.3.2 Manuscript/report review

- Wherever possible, please allow **at least 2 weeks** for all authors and the BHF DSC Director, Project Manager and/or Senior Health Data Science team to perform their review of the draft manuscript prior to publishing as a preprint and/or submission to a journal. To mitigate issues with meeting any desired timelines, you should notify the relevant stakeholders of when you plan to issue your manuscript to them for review, to confirm their availability and to enable them to free up time, as necessary.
- Prior to submission, the manuscript should be circulated to all consortium members for information only and/or for comment. Please allow a minimum of 1 week for this review cycle to take place.

6.3.3 Communications

- The BHF DSC Communications Manager, Kate McAllister (kate.mcallister@hdruk.ac.uk), and PPIE Manager, Samaira Khan (samaira.khan@hdruk.ac.uk) should be given **a minimum of 5 working days' notice** of any expected preprint/publication of results. This allows them the time to plan for and provide support with the promotion of your paper, to link with partners, including your home institution and funder, and to prepare any communications, if appropriate. You should also supply the contact details of any additional communications stakeholders who may need to be informed, e.g. your institution's press office, if you haven't already done so via the Research Outputs Reporting form. As the major funder of the BHF DSC, the BHF Research Media team will also be notified of any upcoming papers and sent a link to the preprint.
- The BHF DSC will not publicise work as a web story or press release that has not been peer-reviewed, unless in extraordinary circumstances where the findings have an immediate implication for public health. Researchers are particularly encouraged to get in touch with our Communications Manager and press offices at their own institute as soon as possible where findings require sensitive messaging or are of potentially high interest to the public. This allows for careful plans to be made across communications teams and the potential involvement of the Science Media Centre.
- For any outputs with funding from the TTS Consortium, led by Prof Munir Pirmohamed (Understanding Mechanisms of Thrombosis and Thrombocytopenia in COVID-19 and with SARS-CoV-2 Vaccines), NIHR ask that any outputs associated with NIHR awards, particularly press releases, media exclusive journalist briefings and accepted journal articles are provided to Alicia Ptaszynska-Neophytou (alicia.ptaszynska-neophytou@mhra.gov.uk) **at least 3 days before** journalistic outreach or publication. This is not for approval but for raising awareness to the NIHR and DHSC Comms teams. Further detail on the NIHR outputs and publication policy can be found [here](#).

6.4 Stage 4: Sharing

In accordance with our [principles](#), analysis plans, protocols, code (analysis and curation), phenotype code lists and manuscripts arising from your project must be made **publicly available** via the BHF DSC website, Innovation Gateway, BHF DSC GitHub organisation and through open-access publications.

This can be performed in parallel with writing your manuscript/report, however, it should be completed **before** a manuscript is published as a preprint and/or submitted to a journal.

6.4.1 GitHub

The following material should be uploaded to the GitHub repository for your project/subproject (created as per [section 6.1.1.2](#)):

- The project protocol document.
- All code (curation and analysis) used to generate analytical outputs, visualisations and results included in the manuscript. Code should be annotated to provide clear description of what each script is doing.
- All phenotype code lists uploaded as csv files ([see section 6.1.6.1](#)) which, as a guide, should contain the following fields:
 - Phenotype name (e.g. Hypertension)
 - Code (e.g. I10)
 - Code term (e.g. Essential (primary) hypertension)
 - Code type (where 1 is incident codes, and 0 indicates codes for historic/prevalent events)
 - Code position (i.e. primary, secondary, or any diagnostic code position in hospital admissions)
 - Code terminology (e.g. ICD-10, Read V2, SNOMED-CT,...)
- A ReadMe including a description of the project.

Please see [CCU002_01](#) as an exemplar GitHub repository, and feel free to contact the BHF Data Science Centre health data science team at bhfdsc_hds@hdruk.ac.uk with any questions on constructing your code-lists for GitHub.

At the completion of analysis and prior to publication it is expected that all phenotype definitions (including codelists) be submitted to the [HDR UK Phenotype Library](#) ([see section 6.1.6.1](#)). We suggest that researchers familiarise themselves with the [information that will be required](#) at that point and keep the appropriate records.

6.4.2 Website and Gateway

The BHF DSC team will add:

- The output title and links to the preprint and GitHub repository below the project's plain English summary on the [programme webpage](#) (a link to the published paper or letter to editor will also be added, once available)

- Details of the output to the downloadable research outputs summary document on the [programme webpage](#)
- An entry reflecting the preprint/paper to the Innovation Gateway and link it to the BHF DSC [collection](#), [TRE dataset](#) and the related [project/data use](#).

6.5 Stage 5: Completion

Within one month of completing Stage 4 for all planned outputs/subprojects of your approved project, you should submit an end of project report.

The BHF DSC team will convene a lessons-learned meeting with the project team and relevant stakeholders.

7 Project planning and status reporting

7.1 Project planning

Once your project has been approved, for each subproject you will be asked to provide estimates on when you expect to complete each stage of the project lifecycle. This will assist the BHF DSC team with managing the overall programme.

7.2 Status reporting

The BHF DSC team will meet with or email project/subproject leads, or another designated member of the project team, every 1-2 months to:

- Obtain an update on the status of each subproject;
- Understand if any timelines have changed;
- Highlight any issues/blockers that you need assistance from the BHF DSC team to mitigate.

You may also be asked to provide verbal updates on the status of your projects/subprojects at the monthly consortium meeting, as appropriate.

For projects accessing data in SAIL, the status of outputs will need to be updated on your project's page in [SAIL Confluence](#) at least once a month (as this goes back to SAIL, TAG and SAGE via agreed routes in Wales). Please complete these updates in a timely manner, on request.

8 Acknowledgements

To appropriately acknowledge our funders and recognise the use of patient data in the respective TREs in our research outputs and publications, please refer to the [acknowledgement guidance](#) document in Box and include the relevant statements.

The BHF DSC team will check that the appropriate statements have been included during their [review of the draft manuscript/report](#).

Appendix A: CVD-COVID-UK / COVID-IMPACT Consortium: Principles for Participation

- CVD-COVID-UK / COVID-IMPACT commits to the ‘Five Safes’ (<http://www.fivesafes.org/>) and to a transparent and inclusive approach, enabling additional researchers, research groups and, where appropriate, institutions, to join and contribute to the consortium as the work progresses and evolves.
- All project analysis plans and protocols, listing data requirements, will go through the CVD-COVID-UK / COVID-IMPACT approvals mechanism process before work commences.
- All analysis plans, protocols, code, phenotype code lists and reports arising from this consortium’s work will be made publicly available via the HDR UK website and/or HDR UK Gateway (linking to additional institutional documentation if appropriate), repositories in the BHF Data Science Centre’s GitHub organisation and through open-access publications.
- All reports for government advisory groups and policy makers, the lay public and academic publications will be written in the name of the collaborative group (“on behalf of the CVD-COVID-UK/COVID-IMPACT Consortium”) with all relevant individual contributions (coordination, writing, analysis, interpretation etc.) listed, the coordinating role of the BHF Data Science Centre and any relevant funding (including BHF Data Science Centre funding) acknowledged. In addition, some data custodians require specific acknowledgement of use of their dataset in reports and publications arising from the work of the consortium. Prior to submission, all reports and manuscripts for potential publication must be checked for compliance with these requirements by the BHF Data Science Centre coordination team who will ensure that these checks are conducted in a timely manner.
- A list of the members of the consortium and their institutional affiliations will be maintained by the BHF Data Science Centre coordination team and placed in the public domain on the BHF Data Science Centre webpages / website.
- Access to data within trusted research environments (TREs) for relevant researchers within the consortium will require institutions to sign various agreements. Relevant members of the consortium should do everything they can to ensure that this is done as quickly as possible so as not to impede the progress of our research. In particular, when institutions are added as joint data controllers to the data sharing agreement required for researchers to access the NHS Digital TRE, or when new data sources are incorporated into that TRE, existing signatories will be required to re-sign the agreement. Sign off is a simple and rapid electronic process. The BHF Data Science Centre coordination team will work with NHS Digital to limit the number of times this agreement needs to be re-signed.
- Accredited researchers working within the TREs agree to only perform analyses for the purpose of producing outputs for projects approved by the CVD-COVID-UK/COVID-IMPACT Approvals & Oversight Board, and within the scope of the project approvals.

Appendix B: Joint data controller responsibilities framework for COVID research in the NHS Digital Trusted Research Environment for England

(V1.0 – 27/05/2021)

1. Background

NHS Digital has developed a trusted research environment (**TRE**) for England, within which approved researchers who are members of the CVD-COVID-UK consortium are conducting COVID-related research on health data from the population of England.

[NHS Digital's TRE service for England](#) is a secure analysis environment that makes de-identified person-level data (i.e. data that has had all directly personally identifying information, such as name, address, exact date of birth, and NHS numbers removed) available for research. This data can still constitute personal data for the purposes of data protection law (including the UK General Data Protection Regulation (UK GDPR) and the Data Protection Act 2018). Access to the data within the TRE is restricted to approved, named researchers with certified training in the safe use of health-related data for research. These approved researchers access the data held in each environment remotely through a highly secure authentication system. The researchers conduct analyses of data for the purposes of approved research in the TRE. Before researchers can share results of their analyses outside the environment (e.g. within published reports or presentations), trained output checkers at NHS Digital first check the results of analyses conducted within the TRE to ensure that no person-level data or potentially identifiable data are included.

2. Access to the NHS Digital TRE for England

Before researchers based in CVD-COVID-UK consortium partner organisations can obtain access to the TRE, the following conditions must be met:

- individual researchers must become consortium members and agree to the consortium's principles (see CVD-COVID-UK Ways of Working section 2);
- the partner organisation must have a Data Sharing Framework Contract (DSFC) in place with NHS Digital;
- the partner organisation must meet the required standards of [the NHS Digital Data Security and Protection Toolkit \(DSPT\)](#); or equivalent and
- the partner organisation must sign the Data Sharing Agreement (**DSA**) between NHS Digital and all partner organisations jointly accessing data within the TRE supporting the CVD-COVID-UK consortium's work. In signing the DSA, each partner organisation agrees to ensure that researchers from that organisation who will access the TRE have appropriate information governance and data protection training in relation to the use and storage of health data and have a confident understanding of their responsibilities (DSA ref: DARS-NIC-381078-Y9C5K).

3. TRE Responsibilities

- NHS Digital will:
 - ensure that an appropriate system security policy and appropriate technical and organisational controls are in place for the TRE in order to ensure the security of all data processing and personal data in the TRE – these measures include the

- provision and operation of a secure authentication system for user access to the TRE and ensuring that only de-identified data is uploaded to the TRE for analysis (i.e. with the direct identifiers removed), as defined within the approved DARS application and associated DSA;
- design and maintain the TRE (including related access processes) in a manner that meets the data minimisation and other principles under data protection law principles and ensures that the data in the TRE is accurate and maintained;
 - make available information to the data subjects to whom the data in the TRE relates (the **Data Subjects**) regarding the processing activities and joint control arrangement in an appropriate and legal manner;
 - be responsible for responding to any requests from the Data Subjects seeking to exercise their rights under data protection law in a timely manner;
 - carry out an assessment of potential high risks to the rights and freedoms of natural persons that may arise from the data processing activities relevant to the TRE and make this data protection impact assessment available to partner organisations upon request;
 - make itself available as the first point of contact for supervisory authorities or individuals that have questions or legal requests relating to the data in the TRE; and
 - be responsible for notifying data protection supervisory authorities and, where applicable, notifying affected Data Subjects in the event of a personal data breach affecting the data in the TRE.
- Each partner organisation that is a party to the DSA will ensure that researchers from that organisation who access the TRE:
 - are aware that they must not share their personal TRE authentication and password with any other person;
 - only access the TRE from a secure device approved by the partner organisation;
 - do not use insecure software when accessing the TRE;
 - do not access the TRE from a public place or other insecure environment (e.g. coffee shop or public transport);
 - follow any guidelines or policies issued in relation to the TRE and wider project;
 - do not attempt to remove any data from the TRE except in accordance with the approved processes for extracting outputs from the TRE (i.e. subject to review of the output by trained NHS Digital personnel);
 - do not attempt to re-identify any of the Data Subjects to whom the data in the TRE relates;
 - only conduct analyses that comply with the relevant legal, ethical and regulatory approvals and that fall within the remit of the project(s)* that they have approval to work on; and
 - promptly notify NHS Digital in the event that they receive a request from a Data Subject seeking to exercise their rights under data protection law.
 - Each of NHS Digital and the partner organisations will also be responsible for maintaining a record of its processing activities relevant to the TRE as required by UK GDPR and complying with their respective obligations under data protection law.
 - Each of NHS Digital and the partner organisations will notify HDR UK before communicating with a supervisory authority in respect of the processing and take into account any reasonable comments made by HDR UK.

- Only aggregated summary data can be exported from and shared outside of the TRE (for example in reports to health policy makers or manuscripts for publication in medical and scientific journals). NHS Digital will ensure that analysis outputs and code are subject to appropriate disclosure control by trained output checkers to ensure that no potentially identifiable data (on patients or practitioners) are released.
- A concise, transparent privacy statement will set out the purposes for processing personal data and explain how data is used for these purposes. This [privacy statement](#) is maintained by the BHF Data Science Centre on behalf NHS Digital and the consortium partner organisations and accessible on the BHF Data Science Centre [webpage](#).

Version Control

Version	Date	Author(s)	Comment
1.0	28/05/2021	RP, JMA	<ul style="list-style-type: none"> Final version incorporating feedback from WW, AA, SD, CS, LM, DR.
1.1	09/06/2021	RP	<ul style="list-style-type: none"> Section 6.3: notice to comms manager changed to a minimum of 5 working days. Included requirement to circulate the manuscript to all consortium members for information/comment. Section 8: Expanded HDR to Health Data Research in Acknowledgements
2.0	25/05/2022	RP, LM, TB, JN, DR, CS, SD	<ul style="list-style-type: none"> Complete refresh of the document
2.1	10/11/2022	RP, LM, KMA, SK, JN, TB	<ul style="list-style-type: none"> Section 4 and 6.3.1: offer of a meeting with the public contributors to discuss outcomes/results dissemination Section 5.2: funding via D&C to March 2023 Section 6.1.3: addition of detail for NHSD TRE induction and notification when users no longer need TRE access Section 6.1.6.1: phenotype code list process changes Section 6.3.3: greater clarity on communications for preprints and publications Section 6.4.1: further guidance on uploading phenotype code lists to GitHub