SOP number	**59.006**	Version	**1.0**
Title	**Creation of Data Extracts from the Safe Haven**		

Prepared by Signature	Charlie Mayor Date
Approved by Signature	Chloë Cowan Date
Released by Signature	Julie Brittenden Date

SOP category	West of Scotland Safe Haven
Staff category	

Staff Category	**R**	**A**	**C**	**I**
West of Scotland Safe Haven Manager		X		
Data Management	X			
Data Processing Staff	X			
Database Development Staff	X			

1. Scope

This procedure applies to the Safe Haven within NHS Greater Glasgow and Clyde.

2. Purpose

The purpose of this SOP is to describe the procedures relating to the creation of data extracts from the Safe Haven.

3. Procedures

3.1. Data Requests

The Safe Haven team will have prepared a Project Data Specification for the extract, with details of the application request, using template Form 59.006A. This project specification will have been reviewed, amended as required and signed off by the project Chief Investigator. This Project Specification will be placed in the 'Safe Haven Projects' folder in the shared Safe Haven network drive. The data request will then be added to the Safe Haven work plan.

If the technical specialist/analyst feels the Project Specification is incomplete in any way, more detail should be requested from the Project Manager. The file should be checked for any inconsistencies or wrong information (such as a field name attributed to the wrong dataset). If anything is deemed to be wrong with the file, this should be highlighted to the Project Manager and de-prioritised until corrections are made. Once a Project Specification has been received and prioritised, the extracts should be worked on in priority order as defined on the waiting list.

3.2. Creating the Project Database

The SH team will create a blank working database using scripts provided by BI, for the project on the TEST server (NHSGGCBISQLTEST), called GSHYYAANNN_Project_Short_Title (using the Safe Haven project number as defined on the Project Specification and created according to SOP 59.007). All

data extracts for the project should go into, and be manipulated within, this database. BI should be notified in advance when creating a new TEST database.

3.3. Creating the Cohort on the TEST server

Using the 'Cohort' tab in the Project Specifications file (there may be more than one, dependent on the project design) create a table which will be used to define the cohort and to store the subjects who match the criteria within these sources.

Create the cohort according to GUI 59.006A_Cohort Criteria Query.

Each stage of the query should be checked separately to quality check the process and sense-check the numbers of subjects included as the cohort is being built. There will be a checklist that the analyst will have to sign off to record that these checks will be done.

If you are carrying out a feasibility or aggregate screen, this stage will be your end point and the expected output should be as detailed in the Project Specifications file (Form 59.006A).
Each of these steps will be recorded in the project Extract Report, Form 59.006B, by the individuals performing them.

3.4. Pulling the Data Requirements of the Project

The Project Data Specifications, Form 59.006A, file will be used as an explicit reference tool. Using the Data Requirements tab (there may be more than one, relating to the cohort tabs, dependent on the project design), query the relevant default datasets and pull any additional fields not in the default datasets. All the queries start with a number as per the instruction on ref document (fill this in). The numbers should reflect the order in which the queries will be run.

All datasets will be produced from the Safe Haven as anonymised datasets unless explicitly stated when the data is requested and approved that identifiers are acceptable for that particular project. Detail of which identifiable fields are allowed will have been added to the project specifications. This would only be for a project that has ethical approval and informed consent from patients specific to the project that allows the researcher to view the data in an identifiable format.

Use SQL scripts to anonymise the data to be extracted. CHI numbers are replaced with a Safe Haven Id before output. A summary characteristics table will be generated which describes the dataset produced from the SQL scripts and this will be the extract report. Another member of the team will be responsible for checking the SQL scripts and data produced at this stage.
Once another Safe Haven analyst has checked the SQL scripts and verified results, the Safe Haven analyst will schedule on the LIVE (NHSGGCBISQLPROD) database a time to run the scripts
Each of these steps will be recorded in the project Extract Report, Form 59.006B, by the individuals performing them.

3.5. Extracting the Data on the LIVE (PROD) server

The Safe Haven Analyst will run the scripts on the live server at the allocated time.
A second summary characteristics table will be generated.
If an error is found in the extract, the Project Manager should be notified as well as the analyst who created the extract. The analyst should correct the error and return the file for checking again before release. The Project Manager should ensure that this process is documented.
Each of these steps will be recorded in the project Extract Report, Form 59.006B, by the individuals performing them.

3.6. Extracted Outputs

Extract files can be produced in various forms depending on what the research request is.
They usually come in one of three forms;

- Aggregate data: A list of numbers indicating how many patients have met the specified criteria. Researchers normally request this kind of data to support a grant application or when they are trying to assess whether the population exists for a study to be carried out.
- Anonymised data: This would be a number of files containing clinical data for a number of patients which has been anonymised through removing any identifiable information.
- Identifiable data: This data would only be released where specific Caldicott permission has been given. Even if data is consented, it would normally be anonymised and a study code used before sending out to a researcher.

Regardless of the type of output produced for the extract, all outputs should be subject to disclosure control before being released. This should be done via the checking process outlined in SOP 59.012. The data will now be considered as suitable for release as per Transfer of Electronic Data SOP 59.005. An email should be sent to the Project Manager to alert them that it is there.

Each of these steps will be recorded in the project Extract Report, Form 59.006B, by the individuals performing them.

4. Referenced documents

- SOP 59.006 - Data Management: Transfer of Electronic Data
- SOP 59.007 - Assigning Local Study Numbers and Subject IDs
- SOP 59.012 - Quality Checking Data Extracts from the Safe Haven
- Form 59.006A - Project Specifications Template
- Form 59.006B - Extract Report
- GUI 59.006A - Cohort Criteria Query

5. Related documents

- Data Management Plan

6. Document history

Version	Date	Description
1.0		Migration to Glasgow Clinical Trials Unit QMS