

SOP number	59.012	Version	1.0
Title	Quality Checking 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			

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 quality checking of data extracts from the Safe Haven

3. Procedures

3.1. Quality Checking of Extract Files

All datasets will be produced from the Safe Haven as fully anonymised datasets, according to SOP 59.006, unless explicitly stated when the data is requested and approved that identifiers are acceptable. If identifiable data is contained within the extract, check it has been requested within the Project Specifications (Form 59.006A). The first quality check will be completed by the Safe haven analyst responsible for extracting the dataset. Subsequently, extracts will be independently checked at two separate stages; first the syntax used for the extract will be checked before the data is extracted and then after the extract has been run the data tables and files will be checked. In both cases, these checks should be completed by a different Safe Haven analyst from the one who conducted the extraction. The checking process must be recorded using Form 59.006B - Extract Report, for audit purposes.

3.1.1. Independent Safe Haven Analyst Check of Syntax

The analyst should sense check the syntax used, paying particular attention to the build of the extract cohort. Date ranges used should be appropriate, as should any codes used for filtering. The analyst should check all requested criteria have been applied and any calculations are correct.

3.2. Independent Safe Haven Analyst Check of Extract

The following checks should be made for all extract tables and files:

Anonymised IDs are in range across all files and that the files can be linked

- All data requirements are contained within the files as specified within the data definitions in the Project Specifications
- Highlight reasons for any data which has been 'extracted' but is NULL, i.e. the data has been queried but there are no results
- Any Date of Birth data points that are out of range/abnormal should be highlighted
- All fields are in the correct format, for date fields this should be YYYY-MM-DD
- If previously indicated within the Project Specifications, the number of records is in the expected output range
- Only patients who are resident within NHSGGC health board (the research population) should be included in the extract unless there are specific permissions to include others outside this population
- If any data files have been provided by the researcher, these do not need to be checked in detail but they should be checked for anonymised ID consistency between files and presence in the summary sheet
- A summary sheet of characteristics, including the date of extraction, has been provided for the data extract and that all relevant information is there
- Potential disclosure within the files arising from combinations of data points
- Any radiology & microbiology or other text reports should be checked for names of staff or patients within the body of text
- The file names are consistent and should contain the project number
- Unless otherwise specified, files should be of .csv format

3.3. Metadata to be provided

Once checking has been completed the following information should be provided about the extract files:

- Any data queried which has produced no results
- Any identifiable fields and what makes them identifiable (i.e. contain person identification numbers or due to a combination of data points) and any potential disclosure risks
- Any data which is out of range/abnormal
- Any text reports which contain names of staff or patients
- Any files which cannot be linked or anonymised IDs which seem incorrect
- Any file names which are inconsistent

4. Referenced documents

- SOP 59.006 - Creation of Data Extracts from the Safe Haven
- Form 59.006A - Project Specifications Template
- Form 59.006B - Extract Report

5. Related documents

- Data Management Plan

6. Document history

Version	Date	Description
1.0		Migration to Glasgow Clinical Trials Unit QMS

This SOP is a controlled document. The current version can be viewed on the GCTU website.
Any copy reproduced from the website may not, at time of reading, be the current version.