



Producing the Source Linkage Analysis Files

Standard Operating Procedure (SOP)

**Synopsis**

*This document provides step by step guidance for the successful production of the source linkage files at episode and individual level.*

**Document Control**

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Date** | **Change** | **Author** |
| 0.1 | 11 August 2016 | First draft | D Hastie |
| 0.2 | 31 August 2016 | Several additions made to process – due to refinements/amendments noted during process update | D Hastie |
| 0.3 | 12 December 2016 | Adding in more detailed notes for the Business Objects section | D Greig (nee Hastie) |
| 0.4 | 28 December 2017 | Draft send to selected H&SC team members | D Greig |
| 1.0 | 08 January 2018 | Final version | D Greig |
| 1.1 | 08 February 2018 | Update for SPARRA Risk scores and for A&E BO query for Glasgow | A George |
| 1.2 | 31 May 2018 | Added notes for syntax Death\_Flag. | A George |

Last updated: August 2018

***Glossary***

| **Acronym / Term** | **Details** |
| --- | --- |
| A&E | Accident & Emergency |
| ACaDMe | Acute, Cancer, Death and Mental Health data marts. All are accessible through one universe (ACaDMe). |
| AE2 | Accident and Emergency data mart  All attendances are recorded for A&E. The majority is recorded as individual hospitals, but there are some hospitals that return aggregate data only (i.e. not patient level). |
| BI | Business Intelligence |
| BOXI | Business Objects XI is a web-based reporting tool that allows users to create reports which can be made available to other users who have the appropriate access permissions. |
| BOXI Universe | The BOXI universe is the reporting environment which is built on and allows access to the associated data mart. |
| CHI | Community Health Index  Health service number to represent a patient. Note that a patient (predominantly pre 2003) could have had more than one CHI number held in the CHI system (if they had moved GP Practice) |
| CSD | NSS IT Customer Support Desk |
| Data Mart | A subset of a data warehouse which has been designed to satisfy a particular function or reporting area within an organisation. A data mart can be designed and developed independently of other data marts. A data warehouse will contain multiple data marts. |
| GLS | Geriatric Long Stay |
| H&SC | Health and Social Care |
| ISD | Information Services Division |
| IT | Information Technology |
| LIST | Local Intelligence Support Team |
| LTC | Long Term Condition  A disease that a patient has contracted and will be impacted by for the rest of their life. |
| NRS | National Records for Scotland |
| PHI | Public Health & Intelligence |
| PLICS | Patient Level Costing – this is a methodology. Source linkage files were previously called PLICS analysis files; but they should not be referenced by this name anymore. |
| SBU | Strategic Business Unit |
| Service Now | Web based tool to raise calls requiring IT assistance (available to all staff) |
| SMR | Scottish Morbidity Recording |
| SMRA | Oracle database that contains views on database tables produced from the MRL process. Note that other data sets (SMR data and non-SMR data) are also held in SMRA. |
| SMR00 | Outpatient appointments at NHS and domiciliary locations. |
| SMR01 | Acute inpatient and day cases discharges from non-obstetric/non-psychiatric NHS hospitals in Scotland.  Geriatric Long Stay records are returned as part of this dataset. They are identified by the significant facility code 1E. |
| SMR02 | Maternity inpatient and day case discharges from NHS hospitals in Scotland. Some home births are also recorded. |
| SMR04 | Mental health inpatient and day cases admissions and discharges from psychiatric NHS hospitals/units in Scotland (includes The State Hospital). |
| SOP | Standard Operating Procedure |
| Source Linkage Files | There are two Source linkage files created for each financial year. Episode level (all episodes from selected data sets) and individual level (aggregated file with information contained on one row per individual). |
| SPARRA | Scottish Patients At Risk of Readmission or Admission |
| UAS | User Access System – used by PHI (and others) to request permission to the majority of data marts available in the CDW. |
| UNIX areas | Network area held on a UNIX server and hosted by NSS IT  e.g. UNIX irf, UNIX hscdiip |
| UPI | Unique Patient Identifier  This is the ‘latest’ CHI number of a patient. The UPI number is applied to all records belonging to one patient and is the classed the unique identifier for a patient. |
| URI | Unique Record Identifier |
| DD | Delayed Discharge |
| DN | District Nursing |
| OOH | Out of Hours – Usually referring to the GP Out of Hours dataset. |
| COSLA | Convention of Scottish Local Authorities. The Convention of Scottish Local Authorities is the national association of Scottish councils and acts as an employers' association for its 32 member authorities. |

Contents

[Scope 6](#_Toc521674600)

[Introduction 6](#_Toc521674601)

[Procedures 7](#_Toc521674602)

# Scope

The purpose of this standard operating procedure (SOP) is to outline the procedures used when producing the Source Linkage files for analysis predominantly in the Health and Social Care Team. Some LIST members will also use these files for analysis for the partnerships they are supporting analytically.

# Introduction

The ACaDMe linked data set in ISD only contains acute (SMR01, SMR01 (GLS)), mental health (SMR04), cancer and death records (NRS). For health and social care integration as many main services that we have patient level costing for is required to support analysis. The data sets that are linked together for the Source Linkage Analysis Files are:

* Acute (including Geriatric Long Stay (GLS)) – ACaDMe;
* Maternity – Maternity Datamart / SMR02;
* Mental Health – ACaDMe / SMR04;
* Outpatient appointments – Outpatients Datamart / SMR00;
* Accident & Emergency Attendances (episode level only) – AE2 Datamart;
* Limited Prescribing information – PIS datamart (extract provided by CSD);
* NRS death registrations – ACaDMe and extract provided by CSD.
* District Nursing – CHAD Datamart
* GP Out of Hours – GP OOH datamart (part of unscheduled care)
* Care Homes – Social Care datamart (data for 17/18 is only partial in old datamart and will only be for Q4 onwards in new datamart)
* Delayed Discharge – File prepared by the Delayed Discharge team.
* Non-service-users – A list of CHIs which otherwise would not be in the linkage files.

In addition, where possible, extra data is linked to, or calculated for, each record including.

* Patient’s deceased status and best estimate of death date from all sources.
* Long Term Condition (LTCs) markers and dates of incidence for 19 LTCs.
* The demographic and cost cohort that patient belonged to that year.
* HRI flags and information.
* Patients SPARRA and HHG predictions for 12-months from the start of the financial year.

There are two Source linkage files: The ‘episode’ file and the ‘individual’ file. The episode analysis file contains all patient records at episode level (the lowest grain of the data), with the exception of Prescribing data, which is included as an aggregate for the year. The individual analysis file is an aggregated file summarising the data in the episode file into one row per individual (where a valid CHI number is available).

Note that the data files are produced for individual financial years and are available from 2010/11 to the current financial year.

# Procedures

This section of the SOP details the procedures for the production of the Source linkage files.

## **Important things to note**

The schedule for updating the Source Linkage Files is here:

[\\Freddy\dept\PHIBCS\PHI\Health & Social Care\Topics\Linkage\Reference Documents\Schedule for updating Source Linkage Files.docx](file:///\\Freddy\dept\PHIBCS\PHI\Health%20&%20Social%20Care\Topics\Linkage\Reference%20Documents\Schedule%20for%20updating%20Source%20Linkage%20Files.docx)

Access to the following data marts/UNIX areas is required by the analyst producing these files.

### Business Objects

Public Folders > Scotland > NHS > NSS > IRF Reports > Standard Reports > 10 Source extract reports

| **BOXI Data Marts** |
| --- |
| ACaDMe |
| AE2 |
| Maternity |
| New\_PIS (Prescribing) |
| Outpatients |
| District Nursing (CHAD) |
| GP Out of Hours |
| Social Care |

A good working knowledge of each of the data sets is desirable, but not essential.

### Procedure for gaining access to this area with the Public Folders via Business Objects

**Use an access to data form and i**n Dataset/Directory, **enter;**

National Power User for the IRF reports folder LDAP group BONHSDIIRFREP-U [BO NHS NSS (ISD) IRF Reports Power Users]

### Confidential network areas on the nssstats01 server

You will also need an access to data form to become a member of the following groups / areas.

UNIX irf

UNIX hscdiip and UNIX hscdiip\_sl (this allows write access to hscdiip)

UNIX sourcedev

Mapping the areas to drive letters will make things easier (e.g. [\\stats\hscdiip](file:///\\stats\hscdiip)).

DO NOT USE UNIX cl-out (/conf/linkage/output) for saving any files, even temporarily.

It is useful to keep checking how much space is available in the confidential network area being used to save any of the SAV files. If the amount of available space becomes 0 during a save command, then the file that was being saved at the time will be lost completely – If a file is being overwritten the existing file will also be lost.

## Prior to starting the process

The mapping process (PLICS – which applies costs to episodes) usually starts at the end of November (when the cost book is published) and can sometimes continue up to July of the following year. The updating of the Source Linkage Files will impact this process as only finalised costs can be used in the Source Linkage Files. For the March and May updates, work through the following process.

Send an email to the costs team;

Hi Enter name(s),

Can you please advise me what stage the mapping process is at with regards to what reference files have been loaded in to the following data marts within the CDW:

Acute;

Maternity;

Mental Health;

Outpatients;

A&E?

As reference files are loaded in to the production environment, can any un-finalised costs in the data marts please be replaced with the finalised costs from the previous financial year costs by INSERT DATE? Ideally, all costs should belong to one financial year.

*Advice on replacing provisional costs with previous finalised costs*

To put the reference costs ‘back’ to the previous financial year, the previous costs reference files data needs to be entered into a copy of the reference files with the same name as the current provisional costs. It is not simply a case of re-loading the previous financial year costs.

E.g. If 2015/16 are finalised and 2016/17 costs are provisional. The reference files with 2016 in the title will need to be populated with the 2015/16 reference data.

As the costs are forward populated, then as soon as reference files for 2016/17, albeit provisional are successfully loaded, then all data from 2016/17 onwards will be costed with the provisional costs. Re-loading the 2015/16 costs with their current file names will not re-calculate the costs for 2016/17 onwards.

Once the finalised costs have been loaded back in, then I will create the relevant data extracts as quickly as possible to minimise the impact on the mapping process.

Much appreciated,

**Enter name**.

## Produce the Source episode level analysis file

The high level steps involved in producing the Source episode level analysis file are summarised below.

1. Run the reports and extract the data (csv) for Acute (inc. GLS), Maternity, Mental Health, Outpatients, A & E, NRS death records, Care Home, District Nursing and GP Out of Hours from Business Objects.
2. The PIS extract, LTC markers and Derived date of death are provided by the BI team.
3. Syntax is run to produce up to date lookups for later matching onto postcode and GP practice.
4. Outputs from BOXI/BI team are read in to SPSS and data items from the different data sets are standardised for name and format. Monthly costs and bed days are created for acute where possible (most datasets).
5. Data sets are then combined, excluding Delayed Discharge.
6. The Delayed Discharges are processed (they need to be compared to the linked file to ensure they match existing records).
7. Fixes and corrections are made to various variables, including the death dates.
8. Pathways and HRIs are calculated.
9. SPARRA and HHG scores are matched on.
10. Geography/deprivation data items are matched on.

## Requesting Large Data Extracts

A week prior to the start of the quarterly update process of the source linkage files, a request should be made via CSD for large data extracts to be produced by the BI team within the IT SBU.

Log a call using the self-service helpdesk (green shopping trolley icon should be on your desktop, otherwise phone CSD on 7777).

*NOT TO BE INCLUDED IN Request - In the request, reference is made to the previous request for extracts 3 & 4 by number. Reference number and month/year last produced will need to be amended every quarter data is requested.*

Hi there,

Can you please log a call with the BI team?

I would like to request 4 data extracts. I cannot get this information from Business Objects myself.

Extract 1 – all patients with one of the 19 LTCS (as defined in the LTC project, which was part of the HSCDIIP project (2015)). Patient UPI numbers and the data of incidence for each LTC is required.

Extract 2 - all patients with a derived date of death. Patient UPI number and derived date of death are required. Please ensure that there are no duplicate UPI numbers.

Extract 3 – same as extract 3 from previous request, SCTASK0036334 for all available PIS data in 2017/18.

Extract 4 – same as extract 4 from previous request, SCTASK0036334 for all available PIS data in 2017/18.

These extracts were produced in November 2017 with the reference number SCTASK0036334. It is an update that is required.

If I could have these extracts by **INSERT DATE** that would be great. I will send Service Manager approval as soon as I have this.

Many thanks,

**Enter name.**

Send another email to the H&SC Service Manager *(currently Susan Frame)*

Dear Susan,

Can you please provide me with authorization for the large data extracts that I’ve requested from the BI team via CSD? (*8* extracts covering, LTCs, Deaths and PIS data for *2016/17, 2017/18 and 2018/19*).

The extracts are required for the update of the Source Linkage Files. You can just send the approval to me and I can forward this on the BI team member when they make contact about the request.

Many thanks,

Enter name.

Please note that the BI team member who is assigned the call may ask for Business Objects reports to be sent to them. If this happens reply to the request with the following:

**Email to BI team member (if BO reports are requested)**

There aren’t any BO reports available for these four extracts as the information cannot be extracted via the Business Objects universes. Please see the previous request, the scripts used should be available and they can be updated where necessary. Linda Sheppard within the BI can advise them as she has provided these data extracts several times.

The BI team member will also ask where the extracts should be saved. Advise the BI team member that they should be placed in [\\stats\sourcedev\#Extracts](file:///\\stats\sourcedev\%23Extracts)\FolderFor This update# (UNIX sourcdev). Ask them for their user id and advise that you will arranged their access to this UNIX area. To do this, work through the following steps:

Open an Access to Data form, complete sections: Member of staff requiring access; Group; Team; Network login name. In Dataset/Directory 1 **enter UNIX sourcedev.** In Reason for Access **enter**: For save large data extracts for the H&SC team

Email Andrew Lee (or Band 7 for Source Linkage)

Hi Andrew,

Can you please approve access to UNIX sourcedev for **INSERT NAME**, who is the BI team member working on my request for large data extracts?

Thanks,

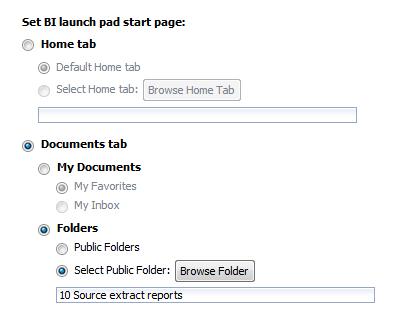
**Enter name.**

### Making data extracts available for analysis file creation

The data extracts received from NSS IT are CSV files that have been zipped up. These files need to be unzipped before being used.

## Scheduling BO reports

Due to the volume of data for some of the data sets (and also to ensure that purged reports only exist in the IRF Reports folder), the BO reports required for extracting the data for the Source Linkage Files are scheduled. A scheduled report can be thought of as a report that is running in the background. Scheduling also allows us to directly extract the data as a csv (instead of having it appear as a Web Intelligence report first).

1. **Log in** to Business Objects (<https://www.bo.scot.nhs.uk/BOE/BI>)
2. **Navigate to** the folder (click on the Folders button): Public Folders > Scotland > NHS > NSS > IRF Reports > Standard Reports > 10 Source extract reports (This can be set as the default folder upon opening BOXI by changing your preferences).
3. **Right click** on the first report (a & e all scotland episode level extract for source file) and select **Schedule**.
4. **In Instance Title –** Add the financial year to the end of the instance title e.g. 201718.
5. **In Prompts** – Click modify, then set the Financial Year (2017 = 2017/18 etc.).
   1. **Click** Apply
6. **In Formats –** Tick ‘Comma Seperated Values (CSV)’
7. **Click** Schedule

**Repeat steps 3 – 7** for all other reports in this folder

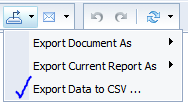
**Right click** on report to check history to see if the report has completed successfully.

Note that all the reports in the folder detailed in point 2 have not been run, except via being scheduled. These reports should not be run and saved with data in them.

### Recommended, but not essential

**Create a**n excel workbook, to keep a check on what has been done. Listing reports to be scheduled and then downloaded. As well as syntax files to be run.

Saving BO instances

Extract the data as csv. If you set the format as CSV when scheduling this can be done by clicking on the relevant report in the query’s history and then picking a suitable download location.

Otherwise click on the report and wait for it to load as a web intelligence doc, then click on ‘Export Data to CSV’ from the export menu.

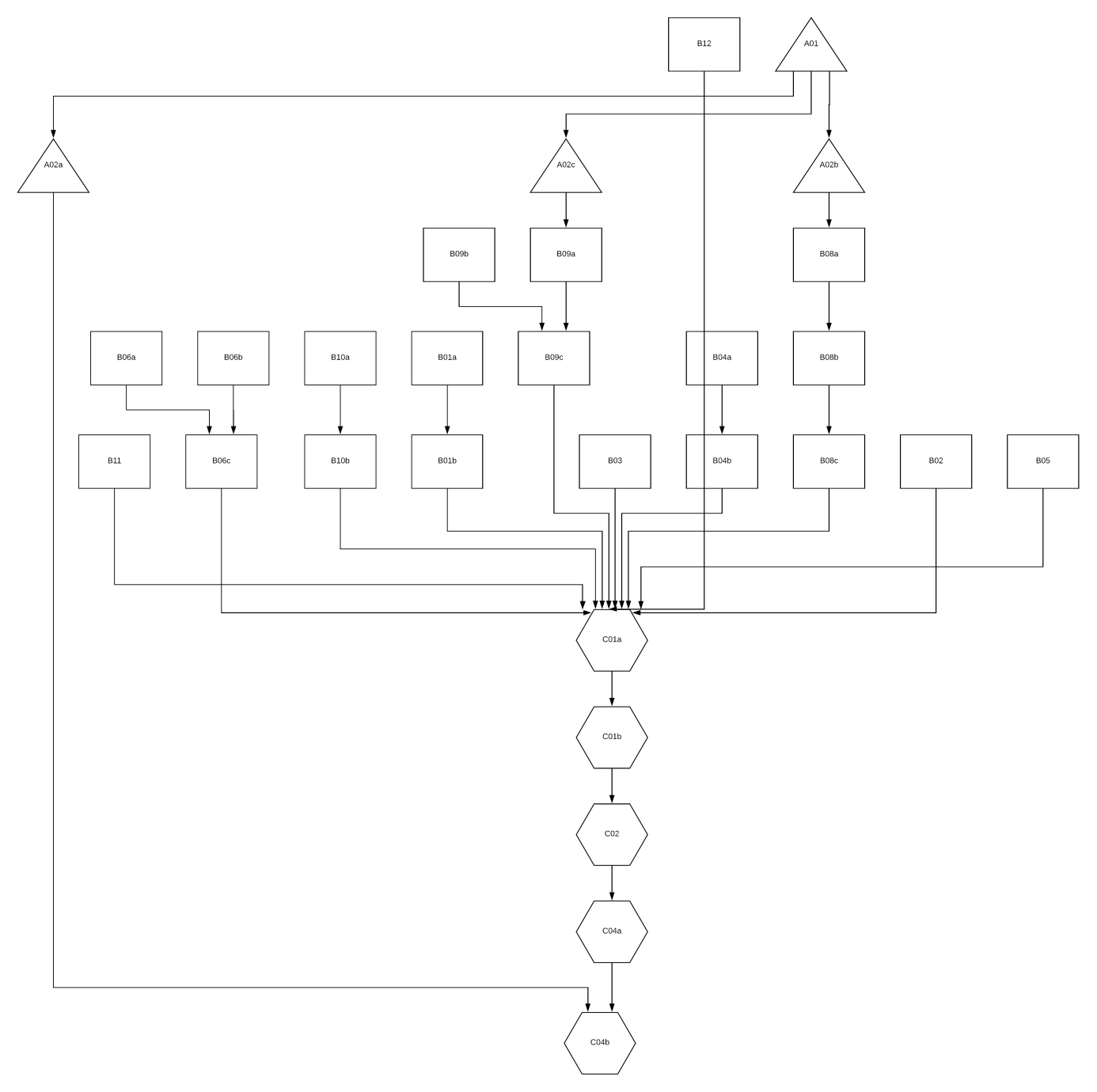
The following BO extracts are required to produce the source linked analysis files. Instances for the latest time period should be produced using the method described above.

|  |  |
| --- | --- |
| **Record type** | **BO extract query/report name** |
| Acute | acute all scotland episode level extract for source file production – excluding stay  acute line number by uri |
| Outpatients | outpatients all scotland episode level extract for source file production |
| Maternity | maternity all scotland episode level extract for source file production |
| Mental Health | mental health all scotland episode level extract for source file production  mental health all scotland los by uri extract for source file production |
| A&E | a & e all scotland episode level extract for source file production |
| NRS Death Registrations | nrs death registrations all scotland episode level extract for source file production |

Note that for any data sets where there are the variables *Length of Stay (days)* and *Occupied Bed Days*, these two measures cannot be returned in the same query, as it results in two sets of data being created in the output. Length of stay should be calculated in the SPSS programs using the date of admission and discharge. Always return the Occupied Bed Days via the Business Objects query.

## Running programs to create files for linking together

The syntax for the source linkage is structured to keep the process as logical and efficient as possible; we try to make sure things are only done once, when needed and that everything that is done serves a purpose. There are a large number of syntax files which should generally be run in sequence, although many in section B can be run in parallel.



The process detailed below describes the steps required for creating a new financial year analysis file. If an update is being carried out for an existing financial year, simply work methodically through the existing programs in the existing financial year folder (which is pretty much the process outlined below but missing out Step 1).

TLDR; Run all of the syntax files in order, making sure to run the macros in A01 first if you close SPSS mid-way through the process.

1. Create a syntax folder for the new year
   1. Open Windows Explorer
   2. Map [\\nssstats01\irf](file:///\\nssstats01\irf) to a network drive (if this is not already mapped)
   3. Navigate through the folders, double click on 11-Development team
   4. Double click on Dev00-PLICS-files (PLICS should be changed to source)
   5. Create a new folder labeled with the financial year (naming convention is YYYY-YY)
   6. Copy the syntax files from the previous financial year folder in to the new folder
2. Log onto to SPSS stats server.
3. Run A01 Set up Macros

This syntax needs to be updated to the relevant financial year. The file locations, for storing files throughout the creation process and the location where the extracts have been saved should also be checked and amended if necessary.

1. Run A02a Create Postcode Lookup and A02b Create GP practice lookup

These files produce a lookup for postcode and gp practice respectively. They start by getting the latest postcode / gpprac list and then matching data we will need onto that. The macros at the top defining file locations should be checked to make sure they are pointing to the most up to date version of files, or the most applicable file for that financial year. If there are any errors it will likely be because of incorrect filenames, these should be checked and updated.

1. Run A03 and A04

These produce lookups which are needed later; it is useful to run them at this point as they may require updated files from other teams. Errors here could be caused by the shapes of the files changing, or files not stored under the expected name in the extract folder.

1. Run all of the syntax prefixed ‘B’

Most of these files can be run in parallel, either by multiple analysts or on multiple computers, with the exception of the ones which have a letter suffix (a, b etc.) as these rely on the previous syntax.

For each syntax file, take care to check the output, looking for any errors, warnings are sometimes ok, depending on the context! If the syntax completes without errors, check the final file, making sure the data looks right (e.g. genders should be mostly 1s and 2s, geog codes should be 9-char starting with S, dates should be mostly filled in and look sensible) use Ctrl + D to skip to the bottom of the file.

**Go to** File > Open > Syntax and **navigate** to mapped samba share for UNIX irf, to the folder 11-Development team > Dev00-PLICS-files > **Relevant Financial year**

**Double click** on Program 10a: 10a-read-in-acute-uri-linenumber-source.sps

**Ensure** the defines for !file and !FY are correct for file location(s) and 4 digit financial year

**Save** any changes (CTRL + S)

**Select** the whole program (CTRL + A)

**Run** the program

Once complete, **check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Close** Program 10a

**Go to** File > Open > Syntax

**Double click** on Program 10b: 10b-produce-acute-file-for-source.sps

**Ensure** the defines for !file and !FY are correct for file location(s) and 4 digit financial year

**Save** any changes (CTRL + S)

**Select** the whole program (CTRL + A)

**Run** the program

Once complete, **check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Close** Program 10b

**Go to** File > Open > Syntax

**Double click** on Program 11: 11-produce-maternity-file-for-source.sps

**Ensure** the defines for !file and !FY are correct for file location(s) and 4 digit financial year

**Save** any changes (CTRL + S)

**Select** the whole program (CTRL + A)

**Run** the program

Once complete, **check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Close** Program 11

**Go to** File > Open > Syntax

**Double click** on Program 12a: 12a-read-in-mh-los-by-uri-create-reference-table-for-source.sps

**Ensure** the defines for !file and !FY are correct for file location(s) and 4 digit financial year

**Save** any changes (CTRL + S)

**Select** the whole program (CTRL + A)

**Run** the program

Once complete, **check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Close** Program 12a

**Go to** File > Open > Syntax

**Double click** on Program 12b: 12b-produce-mental-health-file-for-source.sps

**Ensure** the defines for !file and !FY are correct for file location(s) and 4 digit financial year

**Save** any changes (CTRL + S)

**Select** the whole program (CTRL + A)

**Run** the program

Once complete, **check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Close** Program 12b

**Go to** File > Open > Syntax

**Double click** on Program 13: 13-produce-outpatients-file-for-source.sps

**Ensure** the defines for !file and !FY are correct for file location(s) and 4 digit financial year

**Save** any changes (CTRL + S)

**Select** the whole program (CTRL + A)

**Run** the program

Once complete, **check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Close** Program 13

**Go to** File > Open > Syntax

**Double click** on Program 14: 14-produce-a&e-file-for-source.sps

**Ensure** the defines for !file and !FY are correct for file location(s) and 4 digit financial year

**Save** any changes (CTRL + S)

**Select** the whole program (CTRL + A)

**Run** the program

Once complete, **check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Close** Program 14

**Go to** File > Open > Syntax

**Double click** on Program 15: 15-produce-deaths-file-for-source.sps

**Ensure** the defines for !file and !FY are correct for file location(s) and 4 digit financial year

**Save** any changes (CTRL + S)

**Select** the whole program (CTRL + A)

**Run** the program

Once complete, **check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Close** Program 15

**Go to** File > Open > Syntax

**Double click** on Program 17a: 17a-read-in-pis-measures-file.sps

**Ensure** the defines for !file and !FY are correct for file location(s) and 4 digit financial year

**Save** any changes (CTRL + S)

**Select** the whole program (CTRL + A)

**Run** the program

Once complete, **check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Close** Program 17a

**Go to** File > Open > Syntax

**Double click** on Program 17b: 17b-read-in-pis-patient-file.sps

**Ensure** the defines for !file and !FY are correct for file location(s) and 4 digit financial year

**Save** any changes (CTRL + S)

**Select** the whole program (CTRL + A)

**Run** the program

Once complete, **check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Close** Program 17b

**Go to** File > Open > Syntax

**Double click** on Program 17c: 17c-create-pis-data-file.sps

**Ensure** the defines for !file and !FY are correct for file location(s) and 4 digit financial year

**Save** any changes (CTRL + S)

**Select** the whole program (CTRL + A)

**Run** the program

Once complete, **check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Close** Program 17c

**Go to** File > Open > Syntax

**Double click** on Program 18: 18-create-LTC-patient-ref-table.sps

**Ensure** the defines for !file and !FY are correct for file location(s) and 4 digit financial year

**Update** the filename on line 29

**Save** changes (CTRL + S)

**Select** the whole program (CTRL + A)

**Run** the program

Once complete, **check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Close** Program 18

**Go to** File > Open > Syntax

**Double click** on Program 19: 19-create-Deceased-patient-ref-table.sps

**Ensure** the defines for !file and !FY are correct for file location(s) and 4 digit financial year

**Update** the filename on line 35

**Save** changes (CTRL + S)

**Select** the whole program (CTRL + A)

**Run** the program

Once complete, **check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Close** Program 19

3.2.5 Creating the Source episode level analysis file

This part of the procedure provides the steps to successfully producing the Source Linkage Episode file for the financial year of interest.

**Open** SPSS and **log in** (must be on the nssstats01 server and working in UNICODE mode)

**Go to** File > Open > Syntax and **navigate** to mapped samba share for UNIX irf, to the folder 11-Development team > Dev00-PLICS-files > **Relevant Financial year**

**Double click** on Program 20: 20-create-source-episode-analysis-file.sps

**Ensure** the defines for !file and !FY are correct for file location(s) and 4 digit financial year (Lines 43-55)

**Save** changes (CTRL + S)

**Highlight** from Line 43 to Line 92 (the first save command) and **run**

**Check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Highlight** from Line 96 to Line 106 and **run**

**Check** the date in Lines 108 to 126 and Lines 129 to 147is the end of the financial year that the file is producing the episode level analysis file for.

**Update** dates (if required)

**Highlight** from Line 108 to Line 172 and **run**

**Check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run** (if re-running, may need to start from line 96)

**Highlight** from Line 176 to Line 193 and **run**

**Check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Highlight** from Line 197 to Line 247 and **run**

**Check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Highlight** from Line 251 to Line 462 and **run**

**Check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Check** file location in line 466 – amend if required

**Check** the available space in the network area where the file is to be saved to ensure that there is enough space. To check the available space, do the following:

**Open** Windows Explorer

At the left hand side, **click on** where the user name and PC asset tag number are displayed

**Look** through the Network Location to find the samba share and the space will be noted

If there is enough available space (at least 7GB), **proceed** to Step 26

If there isn’t enough available space, then run Lines 703 - 709 (Erase File commands) – this should create space

**Repeat** similar actions to step e until there is enough space

**Highlight** from Line 466 to Line 500 and **run** (Full save command for the source linkage file)

**Check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Check** available network space as detailed in Step 25

**Check** file location in lines 507 and 522 – amend if required

**Highlight** from Line 507 to Line 528 and **run**

**Check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Check** available network space as detailed in Step 25

**Check** file location in line 536 – amend if required

**Highlight** from Line 536 to Line 669 and **run**

**Check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Check** available network space as detailed in Step 25

**Check** file locations in lines 673, 680, 690 – amend if required

**Highlight** from Line 673 to Line 691 and **run**

**Check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Check** available network space as detailed in Step 25

**Check** file locations in lines 695 - 697 – amend if required

**Highlight** from Line 695 to Line 697 and **run**

**Check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Highlight** from Line 703 to Line 709 and **run** – only applicable if Step 25e has not been carried out

**Check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Close** Program 20

**Check** available network space as detailed in Step 25

**Go to** File > Open > Syntax

**Double click** on Program 22: 22-locality&clusters.sps

**Ensure** the defines for !file and !FY are correct for file location(s) and 4 digit financial year (Lines 1-15)

**Highlight** from Line 17 to Line 68 and **run**

**Check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Close** Program 22

**Check** available network space as detailed in Step 25

**Go to** File > Open > Syntax

**Double click** on Program 23: 23-pathway-cohort-syntax.sps

**Update (to ensure)** the CD working directory is the same as where the source linkage file is saved

**Ensure** the define for !FY is correct for the 4 digit financial year

**Save** changes (CTRL + S)

**Select** the whole program (CTRL + A)

**Run** the program

**Check** the output and syntax windows regularly whilst the program is running to ensure there are no errors

If there are any errors, **resolve** and **re-run** (if not possible to resolve, **speak** to Andrew Mooney)

**Close** Program 23

**Check** available network space as detailed in Step 25

**Go to** File > Open > Syntax

**Double click** on Program 24: 24-add-pathway-cohort-to-source.sps

**Update** Line 7 **(to ensure)** the CD working directory is the same as where the source linkage file and lookups created by Program 23 are saved

**Ensure** the define for !FY is correct for the 4 digit financial year

**Highlight** from Line 7 to Line 26 and **run**

**Check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Close** Program 24

*Updating SPARRA scores in Source Linkage Files*

SPARRA score is updated after the source linkage file is produced.

Request access to SPARRA through UAS: [https://useraccess.nhsnss.scot.nhs.uk/](https://useraccess.nhsnss.scot.nhs.uk/apex5p/f?p=103:LOGIN_DESKTOP:10741387561472).

Once request is granted, request SPARRA extracted data from Rachael Porteous.

Run the syntax to add the SPARRA score to Source Linkage file -32-SPARRA.

*Checking Death dates data in Source Linkage Files*

After the source episode file is created, the accuracy of death data is checked.

The syntax Death\_Flag is run till line 387.

*Move the newly created Source Linkage File to the area where analysts can access the file from*

**Check** available network space for [\\nssstats01\hscdiip](file:///\\nssstats01\hscdiip) - this is the final saving place (as at December 2017) for the Source Linkage Analysis Files

**Go to** the Start Menu > All Programs > PuTTY (folder) > **Double Click** on PuTTY application

**Log in**

**Navigate** to the network area where the source linkage file is currently saved (use the document [\\Freddy\dept\PHIBCS\PHI\Health & Social Care\Topics\Linkage\Reference Documents\Health and Social Care Team - Using UNIX shell to move files.docx](file:///\\Freddy\dept\PHIBCS\PHI\Health%20&%20Social%20Care\Topics\Linkage\Reference%20Documents\Health%20and%20Social%20Care%20Team%20-%20Using%20UNIX%20shell%20to%20move%20files.docx)

**Using** the same document as listed in Step 84, move the Source Linkage Episode file to its ‘home’ on [\\nssstats01\hscdiip](file:///\\nssstats01\hscdiip) in folder 01-Source-Linkage-Files

*Check the UNIX group the files belong to (do this immediately after step 85)*

**Navigate** to the network area where the source linkage file(s) have been moved to on [\\nssstats01\hscdiip](file:///\\nssstats01\hscdiip) and **type** ls -l **press** return

**Check** the fourth column to ensure that the file belongs to the UNIX group hscdiip

If the group is hscdiip, **go to** Step 91

If the group is not hscdiip, **go to** Step 88

**Check** who owns the file (third column indicates file owner) – the owner of the file is the only person (outside NSS IT) who can change the UNIX group for a file.

If you are the file owner, **go to** Step 89

If you are not the file owner, speak to the file owner and **ask** them to work through Steps 82 to 92 (excluding Step 88)

**Type** chgrp hscdiip filename – e.g chgrp hscdiip source-episode-file-201718.sav – **press** return

**Type** ls -l **press** return and **check** that the UNIX group has been successfully changed

*Change the file permissions to read only for UNIX group members, the owner should have read/write access and the ‘world’ should have no access (similar to changing group, these steps must be carried out by the file owner)*

**Type** chmod 640 filename e.g chmod 640 source-episode-file-201718.sav – **press** return

**Type** exit **press** return

**3.3 Produce the Source Individual level analysis file**

Producing the financial year data for the Source individual analysis file is more straightforward than the creation of the Source episode analysis file. The sources required to produce this file are detailed below.

Source episode analysis file (for the latest financial year)

Geography lookup files (held in [\\nssstats01\irf](file:///\\nssstats01\irf))

Deprivation lookup files (held in [\\nssstats01\irf](file:///\\nssstats01\irf))

The process detailed below describes the steps required for creating a new financial year analysis file with summary information for a patient. If an update is being carried out for an existing financial year, simply work methodically through the existing programs in the existing financial year folder (which is pretty much the process outlined below but missing out Steps 5 and 6).

**Note** that the sections for creating the individual data sets to be brought together for the Source Individual Analysis File are very time consuming to run. It is recommended that this when running this program, start as early as possible.

**Open** Windows Explorer

**Map** [\\nssstats01\irf](file:///\\nssstats01\irf) to a network drive (if this is not already mapped)

**Open** SPSS and **log in** (must be on the nssstats01 server and working in UNICODE mode)

**Go to** File > Open > Syntax and **navigate** to mapped samba share for UNIX irf, to the folder 11-Development team > Dev00-PLICS-files > **Relevant Financial year**

**Double click** on Program 21: 21-create-source-linkage-individual-analysis-file.sps

**Ensure** the defines for all file paths (!file etc) and !FY are correct for file location(s) and 4 digit financial year

**Save** any changes (CTRL + S)

**Highlight** from Line 25 to Line 474 (creates individual files for matching) and **run**

Whilst running, **check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Check** available network space as detailed in Section 3.2.5, Step 25 (part (e) is lines 822-829 for the individual analysis file)

**Highlight** from Line 479 to Line 491 (creates temporary individual level file) and **run**

If there are any errors, **resolve** and **re-run**

**Check** available network space as detailed in Section 3.2.5, Step 25 (part (e) is lines 822-829 for the individual analysis file)

**Highlight** from Line 512 to Line 604 and **run**

If there are any errors, **resolve** and **re-run**

**Check** available network space as detailed in Section 3.2.5, Step 25 (part (e) is lines 822-829 for the individual analysis file)

**Highlight** from Line 609 to Line 718 and **run**

If there are any errors, **resolve** and **re-run**

**Check** the date in Lines 724 to 742 and Lines 745 to 763is the end of the financial year that the file is producing the episode level analysis file for.

**Update** dates (if required)

**Highlight** from Line 724 to Line 770 and **run**

**Check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run** (if re-running, start from line 604)

**Check** available network space as detailed in Section 3.2.5, Step 25 (part (e) is lines 822-829 for the individual analysis file)

**Highlight** from Line 773 to Line 800 and **run**

If there are any errors, **resolve** and **re-run**

**Close** Program 21 (this program will be used again later in this procedure).

**Check** available network space as detailed in Section 3.2.5, Step 25 (part (e) is lines 822-829 for the individual analysis file)

**Go to** File > Open > Syntax

**Double click** on Program 22: 22-locality&clusters.sps

**Ensure** the defines for !file and !FY are correct for file location(s) and 4 digit financial year (Lines 1-15)

**Highlight** from Line 73 to Line 117 and **run**

**Check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Close** Program 22

**Check** available network space as detailed in Section 3.2.5, Step 25 (part (e) is lines 822-829 for the individual analysis file)

**Go to** File > Open > Syntax

**Double click** on Program 24: 24-add-pathway-cohort-to-source.sps

**Update** Line 7 **(to ensure)** the CD working directory is the same as where the source linkage file and lookups created by Program 23 are saved

**Ensure** the define for !FY is correct for the 4 digit financial year

**Highlight** from Line 37 to Line 48 and **run**

**Check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Close** Program 24

**Check** available network space as detailed in Section 3.2.5, Step 25 (part (e) is lines 822-829 for the individual analysis file)

**Go to** File > Open > Syntax

**Double click** on Program 25: 25-calculating-HRIs.sps

**Ensure** the defines for all file paths (!file etc) and !FY are correct for file location(s) and 4 digit financial year

**Update** date on line 112 to be the mid-year of the financial year that the program is producing the file for. For example, for 2017/18, the mid-year date is 20170930.

**Save** any changes (CTRL + S)

**Highlight** from Line 101 to Line 438 and **run**

**Check** the output and syntax windows to ensure there are no errors

Whilst running, **check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Check** available network space as detailed in Section 3.2.5, Step 25 (part (e) is lines 822-829 for the individual analysis file)

**Highlight** from Line 438 to Line 449 and **run**

**Check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Close** Program 25

**Check** available network space as detailed in Section 3.2.5, Step 25 (part (e) is lines 822-829 for the individual analysis file)

**Go to** File > Open > Syntax

**Double click** on Program 21: 21-create-source-linkage-individual-analysis-file.sps

**Highlight** from Line 25 to Line 47 and **run**

If there are any errors, **resolve** and **re-run**

**Check** available network space as detailed in Section 3.2.5, Step 25 (part (e) is lines 822-829 for the individual analysis file)

**Highlight** from Line 802 to Line 818 and **run**

If there are any errors, **resolve** and **re-run**

**Highlight** from Line 822 to Line 829 and **run**

**Check** available network space as detailed in Section 3.2.5, Step 25 (part (e) is lines 822-829 for the individual analysis file)

**Close** Program 21

**Go to** File > Open > Syntax

**Double click** on Program 30: 30-hri-simd.sps

**Highlight** from Line 25 to Line 47 and **run** **NEED TO LOOK AT THE PROGRAM TO CHECK WHAT IS ACTUALLY REQUIRED TO BE RUN HERE.**

Whilst running, **check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Close** Program 30.

**Go to** File > Open > Syntax

**Double click** on Program 31: consistency

Run the syntax

Whilst running, **check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Close** Program 31.

**Double click** on Program 32: SPARRA

Run the syntax for source individual (relevant year)

Whilst running, **check** the output and syntax windows to ensure there are no errors

If there are any errors, **resolve** and **re-run**

**Close** Program 32.

After the source individual file is created, the accuracy of death data is checked. The syntax Death\_Flag is run from line 390-410.

**Check** available network space for [\\nssstats01\hscdiip](file:///\\nssstats01\hscdiip) - this is the final saving place (as at December 2017) for the Source Linkage Analysis Files

**Go to** the Start Menu > All Programs > PuTTY (folder) > **Double Click** on PuTTY application

**Log in**

**Navigate** to the network area where the source linkage file is currently saved (use the document [\\Freddy\dept\PHIBCS\PHI\Health & Social Care\Topics\Linkage\Reference Documents\Health and Social Care Team - Using UNIX shell to move files.docx](file:///\\Freddy\dept\PHIBCS\PHI\Health%20&%20Social%20Care\Topics\Linkage\Reference%20Documents\Health%20and%20Social%20Care%20Team%20-%20Using%20UNIX%20shell%20to%20move%20files.docx)

**Using** the same document as listed in Step 47, move the Source Linkage Episode file to its ‘home’ on [\\nssstats01\hscdiip](file:///\\nssstats01\hscdiip) in folder 01-Source-Linkage-Files

**Navigate** to the network area where the source linkage file(s) have been moved to on [\\nssstats01\hscdiip](file:///\\nssstats01\hscdiip) and **type** ls -l **press** return

**Check** the fourth column to ensure that the file belongs to the UNIX group hscdiip

If the group is hscdiip, **go to** Step 91

If the group is not hscdiip, **go to** Step 88

**Check** who owns of the file (third column indicates file owner) – the owner of the file is the only person (outside NSS IT) who can change the UNIX group for a file.

If you are the file owner, **go to** Step 89

If you are not the file owner, speak to the file owner and **ask** them to work through Steps 82 to 92 (excluding Step 88)

**Type** chgrp hscdiip filename – e.g chgrp hscdiip source-individual-file-201718.sav – **press** return

**Type** ls -l **press** return and **check** that the UNIX group has been successfully changed

**Type** chmod 640 filename e.g chmod 640 source-individual-file-201718.sav – **press** return

**Type** exit **press** return

Update Source linkage Individual and Episode in Excel with any new variable that has been added to the source linkage files.

**3.5 Communicating Updates**

Once the two analysis files have been created for the financial year(s), an email should be sent to H&SC team members and LIST advising that updated files are available. Documentation also needs to be updated. Work through the following steps.

Open [\\Freddy\DEPT\PHIBCS\PHI\Health & Social Care\Topics\Linkage\Reference Documents\Schedule for updating Source Linkage Files.docx](file:///\\Freddy\DEPT\PHIBCS\PHI\Health%20&%20Social%20Care\Topics\Linkage\Reference%20Documents\Schedule%20for%20updating%20Source%20Linkage%20Files.docx)

**Update** what data are available on page 4 of this document

**Click** Save

Open **MS Outlook**

**Open** a New Mail Message

In the To line**, add:** Current H&SC team members and LIST

In the Subject line, **enter:** Source Linkage Files update

Body of the email, **enter** (amend as appropriate):

Dear all,

Source linkage files for **INSERT** YEAR have been updated/created and are now available for analysis. Page 4 in the document: [\\Freddy\DEPT\PHIBCS\PHI\Health & Social Care\Topics\Linkage\Reference Documents\Schedule for updating Source Linkage Files.docx](file:///\\Freddy\DEPT\PHIBCS\PHI\Health%20&%20Social%20Care\Topics\Linkage\Reference%20Documents\Schedule%20for%20updating%20Source%20Linkage%20Files.docx) advises what data are available.

Thanks,

**Enter** **name**.

**Click** Send

**3.6 Recording Updates**

Update source linkage Individual and Episode files in Excel with any new variables that has been added to the source linkage files.

**Declaration**

*AUTHORISATION*

As author of this SOP, I declare that the procedures outlined above accurately reflect the procedures used and that this document if fit for purpose.

Please see version history for the version number of this SOP.

Denise K Greig (nee Hastie)

Senior Statistician

Date: January 2018