HSC Business Services Organisation

QUEEN'S UNIVERSITY BELFAST

PATHWAYS TO A CANCER DIAGNOSIS:

Monitoring variation in the patient journey across Northern Ireland 2012 – 2016

Technical Document

1. Introduction

This document provides a summary of the data sources and methodology used for the first 'Pathways to Cancer Diagnosis in Northern Ireland' project. Extracted data covers tumours diagnosed in 2009 to 2016.

The origin of Pathway to Cancer Diagnosis Analysis

The National Cancer Registration and Analysis Service (Public Health England) have produced an award winning Routes to Diagnosis study which defined a methodology to determine the route a patient took through the healthcare system before receiving a cancer diagnosis. It is from this that the Pathways to Cancer Diagnosis project in Northern Ireland has taken its lead.

Routes to diagnosis are an important and often modifiable prognostic indicator in cancer care. Patients who present as an emergency have poorer one year survival than those who present via the 'red flag' route or via screening.

Quantifying incidence by route or pathway to diagnosis data has potential to highlight variation across a range of different variables such as age, deprivation, health geography.

The 'Pathways to Cancer Diagnosis in Northern Ireland' project adapts the Public Health England methodology to provide the first analysis of this type for cancer patients in Northern Ireland.

2. Overview of the Routes to Diagnosis project

2.1 Project goals

The aim of this project is to produce, for the first time, measures on pathways to diagnoses in Northern Ireland for all cancers and by individual cancer site, as well as by Trust of residence, urban/rural and deprivation category. The diagnosis pathways used are those categorised by the National Cancer Registration and Analysis Service (Public Health England).

The study examines first whether it is feasible to apply the Public Health England Routes to Diagnosis methodology and algorithm to the Northern Ireland setting.

On the basis that this is feasible, for at least some years of data, the project looks to identify whether there is an association between pathways to diagnosis and survival for cancer patients. Also, do age, sex, deprivation, cancer stage and geographical area of residence on diagnosis have any influence on routes?

The study uses funnel plots to explore the variation in pathways to diagnosis in order to aid interpretation and identify areas for follow-up analysis and subsequent development of intervention strategies.

2.2 Technical overview

Data from the Northern Ireland Cancer Registry is combined with Patient Administration System data for Inpatient and Outpatient activity and Cancer Patient Pathway System data. Using these datasets cases of cancer registered in Northern Ireland diagnosed in 2009 to 2016 are categorised into one of 8 principle 'Routes to Diagnosis' as identified by the National Cancer Registration and Analysis Service (Public Health England) Routes to diagnosis study.

2.3 Policy Context

The project forms part of the Health Foundation's Applying Advanced Analytics Project.

The Health Foundation is an independent charity committed to bringing about better health and health care for people in the UK. Under its 'Applying Advanced Analytics Project', the charity is funding pilot projects, to be carried out from within the UK health-service, to build up analytical capacity to drive monitoring and improvement of service delivery.

The Pathways to Cancer project proposal was submitted by BSO in partnership with the Centre for Public Health in Queens's University of Belfast and was supported by the Northern Ireland Cancer Registry and the Director of Commissioning for the Health and Social Care Board.

3. Methods

The methodology used to produce the pathways to cancer diagnosis for Northern Ireland is detailed in the sections below.

The original algorithm which assigns a route to diagnosis was produced by the National Cancer Registration and Analysis Service (Public Health England). An overview of the original Routes to Diagnosis algorithm is set out within the PHE technical documentation.

> "The algorithm takes as a starting point the date of cancer diagnosis, as defined by the UK and Ireland Association of Cancer Registries (UKIACR) using European Network of Cancer Registries (ENCR) rules. Routine data immediately prior to this date are examined and a series of rules is used to classify the 'Route to Diagnosis' for each case. The routes are categorised in detail by 3 variables: the end-point, the pathway group, and the start-point. These detailed routes have been aggregated into 8 broader categories to facilitate analysis.
>
> It is important to note that patient records being used to describe the route to diagnosis may not have a cancer code assigned to them, as the episodes and attendances will have taken place before a cancer diagnosis has been coded. It is therefore not possible to be absolutely certain that the episodes and attendances related to the patient prior to diagnosis were directly related to the process of diagnosis of cancer. However, the frequency of hospital attendance and admission in the period immediately before diagnosis greatly exceeds the 'background' rate making the assumption that they are related to the cancer diagnosis reasonable."
>
> **Source: http://www.ncin.org.uk/publications/routes_to_diagnosis**

This PHE methodology has been followed as closely as possible to allow for potential comparisons of data across regions. However, a number of variations have been introduced in order to best fit Northern Ireland's different recording systems.

As with the PHE hospital data extracts, hospital information used to assign routes cannot be matched to cancer diagnoses from the NICR with absolute certainty as, having taken place before diagnosis, they may not have a cancer code assigned to them. As with the PHE project, the decision has been taken that the proximity of the hospital attendance or admission to the cancer diagnosis makes it reasonable to assume that they are related.

4. Data Sources

4.1 Northern Ireland Cancer Registry (NICR)

The NICR is a population based cancer registry collecting data on all malignant and certain non-malignant tumours diagnosed in Northern Ireland.

For this project, all cancer registrations across Northern Ireland between 2009 and 2016 inclusive, with ICD-10 diagnosis codes **C00–C97, D00-D09, D13, D15, D27, D29, D32, D33, D35** and **D37-D48** (all neoplasms) were extracted from the NICR.

A subset of this data for tumours with ICD-10 diagnosis codes **C00-C97** excluding **C44** which were diagnosed in calendar years 2009 to 2016 was used for reporting. A number of other exclusions were also made from the NICR extract following validations and based on experience of the PHE research. These are detailed below:

- all D codes with the exception of **D05, D06, D090, D32, D330-D332, D333, D334, D352-D354, D42, D430-D432, D433-D434, D437-D439** and **D443-D445** were excluded from the reporting dataset
- the records for patients with non-melanoma skin cancer
- records where the sex code and cancer site code are incompatible
- records with invalid ages
- duplicate records identified
- records with invalid ICD10 site codes

Routes were derived for all tumours fitting the criteria specified above. This includes incidences where more than one tumour is recorded in the same person.

4.2 Patient Administration System (PAS)

Patient Administration Systems are principally used to manage and record inpatient, day case and outpatient activity within Health and Social Care hospital sites in Northern Ireland.

4.2.1 Hospital Inpatient System (HIS)

The Hospital Inpatient System (HIS) is formed in the HSC Data Warehouse from PAS data and provides information on admitted patient care delivered by health and social care hospitals in Northern Ireland. It is a patient level administrative data source and each record relates to an individual consultant episode.

For the pathways to diagnosis project, an extract containing 2008/09 to 2015/16 records was used to identify patients from the NICR with a hospital admission for any cause during the six months prior to their cancer diagnosis.

4.2.2 Outpatient Universe

The Outpatient Universe is formed in the HSC Data Warehouse from PAS data and provides information on outpatient appointments at Health and Social Care (HSC) hospitals in Northern Ireland. It is a patient level administrative data source and each record relates to an appointment. This can include a change in appointment details.

For the pathways to diagnosis project an extract containing 2008/09 to 2015/16 records was used to identify patients from the NICR with an outpatient attendance for any reason during the six months prior to their cancer diagnosis.

4.3 Cancer Patient Pathway System (CaPPS)

The Cancer Patient Pathway System (CaPPS) is a bespoke data system used to administer cancer treatment services within Health and Social Care (HSC) Trusts in Northern Ireland. This dataset contains patient level information and is used to monitor and report on the number of patients treated for cancer following a decision to treat being taken.

For the pathways to diagnosis project an extract containing 2008/09 to 2015/16 records was used to identify patients from the NICR who received an urgent GP referral for suspect cancer.

5. Data Preparation

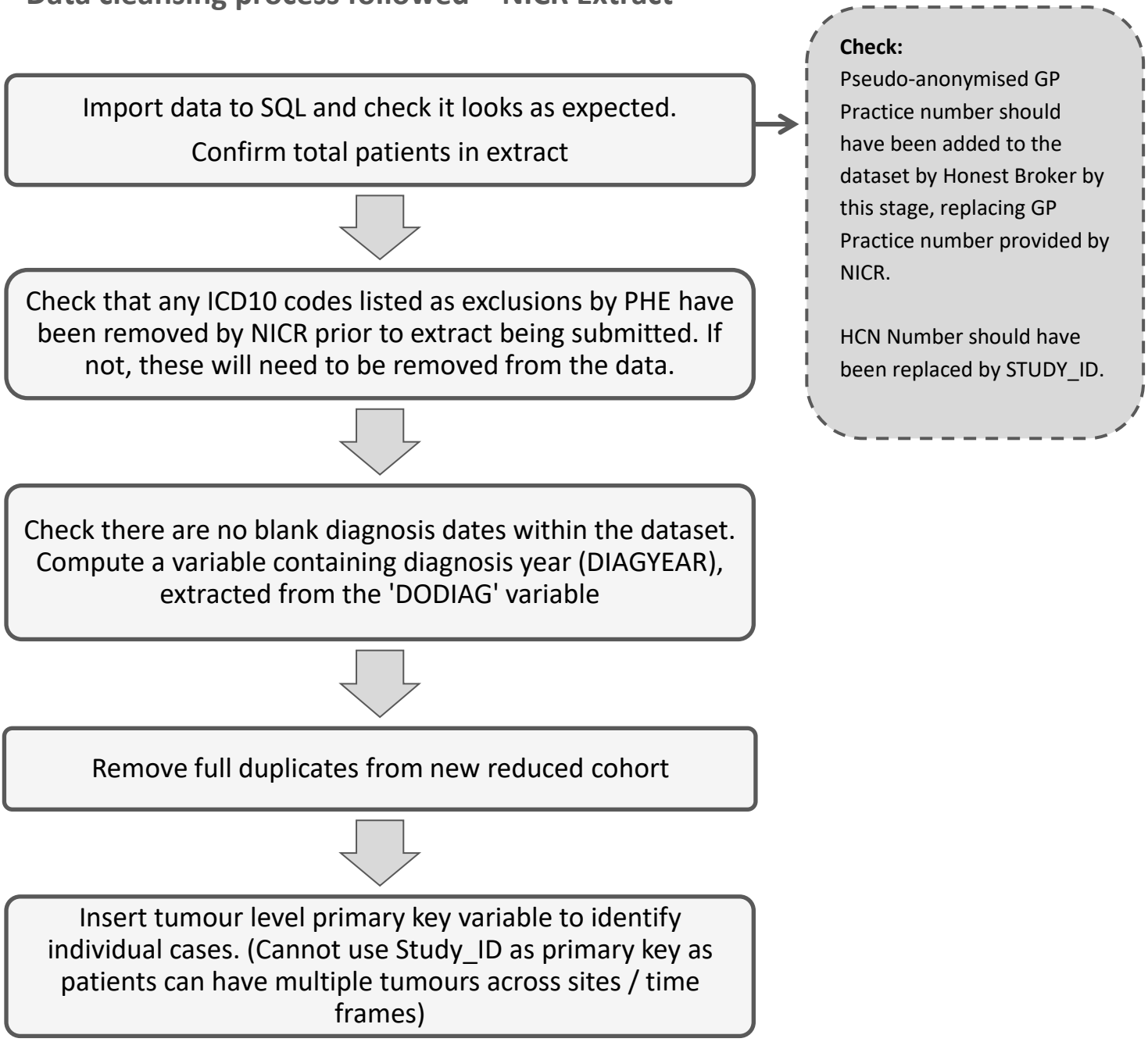
5.1 Northern Ireland Cancer Registry (NICR)

Variables requested from the Data Warehouse were as follows –

HCN – Health and Care Number, replaced by honest broker with **STUDY_ID**
SEX – Sex of patient
AGE – Age of patient
DEPRIV2010 – deprivation quintile, based on patient postcode (2010 values)
URBAN_RURAL – urban/rural indicator, based on patient postcode
LGD2014 – Local Government District, based on patient postcode
TRUST – HSC Trust of residence of the patient
GP_Practice – replaced by **GP_ANON**. Produced by Honest Broker, anonymised GP practice of the patient at time of diagnosis.
DODIAG – Date of diagnosis of tumour
SCREEN – Screen detected flag placed on data by cancer registry
BASIS – Clinical variable relating to tumour
SITE_ICD10 – Site of tumour with 4 character ICD10 code
SITECLASS – Aggregated tumour site grouping
MORPH – Clinical variable relating to tumour
GRADE – Clinical variable relating to tumour
STAGE – Clinical variable relating to tumour
DODEATH – Date of death of patient
CAUSED – Flag to identify patient cause of death (*Cancer / Other*) or if patient is alive (*Alive*)

The above information was extracted for all patients appearing on the Northern Ireland Cancer Registry between the years 2009 and 2016.

Data cleansing process followed – NICR Extract



Preparing the Northern Ireland Cancer Registry (NICR) extract

The NICR extract was submitted to the Honest Broker Service (HBS) where the data was anonymised with each HCN being replaced with a variable which will enable linkage across extracts (**STUDY_ID**) before being made available for use.

A variable was added to identify the year of diagnosis (**DIAG_YEAR**) for each tumour. The pathways have been calculated on individual years of data so this variable was used to filter the data accordingly.

The original data extract included some codes which had been excluded by the PHE and so had to be removed prior to analysis (list can be found in the NICR section above as well as in the SQL code). However in the updated extract these codes were removed by the NICR. It would be good practice in future to check that this have been done in any future extracts, for that reason this portion of code on has purposely been left in.

A check was completed to confirm there were no missing values against the date of diagnosis (any such cases would need removed) and that the number of cases across each year was broadly consistent.

All full duplicates were removed from the file. There were a small number of cases where the only difference was the marital status variable. Since this variable was not of interest it was removed and the remaining duplicate cases deleted. In future extracts there is no need to ask for marital status information and as such have removed it from the list of extract variables above.

Finally a primary key variable (**TUMOUR_ID**) was added for analysis.

The file has been saved as **'HONESTBROKER.E024.NICR_FINAL'**

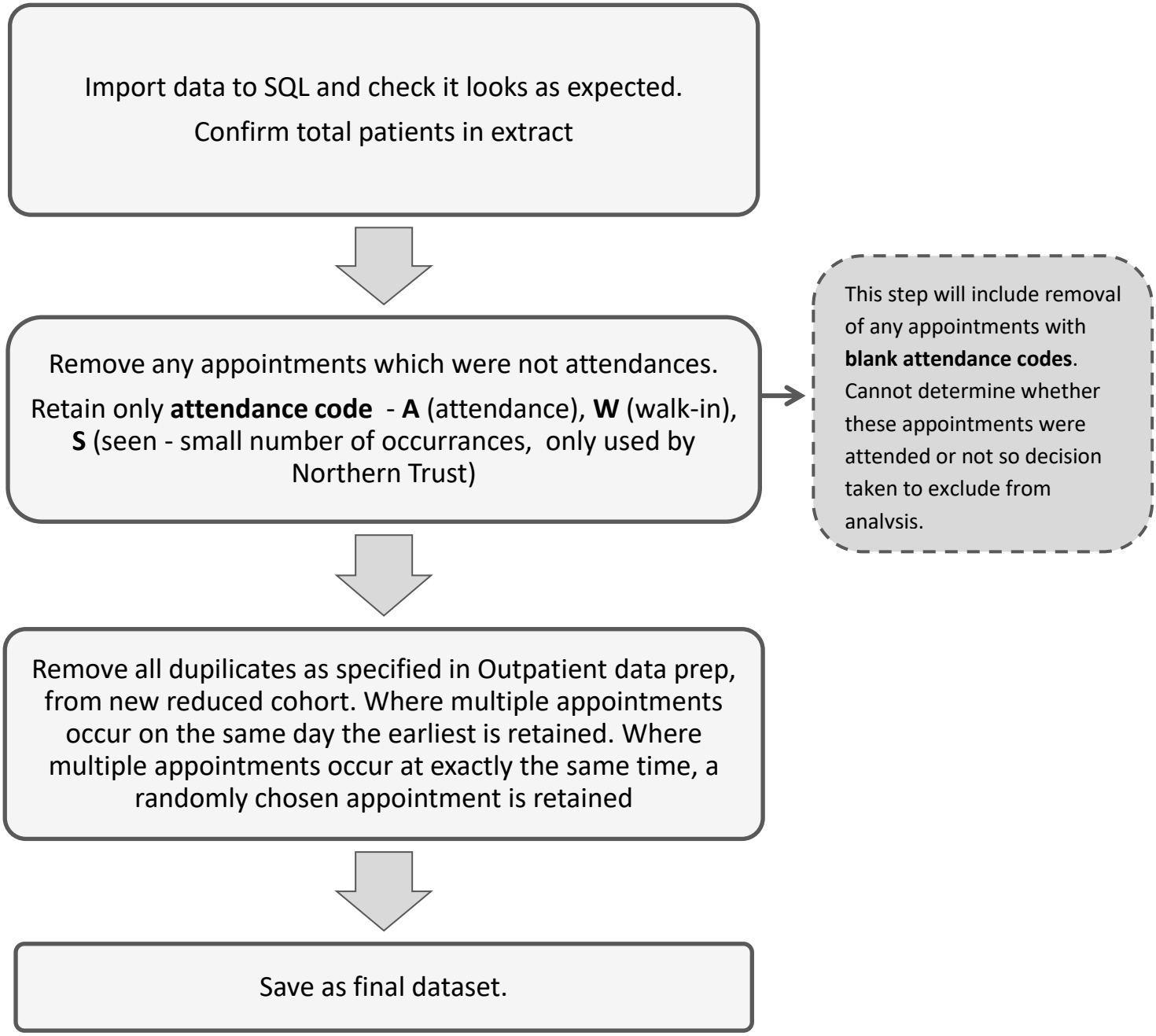
5.2 Outpatient Universe

Variables requested from the Data Warehouse were as follows –

HCN – Health and Care Number, replaced by Honest Broker with **STUDY_ID**
Appointment_date – date and time of appointment
Attendance_code – denotes whether the appointment took place or not
Attendance_desc – descriptor variable for attendance code
Appointment_type_r – regional variable classifying appointment as first or review
Referral_code – local code identifying the source of the referral
Referral_desc – local descriptor variable for source of referral code
CMDS_value – closest matching value to the PHE referral source mapping

The above information was extracted for all patients appearing on the Northern Ireland Cancer Registry between the years 2009 and 2016. Outpatient appointments were obtained for the years 2008-2016 to enable identification of appointments attended during the 6 month period prior to cancer diagnosis.

Data cleansing process followed – Outpatient Extract



Preparing the Outpatient data extract

The outpatient extract was submitted to the Honest Broker Service (HBS) where the data was anonymised with each HCN being replaced with a variable which will enable linkage across extracts (**STUDY_ID**) before being made available for use.

Unnecessary variables were removed, for the next extract we could exclude these from the DAA.

Only appointments that were actually attended were needed for analysis so I have removed any that were cancelled or missed. I also removed any appointments where the attendance code was blank. This left only appointments that were attended (A), seen (S) or walk-ins (W).

I have then looked at the new reduced cohort and carried out a validation of the referral source variable (CMDS Value) using the local descriptions. Any regional codes that looked to be incorrect I have recoded in line with their local code. I have also recoded all action cancer cases to be picked up as screen detected. This recoding will need done manually again with any new extracts.

Using the recoded referral information, I have checked the file for full duplicates and have removed any such cases.

There are a number of other types of duplicates that need to be identified and removed from the file. To help with identifying them I have created an extra variable which only has the date of the appointment and drops the time element (**APP_DATE_ONLY**). This variable will also be used later when calculating proximity of outpatient activity to cancer diagnosis in the NICR extract.

Where there appear to be multiple appointments attended on the same day at exactly the same time but with different referral sources I have selected a random one to retain.

Where a patient attended more than one appointment identical appointment on the same day I have retained the earliest appointment in the day.

Where a patient attended more than one appointment on the same day with different referral sources I have retained the earliest appointment in the day.

Where a patient attended more than one appointment on the same day with different referral sources and different attendances codes I have retained the earliest appointment in the day.

Where a patient attended more than one appointment on the same day at the same time but with different referral sources and different attendances codes I have randomly selected an appointment to retain.

The file has been saved as **'HONESTBROKER.E024.OUTPATIENT_FINAL_FILE'**

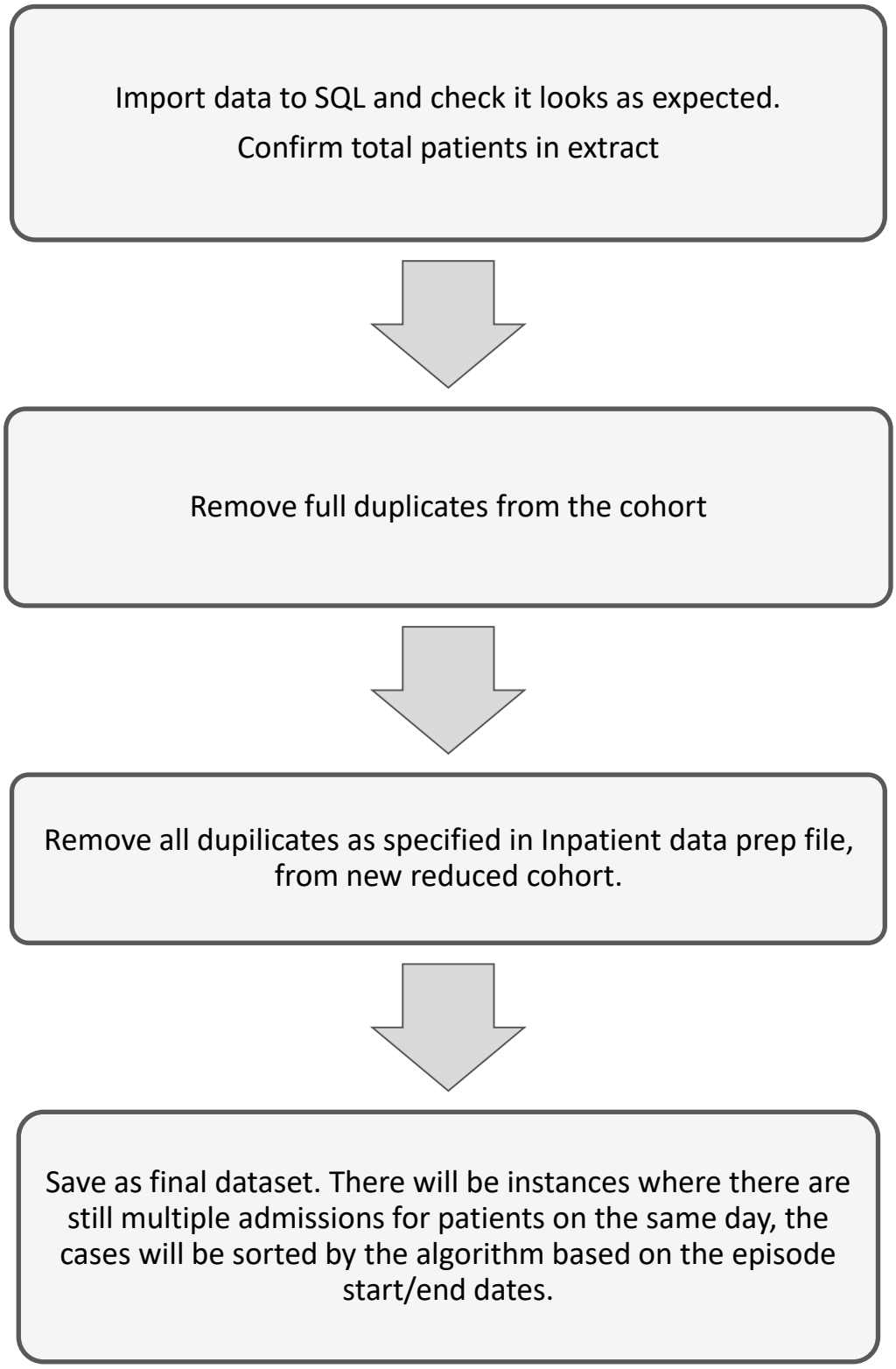
5.3 Inpatient Universe

Variables requested from the Data Warehouse were as follows –

HCN – Health and Care Number, replaced by honest broker with **STUDY_ID**
EPS_START_DATE – This field contains the episode start date (without time) on which a patient was under the continuous care of one consultant. For each new episode, which may be due to transfer to another consultant or hospital, there will be a new episode start date.
EPS_END_DATE - This field contains the episode end date (without time) on which a patient finished under the continuous care of one consultant. It may be that the patient was discharged, but also may be due to transfer to another consultant or hospital, for which there will be a new episode start date.
EPS_ORDER - This field contains the order of the episode within the hospital admission.
ADMIT_DATE - Date of admission with time of the first or only episode within a hospital stay. ADMIT_DATE is recorded on all episodes within a hospital admission.
ADMIT_METHOD_CODE - Method of Admission. This is recorded on the first and all subsequent episodes within the admission.

The above information was extracted for all patients appearing on the Northern Ireland Cancer Registry between the years 2009 and 2016. Inpatient admissions were obtained for the years 2008-2016 to enable identification of admissions occurring during the 6 month period prior to cancer diagnosis.

Data cleansing process followed – Inpatient Extract



Preparing the Inpatient data extract

The Inpatient extract was submitted to the Honest Broker Service (HBS) where the data was anonymised with each HCN being replaced with a variable which will enable linkage across extracts (**STUDY_ID**) before being made available for use.

All full duplicates were removed from the file.

Local codes for admission method had been requested in the original extract but they weren't needed for the analysis, regional coding was used which is more consistent and matches the categories used by PHE. Local codes were deleted and any duplicates that were created by doing this were then removed.

There were a number of cases where all information was identical apart from the admission method. In other words it looked as though the patient was admitted via two different methods at exactly the same date/time. In these cases only one record was retained and this was selected at random.

The file has been saved as **'HONESTBROKER.E024.INPATIENT_FINAL_FILE'**

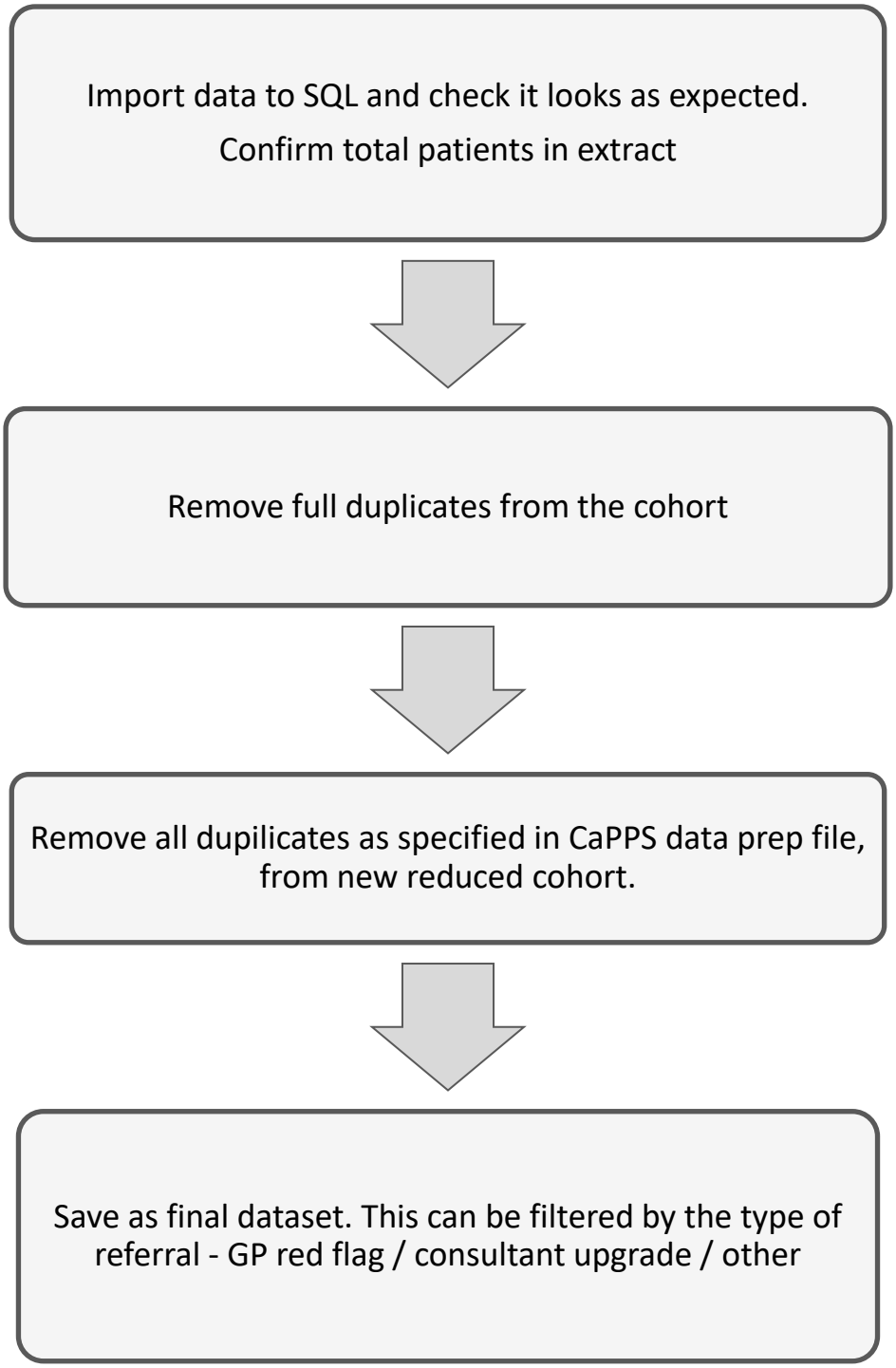
5.4 Cancer Patient Pathway System (CaPPS)

Variables requested from the Data Warehouse were as follows –

HCN – Health and Care Number, replaced by honest broker with **STUDY_ID**
DATEDECISIONTOTREAT – Date of decision to treat
TUMOURSITE – Suspected tumour site code contained in the referral
TUMOURSITE_DESC – Suspected tumour site description contained in the referral
CONFIRMED TUMOUR_SITE – Confirmed tumour site (may differ from suspect)
PRIMARY_REF_TYPE_DESC - Description for the primary referral type, grouped into regional categories

The above information was extracted for all patients appearing on the Northern Ireland Cancer Registry between the years 2009 and 2016. CaPPS records were obtained for the years 2008-2016.

Data cleansing process followed – CaPPS Extract



Preparing the CaPPS data extract

The CaPPS extract was submitted to the Honest Broker Service (HBS) where the data was anonymised with each HCN being replaced with a variable which will enable linkage across extracts (**STUDY_ID**) before being made available for use.

Any unnecessary variables were removed, for future extract these should be excluded from the DAA.

All full duplicates within the extract were removed to leave only single instances of each.

Where there is only a difference in tumour site these duplicates were removed to leave only one instance. The tumour site is not used for the linkage so a random case has been retained.

The file has been saved as **'HONESTBROKER.E024.CAPPS_FINAL3'**

6. Calculating the Pathways to Cancer

6.1 Data Matching

All records have been allocated a STUDY_ID by the Honest Broker Service to enable linkage to take place. This anonymised 'STUDY_ID' variable was added to all datasets based on the HCN from the original extracts. HCN details have been removed prior to their release by the Honest Broker Service.

Records from the NICR, Inpatient and Outpatient datasets have been matched based on STUDY_ID and their proximity to each other.

Records from NICR and CaPPS have been matched on STUDY_ID. Only CaPPS records with a valid decision to treat date have been matched; where the date of diagnosis lies between 62 days before and 31 days after the decision to treat date.

6.2 Producing the Pathway to Diagnosis

The Pathways to Diagnosis algorithm has been constructed to replicate, as closely as possible, the methodology of the PHE Routes to Diagnosis algorithm. The algorithm assigns a three part code to each tumour within the NICR based on available inpatient and outpatient hospital data. The three part codes are then mapped into broad categories for analysis purposes. Finally, the presence of screening or CaPPS information is checked for. This data will take precedence over assigned route.

Calculating the end point of the pathway to cancer diagnosis

End points have been identified for each tumour based on the proximity of the patient's inpatient and outpatient hospital interactions to the date of diagnosis. The end point is assumed to be the clinical care event that led most immediately to diagnosis. The following rules have been applied, in line with Public Health England.

- Where both inpatient and outpatient activity occurred on the date of diagnosis the inpatient episode was defined as the end-point of the route.
- Otherwise, if there was an episode within 28 days prior to the date of diagnosis then this was assigned as the end-point of the route, with inpatient episodes taking precedence over outpatient episodes and the most recent episode taking precedence if there were multiple episodes.
- If there was no hospital activity within 28 days of diagnosis then the most recent episode within 6 months (inpatient or outpatient) was used as the end-point of the route.

Based on these rules, the one of five end-point codes were assigned:

Special cases (SC) – patients with a cancer diagnosis date on the same day as an inpatient admission date and an outpatient attendance date, or whose closest hospital interactions to diagnosis are an inpatient and outpatient record occurring on the same date. These are a special case of inpatient diagnosis.

Inpatient diagnosis (IP) – patients with a cancer diagnosis date related to a preceding inpatient episode (excluding patients already defined as special cases). An inpatient diagnosis is defined where the cancer diagnosis date is within the start and end of an episode. In addition, due to the potential for diagnosis to be confirmed following a relevant inpatient episode, a cancer diagnosis date that is within 6 months after the end of an episode and with no outpatient activity between would also be regarded as an inpatient diagnosis.

Outpatient diagnosis (OP) – patients with no inpatient episodes preceding the cancer diagnosis date (as defined above) but with an outpatient attendance preceding the cancer diagnosis date or with an inpatient elective admission, or were emergencies via an outpatient clinic, or were unresolved inpatient transfers.

Unknown (UN) – Unable to match cancer diagnosis date to any informative inpatient episodes or outpatient attendances within the valid timeframe. It is likely that, for these patients, the cancer diagnosis date was obtained from pathology records only, indicating diagnosis or treatment that only took place outside of a hospital setting (eg NHS patients seen in primary care, independent treatment centres or a community setting, and private patients seen and treated only in private hospitals).

Death Certificate Only diagnosis (DCO) - The cancer registry receives a small number of cancer related death notifications, for which they are unable to obtain additional information to register the disease details fully. This registration is regarded as Death Certificate Only (DCO) and the date of diagnosis is the same as that of the date of death.

Calculating the start point of the pathway to cancer diagnosis

The start point has been calculated by working backwards from the end point of each pathway. The start-points have been categorised based on the following rules, in line with Public Health England.

- Routes that originated in an outpatient attendance use the outpatient source of referral of that attendance as the 'start-point' code.
- Routes that originated in an inpatient episode use the inpatient method of admission as the 'start-point' code.
- Routes where inpatient or outpatient data were unavailable the start-point codes may be assigned as null or unknown (this also includes DCOs).

The start point will determine the initial source of the pathway – ie the outpatient referral source or inpatient admission method of the initial appointment. A breakdown of all possible start codes is provided in the following table.

Available inpatient admission methods
Not known (0, 99)
Elective – Waiting List (11)
Elective – Booked (12)
Elective – Planned (13)
Emergency – A&E Department in the same Board (21)
Emergency – General Practitioner (22)
Emergency – Bed Bureau / Central Bed Bureau (23)
Emergency – Consultant Outpatient Clinic (24)
Emergency – Domiciliary Visit by Consultant (25)
Emergency – Other means (28)
Maternity – Ante Partum (31)
Maternity – Post Partum (32)
Other – Patients from another hospital (81)
Other – Babies born in the hospital (82)
Other – Baby born outside the provider (83)

Available outpatient referral sources
Initiated by consultant responsible for the appointment
Consultant – Following emergency admission (1)
Consultant – Following a domiciliary visit (2)
Consultant - Following A&E Attendance (10)
Consultant – Other (11)
Not initiated by consultant responsible for the appointment
Referral from General Practitioner (GP) (3)
From an A&E Department (4)
From consultant other than in A&E Department (5)
Self-referral (6)
Referral from prosthetist (7)
Referral from GP with special interest (12)
Referral from Specialist Nurse (13)
Referral from allied health professional (14)
Referral from optometrist (15)
Referral from orthoptist (16)
Referral from a national screening programme (17)
Referral from general dental practitioner (92)
Referral from community dental service (93)
Other source of referral (8)
Other (97)
Not known (99)

Pathway grouping of the cancer diagnosis

Each tumour has been assigned one of the following pathway group codes based on the available inpatient and outpatient data.

Pathway Group	Description
A	Inpatient hospital interaction only within 6 months prior to diagnosis
B	Outpatient hospital interaction only within 6 months prior to diagnosis
C	Special case, an inpatient elective or Emergency via outpatient clinic, and there is outpatient interaction within 6 months prior to diagnosis
D	There are no hospital data 6 months prior to diagnosis
E	No hospital data at all prior to diagnosis

Assigning the final route and category for each cancer diagnosis

For each record in the Northern Ireland Cancer Registry, the route end-point, the pathway group and the route start-point are concatenated to produce a final route code. These codes are then aggregated up into eight broad categories. These categories have been classified in line with PHE Routes to Diagnosis. Detail on the categories and how they are mapped is included below.

GP Referral: includes routine and urgent referrals where the patient was not flagged as a suspect cancer referral

Urgent GP Red Flag Referral: urgent GP referrals with a suspicion of cancer (Reported as 'Two Week Wait' by

Emergency Presentation: an emergency route via A&E, emergency GP referral, emergency transfer, emergency admission or attendance

Other Outpatient: an elective route starting with an outpatient appointment that is either a self-referral, consultant to consultant referral, other or unknown referral (these referrals would not include patients originally referred under the Two Week Wait referral route)

Screen Detected: flagged by the cancer registry as detected via the breast or cervical screening programmes

Inpatient Elective: where no earlier information can be found prior to admission from a waiting list, booked or planned

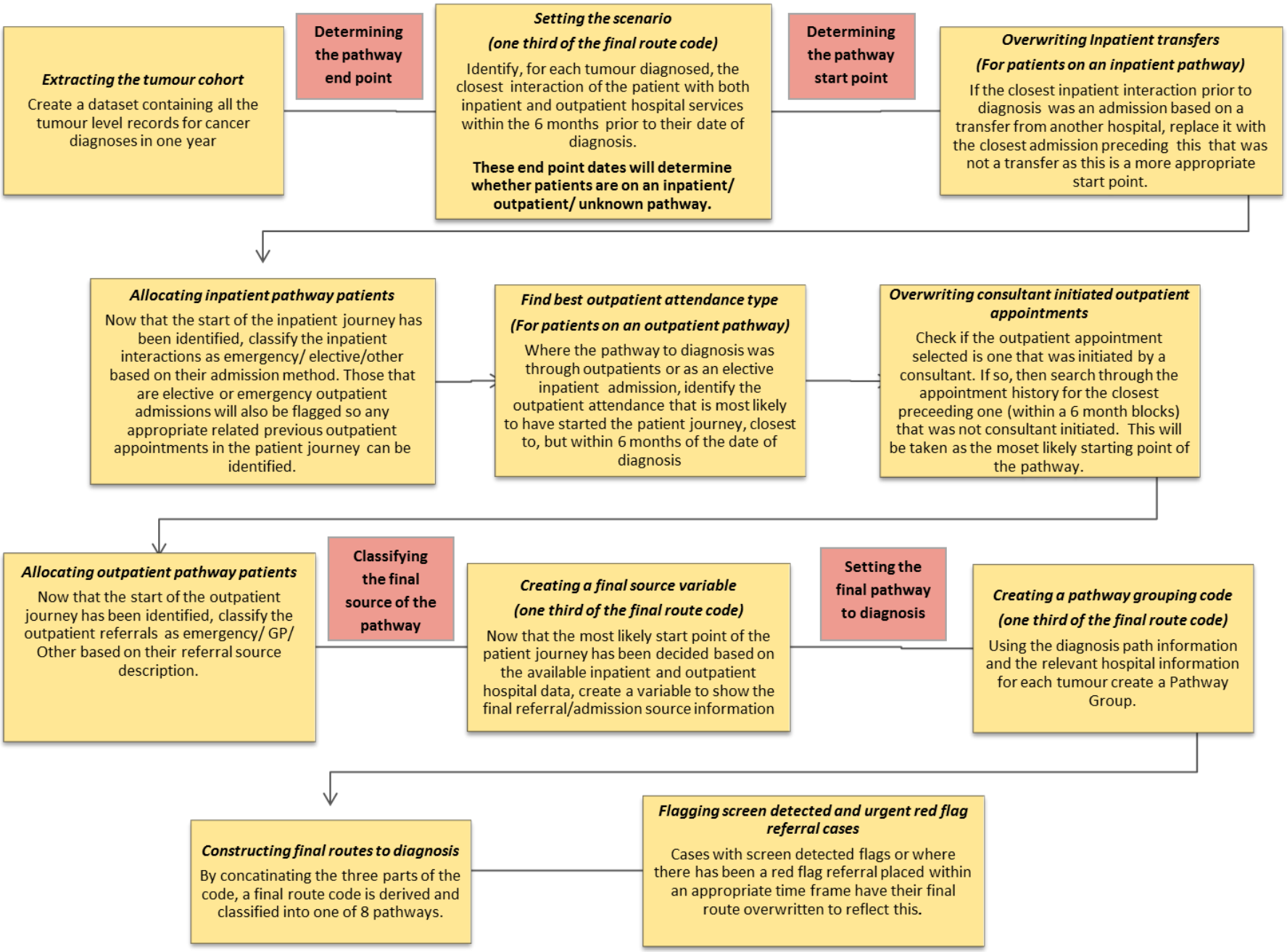
DCO: diagnosis by death certificate only

Unknown: no relevant data available from hospital inpatient or outpatient records, CaPPS or screening information on the Cancer Registry.

Final Route to Category Mapping

No.	Route code	Route group	No.	Route code	Route group
1	DC-D-DCO	DCO	**43**	OP-B-O3	GP referral
2	DC-E-DCO	DCO	**44**	OP-B-O4	Emergency presentation
3	IP-A-I0	Unknown	**45**	OP-B-O5	Other outpatient
4	IP-A-I11	Inpatient Elective	**46**	OP-B-O6	Other outpatient
5	IP-A-I12	Inpatient Elective	**47**	OP-B-O7	Other outpatient
6	IP-A-I13	Inpatient Elective	**48**	OP-B-O8	Other outpatient
7	IP-A-I21	Emergency presentation	**49**	OP-B-O10	Emergency presentation
8	IP-A-I22	Emergency presentation	**50**	OP-B-O11	Other outpatient
9	IP-A-I23	Emergency presentation	**51**	OP-B-O12	GP referral
10	IP-A-I24	Emergency presentation	**52**	OP-B-O13	Other outpatient
11	IP-A-I25	Emergency presentation	**53**	OP-B-O14	Other outpatient
12	IP-A-I28	Emergency	**54**	OP-B-O15	Other outpatient

		presentation			
13	IP-A-I31	Inpatient Elective	**55**	OP-B-O16	Other outpatient
14	IP-A-I32	Inpatient Elective	**56**	OP-B-O17	Screening
15	IP-A-I81	Inpatient Elective	**57**	OP-B-O92	Other outpatient
16	IP-A-I82	Inpatient Elective	**58**	OP-B-O93	Other outpatient
17	IP-A-I83	Inpatient Elective	**59**	OP-B-O97	Other outpatient
18	IP-A-I99	Unknown	**60**	OP-B-O99	Unknown
19	IP-C-O0	Unknown	**61**	SC-C-null	Unknown
20	IP-C-O1	Emergency presentation	**62**	SC-C-O1	Emergency presentation
21	IP-C-O2	Other outpatient	**63**	SC-C-O2	Other outpatient
22	IP-C-O3	GP referral	**64**	SC-C-O3	GP referral
23	IP-C-O4	Emergency presentation	**65**	SC-C-O4	Emergency presentation
24	IP-C-O5	Other outpatient	**66**	SC-C-O5	Other outpatient
25	IP-C-O6	Other outpatient	**67**	SC-C-O6	Other outpatient
26	IP-C-O7	Other outpatient	**68**	SC-C-O7	Other outpatient
27	IP-C-O8	Other outpatient	**69**	SC-C-O8	Other outpatient
28	IP-C-O10	Emergency presentation	**70**	SC-C-O10	Emergency presentation
29	IP-C-O11	Other outpatient	**71**	SC-C-O11	Other outpatient
30	IP-C-O12	GP referral	**72**	SC-C-O12	GP referral
31	IP-C-O13	Other outpatient	**73**	SC-C-O13	Other outpatient
32	IP-C-O14	Other outpatient	**74**	SC-C-O14	Other outpatient
33	IP-C-O15	Other outpatient	**75**	SC-C-O15	Other outpatient
34	IP-C-O16	Other outpatient	**76**	SC-C-O16	Other outpatient
35	IP-C-O17	Screening	**77**	SC-C-O17	Screening
36	IP-C-O92	Other outpatient	**78**	SC-C-O92	Other outpatient
37	IP-C-O93	Other outpatient	**79**	SC-C-O93	Other outpatient
38	IP-C-O97	Other outpatient	**80**	SC-C-O97	Other outpatient
39	OP-B-null	Unknown	**81**	SC-C-O99	Unknown
40	OP-B-O0	Unknown	**82**	UN-D-UNK	Unknown
41	OP-B-O1	Emergency Presentation	**83**	UN-E-UNK	Unknown
42	OP-B-O2	Other outpatient			


Extracting the tumour cohort
Create a dataset containing all the tumour level records for cancer diagnoses in one year
Determining the pathway end point
Setting the scenario
(one third of the final route code)
Identify, for each tumour diagnosed, the closest interaction of the patient with both inpatient and outpatient hospital services within the 6 months prior to their date of diagnosis.
These end point dates will determine whether patients are on an inpatient/ outpatient/ unknown pathway.
Determining the pathway start point
Overwriting Inpatient transfers
(For patients on an inpatient pathway)
If the closest inpatient interaction prior to diagnosis was an admission based on a transfer from another hospital, replace it with the closest admission preceding this that was not a transfer as this is a more appropriate start point.
Allocating inpatient pathway patients
Now that the start of the inpatient journey has been identified, classify the inpatient interactions as emergency/ elective/other based on their admission method. Those that are elective or emergency outpatient admissions will also be flagged so any appropriate related previous outpatient appointments in the patient journey can be identified.
Find best outpatient attendance type
(For patients on an outpatient pathway)
Where the pathway to diagnosis was through outpatients or as an elective inpatient admission, identify the outpatient attendance that is most likely to have started the patient journey, closest to, but within 6 months of the date of diagnosis
Overwriting consultant initiated outpatient appointments
Check if the outpatient appointment selected is one that was initiated by a consultant. If so, then search through the appointment history for the closest preceeding one (within a 6 month blocks) that was not consultant initiated. This will be taken as the moset likely starting point of the pathway.
Allocating outpatient pathway patients
Now that the start of the outpatient journey has been identified, classify the outpatient referrals as emergency/ GP/ Other based on their referral source description.
Classifying the final source of the pathway
Creating a final source variable
(one third of the final route code)
Now that the most likely start point of the patient journey has been decided based on the available inpatient and outpatient hospital data, create a variable to show the final referral/admission source information
Setting the final pathway to diagnosis
Creating a pathway grouping code
(one third of the final route code)
Using the diagnosis path information and the relevant hospital information for each tumour create a Pathway Group.
Constructing final routes to diagnosis
By concatinating the three parts of the code, a final route code is derived and classified into one of 8 pathways.
Flagging screen detected and urgent red flag referral cases
Cases with screen detected flags or where there has been a red flag referral placed within an appropriate time frame have their final route overwritten to reflect this.

Running the Pathways to diagnosis SQL Code

There are 13 files to be run to produce the pathway to diagnosis for each of the tumour level records. Each of these process files will need to be run separately for each applicable year of diagnosis

All the files are stored and numbered in the **Project 024** folder in the subfolder called **'Jen Algorithm Replication'.**

1. Extract Tumour Cohort from NICR dataset

(You need to update the selected year in this section)

This section of code extracts the cohort and variables from the pre-prepared NICR dataset that are required for determining the pathway to diagnosis. These variables are saved into a new table which matches the PHE structure and this can then be updated to produce the finalised Pathways dataset.

2. Set Scenario

(You don't need to update any code in this section)

This file calculates the 'Route Scenario' for each record in the file. The Route Scenario is the end point for each patient that is the final contact point of the patient before they are diagnosed with cancer (based on the NICR date of diagnosis). This will form one third of the final route code.

There are a number of additional variables that will be used for the calculation; these are created in the first section. Once this has been done the code takes the Outpatient and Inpatient extracts and joins them to each tumour record in the NICR based on the STUDY_ID. Next the code selects the dates of the hospital inpatient and outpatient interactions which happened closest in time to the date of diagnosis of each record in the NICR extract. We only check within 180 days prior to the diagnosis date, anything outside this timeframe is not assumed to be connected.

The number of days between the closest Outpatient appointment and Inpatient admission for each tumour diagnosis is calculated. This information is then used to determine the route scenario number. The route scenarios are then aggregated into five different end point categories according to the rules

detailed above – Inpatient, Outpatient, Special Case, Death Certificate and Unknown.

3. Inpatient Transfer Overwrite

(You don't need to update any code in this section)

This file identifies instances where the closest Inpatient admission has an admission method of '81' meaning they were a transfer from another hospital. For these cases we want to look back to the inpatient admission that was closest before this that was not a transfer. In other words we want to try and identify the start point of the patient entering hospital. This change is only needed for patients who have been assigned to an inpatient pathway in the previous section.

There are a number of additional variables that will be used for the calculation; these are created in the first section. Once added, the code then places a flag against all cases where the admission method is '81' and the diagnosis path is classified as inpatient ('IP').

For this subset only, the code now identifies the closest admission prior to the one flagged that has an admission method other than '81' (i.e. not a transfer from another hospital). We only check within 180 days prior to the flagged date, anything outside this timeframe is not assumed to be connected. As a patient can have multiple admissions within the six month time period, the code uses the rank function order by clause to sure identify the closest. If there are two admissions on the same day episode end date then episode order are used to rank them.

The relevant information relating to these newly selected admissions is populated into the 'IP_PreTransfer_Admimeth',' IP_PreTransfer_Epistart' and 'IP_Elective_Admidate' variables.

Next this section of code assigns a 'final' admission method to each of the cases on the inpatient diagnosis pathway. For the cases above where we

identified a non-transfer admission, that admission method is used. Otherwise, the original method is used. There will likely be some of the '81' cases where we weren't able to override the admission with a more appropriate earlier one, in these cases the '81' is retained.

Finally, the final admission date information is updated for the cases where we now have a different admission to the one originally identified.

We have now identified he start point for all the inpatient pathway cases.

4. Allocate Inpatient Records

(You don't need to update any code in this section)

Now that we have identified a final admission method for all patients on the inpatient diagnosis pathway, this section of code buckets all the cases into 'emergency', 'elective' or 'other' categories.

Each group is populated based on the final admission method codes. There is some divergence from the PHE code here. For **emergency inpatients** codes 2A 2B 2C 2D have been removed as they don't exist in NI, all of these methods are included under code 28 instead. Also, code 25 (emergency admission following domiciliary visit by consultant) has been included; this doesn't appear in the English HES codes.

Where the inpatient method was non-elective then we need to remove the 'IP_Elective_Admidate' information. The next section of the code performs this step. Admission method code '24' needs to have the 'IP_Elective_Admidate' retained; this is an emergency code but is for outpatients so different to the other emergency categories.

Where an inpatient admission was elective (or via and emergency outpatient referral) we will go on to look for preceding outpatient appointments. For this reason these cases are now flagged so they can be identified easily.

Elective inpatients are now identified and flagged, including the emergency outpatients as well as the transfers that we were unable to overwrite before.

Finally, all remaining cases falling into the ‘other’ category are flagged. Again there is some divergence from PHE. Codes 84 and 89 which were in the PHE code have been removed as they don’t apply in NI; these are included under 0 which has been included here instead. These codes are all used to classify 'Unknown' admission method.

5. Find Best Outpatient Attendance Type

(You don’t need to update any code in this section)

Having dealt with those patients on the Inpatient pathway this section is addresses the outpatient initiated cases. This will be those bucketed into the ‘OP’ or ‘SC’ categories of those flagged as elective inpatients in the previous section. The code is used to find the most appropriate outpatient appointment prior to the date of diagnosis (i.e. the one most likely to have initiated the pathway to diagnosis).

There are a number of variables used by this section of code which must be set to null, also the ‘OP_Step_Flag’ must be set to 1 to identify the subset being examined.

It is important to note that some of the emergency outpatient cases were flagged for this in the previous section so it is important not to clear this flag or they will be dropped from the subset.

The code splits the outpatient appointments into first, review and unknown and examines each of these groups in relation to every tumour to find the closest in time to date of diagnosis, within a six month period. The appointments are placed in separate tables.

When selecting final appointments for the NICR extract preference is ranked in the following order; first appointments, then review and finally unknown (our dataset currently has no unknown cases). The appointment details are inserted into the appropriate columns in the main table and the temporary tables dropped as they are no longer required.

6. Outpatient Consultant Overwrite

(You don’t need to update any code in this section)

This section is used to flag outpatient appointments where the appointment identified as the start point closest to the date of diagnosis was initiated by a consultant. For these appointments we will look to see if there is an earlier more appropriate one that wasn't initiated by a consultant.

There are a number of additional variables that will be used for the calculation; these are created and populated as either 0 or null in the first section. Once this has been done the code flags all the selected outpatient appointments where the referral source is consultant initiated.

7. Outpatient Consultant Overwrite

(You don’t need to update any code in this section)

Now that the consultant initiated outpatient appointments have been flagged, this section of code is used to overwrite these, where possible, with the closest non-consultant initiated appointment that preceded it (again within a six month time frame). As before, the appointments are split into type and ranked with first being preferable, followed by review and unknown last.

The cases that are selected for the Outpatient Consultant Overwrite have specific referral source codes and cases are only selected where the outpatient appointment has not been initiated by the consultant responsible for the appointment.

In Northern Ireland we have noticed that there are a number of cases with a final Outpatient Route-to-Diagnosis who have a referral Source of 11.

This equates to “Initiated by consultant responsible for the appointment – Consultant – Other (11)”.

These cases are not selected as part of the consultant overwrite and maintain their “outpatient” route-to-diagnosis status.

The code has been written deliberately this way by PHE however this could be investigated to see if the same rationale in terms of coding applies in Northern Ireland.

Examination of a sample of these cases does show that earlier GP Referrals were present for a number of cases so altering the code to include these cases in the consultant overwrite could potentially lead to a decrease in the outpatient routes and an increase in GP Referrals.

However before making this amendment consultation with PAS Systems managers and clinicians would be required to assess the type of cases that are being assigned this referral source in the outpatient universe.

Finally, the PHE code preforms a looping process to keep looking back in 6 month blocks, at present this has not been automated for NI but was done on a manual basis. There are small numbers of cases involved. Further development is required to introduce a loop to the NI analysis.

8. Allocate Outpatient Records

(You don't need to update any code in this section)

This section of code is used to 'bucket' the final selection of outpatient appointments based on their referral source. Appointments are classified into three groups – 'GP referral', 'Emergency referral' and 'Other source of referral'. This information is not necessary for the pathway definition but is a useful extra insight into the data.

9. Consultant HES Related Route Code

(You need to update the year in the section used to code pathway group 'B')

This portion of code is used to construct the pathway grouping of the route to diagnosis and the final referral source. These are the other two thirds of the

final pathway to diagnosis code. This information will later be joined to the start point derived at the beginning of the algorithm.

The code creates a variable ('IPOP_Merged') which merges an Inpatient or Outpatient identifier with the final referral/admission source to form one third of the final pathway code.

Next the code creates a Pathway Grouping which will form the last third of the final pathway to diagnosis code. This grouping is based on the presence of inpatient and/or outpatient activity in the 6 months prior to the date of diagnosis. The pathway groups, A-E, have been detailed above.

With the three elements of the final pathway code now calculated ('Diagnosis_path', 'Pathway_Grouping' and 'IPOP_Merged_Source') the code concatenates this information into a single variable (route_code).

The code next aggregates these codes into 8 categories that can be used for analysis. The code uses a lookup file (RTD_Source_to_Route_Lookup) to carry out this aggregation. Details on the mapping of these categories have been included above.

If any cases have not been assigned they should be checked manually to determine which portion of the code contains an error.

10. Check for Screening Data

Screening information for NI cases is being collected from two sources – the outpatient hospital activity file and the NICR extract. This is different to PHE.

The majority of the screen flags come from the NICR who have validated cases directly with the breast, cervical and colorectal screening services. There are also a number of extra cases picked up through the outpatient activity file. These cases should primarily be cases referred in from Action Cancer (breast cancer only). There is no equivalent to these cases in England but they are being counted for NI as they constitute a valid route to diagnosis.

The code looks for any cases where the ‘Screen’ variable on the NICR extract that are classified as ‘screen detected’ and flags them.

11. Check for Suspect Red Flag Data

This code is used to identify cases which have been in receipt of a red flag cancer referral. For the PHE research this pathway was known as the Two Week Wait (TWW) pathway, reflecting the English target. For NI all red flags have been matched, including consultant upgrades and 'other'. The other group may include cases coming from other consultants or the likes of Action Cancer.

The code matches the NICR tumour level data to the Cancer Patient Pathway System data and identifies red flag referrals where there the date of decision to treat within 62 days prior to or 31 days after the date of diagnosis. These patients are flagged as red flag referrals.

The 62/31 day thresholds around these dates could potentially be revised in future iterations of the Northern Ireland analysis to better fit average waiting times in the Northern Ireland health care system.

12. Derive Final Route

The final step carried out here is to override and previously defined routes with the screening and red flag referral information.

The code then produces a table to show the number of cases for each pathway for the year in question.

Once a year has been done, rename the table at with the year appended onto the end. Repeat the above steps for each year being analysed.

Finally, run the code entitled **'final working dataset creation',** this will combine all the individual years into one dataset called **'finalworkingroutes'.** The code then takes the original NIRC extract and appends the final route code and aggregated pathway category on. This is saved as **'COMPLETED_DATASET';** it is this file that the next stage of analysis will be run from.