South West Public Health Observatory

Routes to Diagnosis – Technical Supplement

30 September 2010

NCIN national cancer intelligence network
Using information to improve quality & choice

Avon, Somerset and Wiltshire NHS
Cancer Services

Contents

Acknowledgements

This project has been led and managed on behalf of the National Cancer Intelligence Network (NCIN) by Lucy Elliss-Brookes, Associate Director Commissioning and Cancer Intelligence, Avon Somerset and Wiltshire Cancer Services. Professor Sir Mike Richards, National Cancer Director, has acted as clinical adviser to the project. Analytical and technical support, input and leadership have been provided by members of the South West Public Health Observatory, in particular Alexander Ives, Dr Julia Verne, Tariq Malik, Matt Greenslade, Luke Hounsome, Andy Pring, Davidson Ho and Roy Maxwell.

This work would not have been possible without collaboration and support from numerous individuals within a number of organisations including the NCIN Coordinating Team, Cancer Research UK, the National Cancer Services Analysis Team, the Department of Health and the eight English Cancer Registries. The authors would like to specifically thank Catherine Thomson, Andy Smith and Chris Carrigan for their contributions.

The vision, concept and approach to the 'Routes to Diagnosis' project was shaped from the outset by the late Dr Brian Cottier.

1.0 Introduction

1.1 Purpose of the report

The aims of this report are to summarise the data sources and methodology used for the `Routes to Diagnosis' project.

The 'Routes to Diagnosis' project supports the evolution of the National Awareness and Early Diagnosis Initiative (NAEDI) whose aim is to promote earlier diagnosis of cancer and thereby improve survival rates and reduce cancer mortality. Successful implementation of NAEDI will make a major contribution to the Cancer Reform Strategy goal of achieving world class cancer outcomes in this country.

The 'Routes to Diagnosis' project is the first to explore the feasibility of using routine data to evaluate how cancer patients access the health service for diagnosis and whether the routes are associated with survival differences. This in turn could be used to inform strategy in terms of: improved patient education regarding signs and symptoms, medical practitioner education and routes of referral. The outputs will help to inform awareness and early diagnosis initiatives locally and nationally, ideally resulting in more appropriate referrals and earlier diagnosis of cancer as well as eventually improving the cost effectiveness of NHS.

The hypotheses tested in the 'Routes to Diagnosis' project are described below:

- Is it feasible to use routinely available data sources to define the routes to diagnosis for patients diagnosed with cancer? For example, whether they present through inpatients, outpatients, screening or emergency presentation.
- If the first is feasible, can the influence of age, sex, ethnicity, deprivation and geographical area of residence on referral routes and pathways be examined?
- Is there an association between routes to diagnosis and survival for cancer patients?

A pilot study was conducted in the South West region to test the feasibility of the approach and based on lessons learned, the national study was conducted (reported in the NCIN Data Briefing, *Routes to Diagnosis*, www.ncin.org.uk).

1.2 Data sources and matching algorithms

Cancer registration

The National Cancer Data Repository holds cancer registration data for the whole of England. The repository contains over 8.5 million cancer registry records. This provides an unparalleled resource for exploring hypotheses and evaluating quality of care. Further information about the National Cancer Data Repository is available from the NCIN website www.ncin.org.uk.

For the national analysis, all 2007 cancer registrations across England with ICD-10 diagnosis codes C00–C97 (malignant cancers) were obtained from the National Cancer Data Repository.

On the basis of observations made during the pilot study, the following exclusions were made:

- The records for patients resident outside England.
- The records for patients in whom multiple tumours were diagnosed during 2007 as it was difficult to distinguish the routes to diagnosis.
- The records for patients with ICD-10 codes D05 and D06 (in situ breast and in situ cervical cancers) because during the pilot phase it was shown that the data for these is of very variable quality with substantial missing information.
- The records for patients with non-melanoma skin cancer, as most of these are diagnosed and treated immediately in outpatients or in primary care and Basal Cell Carcinomas are not subject to the Two Week Wait referral process.

Hospital Episode Statistics (HES)

Hospital Episode Statistics (HES) is a data warehouse containing details of all admissions (day case and inpatient) to NHS hospitals in England. It includes details of private patients treated in NHS hospitals, patients who were resident outside England and of care delivered by treatment centres (including those in the independent sector) funded by the NHS. HES also contain details of all NHS outpatient appointments (attendances for patients who are not formally admitted) in England. It contains admitted patient care data from 1989 onwards, with more than 12 million new records added each year, and outpatient attendance data from 2003 onwards, with more than 40 million new records added each year. Further information about HES is available from the HES online website www.hesonline.org.uk

Inpatient Hospital Episode Statistics

For the national analysis, Inpatient (IP) Hospital Episode Statistics (HES) for 2005 to 2007 were used to identify patients with a hospital admission for any cause during this time period.

Outpatient Hospital Episode Statistics

For the national analysis, Outpatient (OP) Hospital Episode Statistics (HES) for 2004/05 to 2007/08 were used.

Matching algorithm for cancer registration data and HES data

Initially, patient records from the National Cancer Data Repository were matched to HES records using the NHS Number only. Then a more stringent system of matching was applied as described below. It is important to note that all the matching was exact on each field, and no 'fuzzy' matching was undertaken. Patient records were matched using four data fields: NHS Number, Date of Birth, Postcode, and Sex. The data sets to be matched have these four fields identically formatted – this is particularly important with the postcode field as formats vary between data sets.

The algorithm runs in four distinct batches:

- The first batch matches patients on all four fields where none of the fields is a NULL value (No data). This is therefore the most accurate match and is assigned a 'match rank' of 1.
- The second batch repeats the process without the Postcode field being used but only on records not already matched. This is because there is a chance that people may have changed address between data sets being compiled, especially if the data span a long time frame. This is assigned 'match rank' 2.

- The third batch matches any records not matched in batches 1 and 2 using only NHS Number and Postcode. This is assigned 'match rank' 3. This attempts to catch anyone whose date of birth details were unknown or incorrectly entered or where there was an incorrect entry for their sex.
- The final batch runs against any records still unmatched and uses Postcode, Date of Birth and Sex. This does not require NHS Numbers to match. This is the least reliable match and is assigned a 'match rank' 4.

We have decided to exclude 'match rank' 4 HES data from the national study. Therefore, HES data were excluded where there was a possibility that the same person had been given two different NHS Numbers across the HES and National Cancer Repository datasets. The number of patients from the National Cancer Data Repository matched to HES data at each rank is shown in Tables 1.1 and 1.2.

Table 1.1: Patients from the National Cancer Data Repository matched to Inpatient HES by rank

Rank	Patients	Percentage
1	272,565	79.5%
2	42,139	12.3%
3	7,435	2.2%
4	1,470	0.4%
Unmatched	19,406	5.7%
Total	343,015	100.0%

Table 1.2: Patients from the National Cancer Data Repository matched to Outpatient HES by rank

Rank	Patients	Percentage
1	229,295	66.8%
2	43,452	12.7%
3	5,225	1.5%
4	0	0.0%
Unmatched	65,043	19.0%
Total	343,015	100.0%

National Cancer Waiting Times

For the national analysis, National Cancer Waiting Times (NCWT) data for 2005 to 2007 were used.

The NCWT system is hosted nationally on NHSNet (Open Exeter) and allows NHS providers to record data derived from patient care activity. These data are used to monitor performance against the NCWT standards specified in the NHS Cancer Plan 2000 and the Cancer Reform Strategy 2007. As a patient moves through the stages of their treatment pathway, data on referrals, treatments and diagnosis are derived from care records locally. NHS providers are mandated by Data Set Change Notice (DCSN) 20/2008 to collect data concerning all patients covered by the NCWT standards, including patients referred with suspected cancer and patients diagnosed with and treated for new and subsequent cancer. Further information about the NCWT system is available from the Department of Health website www.dh.gov.uk.

The data extract provided by Connecting for Health via Trent Cancer Registry contained three data fields (NHS Number, Referral Priority, Decision to Treat Date)

for all patients with a Date First Seen between 1st January 2005 and 31st December 2007.

Matching algorithm for cancer registration data and National Cancer Waiting Times Data

Records were selected that had a referral priority of Two Week Wait and a valid Decision to Treat Date. These were then matched to the records in the National Cancer Data Repository using NHS Number and having a Cancer Diagnosis date within 31 days (before or after) of the Decision to Treat Date.

The Decision to Treat Date was used as a proxy for date of diagnosis. The 31 days timeframe is supported by the National Cancer Waiting Times standards which demand 31 days from decision to treat to treatment. The validity of this method was checked by comparing the data for the South West from the national extract with South West Cancer Intelligence Service data because the South West Cancer Registry already includes Two Week Wait data. This resulted in a similar number of matched cases.

Screening

The screening route of diagnosis only applies to breast and cervical cancer patients. An extract of the National Cancer Data Repository was sent to each of the English cancer registries. The extract contained the breast and cervix cancer patients diagnosed in 2007 in the relevant registry catchment area. Each registry was asked to complete the screen detected field (indicating whether the tumour was detected via the National Screening Programme), which was then used to update the extract from the National Cancer Data Repository.

The following fields were sent to each of the cancer registries for screening extract: NHS Number; Date of Birth; Postcode; Diagnosis Date; ICD-10 Code; Source. Additional identifiers (Patient ID, Cancer ID, Forename and Surname) were requested by the Eastern Cancer Registration and Information Centre.

Screening data were provided by the cancer registries, based on the records held internally as a result of local data exchanges between the cancer registries and the screening Quality Assurance Reference Centres (QARCs).

Accident and Emergency data

At the start of the project, and in particular following the pilot study in which it was observed that overall 22% of patients presented as emergencies (range 61% for acute leukaemia to 4% melanoma), it was hoped to include Accident and Emergency (A&E) data. Unfortunately this could not be obtained for either phase of the project.

2.0 Methods

The analysis takes as a starting point the date of cancer diagnosis. A set of rules was defined to identify the sequence of events that make up the different routes to diagnosis. The routes can be categorised by three variables: setting for diagnosis, the pathway, and the source of first contact. These routes have been grouped to facilitate analysis. The approach follows the patient rather than the tumour.

For patients classified as being diagnosed as inpatients the approach has defined a sequence of events referenced back to an outpatient referral, if possible, and therefore applying source of referral to the pathway. Alternatively, the diagnosis has been referenced back to an emergency admission which is then broken down more specifically. For patients classified as being diagnosed in outpatients the approach identified the source of referral.

It is important to note that patient records being used to describe the route to diagnosis will 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.

2.1 Assigning the routes to diagnosis codes

For each patient, a setting for diagnosis, the pathway and source of first contact were derived and an overall route was defined by the concatenation of these three codes in the specific order: setting of diagnosis–pathway–source of first contact (e.g. IP-02-O03). This resulted in a total of 269 distinct routes to diagnosis codes, listed in Appendix 1. The rules for deriving the setting for diagnosis, the pathway and source of first contact codes are specified in the following three sections.

Cancer screening and Two Week Wait data (TWW) (from NCWT) are the last information to be added to the analysis of routes. When these are applied, screening has preference over TWW, which in turn has preference over any previously assigned route to diagnosis.

A key element of this project is to examine the routes to diagnosis and outcomes for patients who are not referred via the TWW referral route. The overwriting of the codes with the TWW flag is required because most cancer registries do not record whether a patient was originally referred under the TWW rule, and these data are not recorded in HES, hence the need to apply the code separately. These patients may have been referred by their GP to outpatients or may have been referred straight to diagnostic test.

The overwriting of the codes with the screening flag is based on the assumption that the quality of the matching of QARC data to cancer registry data is better than the information being taken from NCWT and HES and matched to the National Cancer Data Repository.

2.2 Assigning the setting for diagnosis code

Setting for diagnosis codes were assigned to each patient record depending on whether data existed in inpatient or outpatient HES and how the diagnosis date related to the date ranges in those datasets. Codes were assigned according to the following hierarchy:

- **Multiple diagnoses** (MD, MI, MO, MS, MU) – patients with multiple malignant diagnosis records in 2007 in the cancer registry, but with the same diagnosis date for each diagnosis.
- **Multiple not defined** (NULL) – patients with multiple malignant diagnosis records in 2007 in the cancer registry, but with multiple different cancer diagnosis dates.

- **Special cases** (SC) – patients with a cancer diagnosis date on the same day as an inpatient admission date and an outpatient attendance date. These are a special case of inpatient diagnosis.
- **Inpatient diagnosis** (IP) – patients with a cancer diagnosis date related to a preceding inpatient HES episode (excluding patients already excluded as special cases). Patient may have multiple inpatient episodes. 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 28 days after the end of an episode would also be regarded as an inpatient diagnosis.
- **Outpatient diagnosis** (OP) – patients with no inpatient HES episode preceding the cancer diagnosis date (as defined above) but with an outpatient HES attendance preceding the cancer diagnosis date.
- **Unknown** (UN) – Unable to match cancer diagnosis date to any inpatient or outpatient HES episode. 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 (e.g. 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** (DC) - The cancer registry receives a small number of cancer related death notifications, for which, despite extensive enquiries, 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.

2.3 Assigning the pathway code

A set of rules was devised to identify the routes to diagnosis in a cohort of all patients resident in the South West cancer registry region diagnosed in 2007, using matched data from the registry, inpatient and outpatient Hospital Episode Statistics for 2005/06 and 2006/07 and Cancer Waiting Times. By working backwards from diagnosis, through their cancer journey, the sequence of events leading to diagnosis were ascertained. These rules were applied to the national data.

The diagram in Appendix 2 illustrates the algorithm used to identify the pathway codes, each of which is described below.

Some of the pathways refer to an outpatient consultant loop. This means that an outpatient attendance has been identified that has a source of referral from another consultant. In these instances, an earlier outpatient attendance that might have preceded that attendance is sought, so the pathway works backwards where possible until a non-consultant referral is identified as the originating source of referral for the pathway.

- Pathway 1 (outpatient to inpatient): These will all be inpatient diagnoses with inpatient admissions that are preceded by an outpatient attendance that does not involve an outpatient consultant loop.
- Pathway 2 (outpatient to consultant outpatient to inpatient): These will all be inpatient diagnoses with inpatient admissions that are preceded by an outpatient attendance that involves an outpatient consultant loop.
- Pathway 3 (inpatient): These will all be inpatient diagnoses with inpatient admissions with no preceding outpatient information. The inpatient method of admission will not be a transfer.

- Pathway 4 (outpatient to inpatient via transfer): These will all be inpatient diagnoses with inpatient admissions that have an admission method of transfer. This inpatient admission will then be preceded by an outpatient attendance that does not involve an outpatient consultant loop.
- Pathway 5 (outpatient to consultant outpatient to inpatient via transfer): These will all be inpatient diagnoses with inpatient admissions that have an admission method of transfer. This inpatient admission will then be preceded by an outpatient attendance that involves an outpatient consultant loop.
- Pathway 6 (inpatient via transfer): These will all be inpatient diagnoses with inpatient admissions that have a method of admission of transfer. This inpatient admission will be preceded by an inpatient admission with a method of admission that is not transfer or emergency. There will be no preceding outpatient attendances.
- Pathway 7 (outpatient to inpatient via emergency: via consultant outpatient clinic): These will all be inpatient diagnoses with inpatient admissions that have an admission method of emergency via a consultant outpatient clinic. This admission will then be preceded by an outpatient attendance that does not involve an outpatient consultant loop.
- Pathway 8 (outpatient to consultant outpatient to inpatient via emergency: via consultant outpatient clinic): These will all be inpatient diagnoses with inpatient admissions that have an admission method of emergency via a consultant outpatient clinic. This inpatient admission will then be preceded by an outpatient attendance that involves a further outpatient consultant loop.
- Pathway 9 (inpatient via outpatient clinic (no outpatient attendance)): These will all be inpatient diagnoses with inpatient admissions with no preceding outpatient information.
- Pathway 10 (no IP HES): These will all be a mixture of outpatient diagnoses and unknowns. The outpatient diagnoses will have no inpatient admissions but will have an outpatient attendance prior to cancer diagnosis. The unknowns will have no inpatient admissions and no outpatient attendances.
- Pathway 11 (IP pre-diagnosis): These will all be a mixture of outpatient diagnoses and unknowns. The outpatient diagnoses will have inpatient admissions earlier than 28 days prior to diagnosis but will have an outpatient attendance closer to cancer diagnosis. The unknowns will have inpatient admissions earlier than 28 days prior to diagnosis but no intervening outpatient attendances, and the method of admission will be unknown.
- Pathway 12 (IP post-diagnosis): These will all be a mixture of outpatient diagnoses and unknowns, all of which will only have an inpatient admission after diagnosis. The outpatient diagnoses will have no inpatient admissions prior to cancer diagnosis but will have an outpatient attendance prior to cancer diagnosis. The unknowns will have no inpatient admissions prior to cancer diagnosis and no outpatient attendances.
- Pathway 13 (No IP HES -Outpatient post-diagnosis): These will all be unknowns as there is no inpatient admission and the outpatient attendance is after cancer diagnosis.
- Pathway 14 (IP pre-diagnosis - Outpatient post-diagnosis): These will all be unknowns as there is no inpatient admission within 28 days prior to cancer diagnosis and the outpatient attendance is after cancer diagnosis.
- Pathway 15 (IP post-diagnosis - Outpatient post-diagnosis): These will all be unknowns as the inpatient admissions and outpatient attendances are after cancer diagnosis.

- Pathway 16 (No IP HES via consultant outpatient): These will all be outpatient diagnoses with outpatient attendances but no inpatient admission.
- Pathway 17 (IP pre-diagnosis via consultant outpatient): These will all be outpatient diagnoses which have inpatient admissions earlier than 28 days prior to diagnosis but will have an outpatient attendance closer to cancer diagnosis. This outpatient attendance will then involve a further outpatient consultant loop.
- Pathway 18 (IP post-diagnosis via consultant outpatient): These will all be outpatient diagnoses with outpatient attendances prior to diagnosis but no inpatient admission prior to cancer diagnosis.
- Pathway 19 (No route defined (multiple diagnoses not on same day)): These will be a mixture of all of the different diagnoses and routes and excluded from the main analysis as they refer to patients with multiple malignant diagnoses on different days. Further analysis of this group of patients is required.
- Pathway NULL (Special cases or transfers): These will all be inpatient diagnoses. The special cases will all be inpatient diagnoses with inpatient admissions that are preceded by an outpatient attendance, where the cancer diagnosis date is on the same day as the inpatient admission date and the outpatient attendance date. The transfers will all be inpatient diagnoses with inpatient admissions that have a method of admission of transfer but no preceding inpatient admissions or outpatient attendances.

2.4 Assigning the source of first contact code

For pathways that originated in an outpatient attendance, the outpatient source of referral of that attendance has been assigned as the 'source of first contact' code.

For pathways that originated in an inpatient episode, the inpatient method of admission has been assigned as the 'source of first contact' code.

For pathways where inpatient or outpatient data were unavailable the source of first contact codes may be assigned as null or unknown (this will also include DCOs).

A list of all possible 'source of first contact' codes is provided in Appendix 3.

2.5 Routes to diagnosis categorised for analyses

For the analysis, 21 tumour types were identified (22 including 'other'), primarily based on their relevance to the NAEDI agenda. These were then broken down by age band, sex, deprivation quintile and Cancer Networks. The list of tumour types by International Classification of Disease (ICD-10) codes is provided in Appendix 4.

Based on the component setting for diagnosis code, pathway code and source of first contact code, the 269 individual routes to diagnosis codes have been categorised into eight broad routes to diagnosis categories:

- **GP/outpatient referral**: includes routine and urgent referrals where the patient was not referred under the Two Week Wait referral route.
- **Two Week Wait**: urgent GP referrals with a suspicion of cancer.
- **Emergency presentation**: an emergency route via A&E, emergency GP referral, emergency consultant outpatient 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 data available from IP or OP HES or from NCWT or screening.

3.0 Presentation of results and statistical testing

The results are available in spreadsheet format on the NCIN website (www.ncin.org.uk). The majority of results are presented in tabular form with a continuous colour gradient applied such that larger percentages are a darker colour.

One-year relative survival was calculated using tools developed by the London School of Hygiene and Tropical Medicine (LSHTM)[1] run using Stata version 10. Relative survival is the ratio of the observed cumulative probability of survival in the study group and the survival that would have been expected if the group had only been subject to the background mortality in the general population (obtained from life tables). The particular life table used allowed for variations in background mortality by age, sex, region and social deprivation.

The survival figures which are statistically significantly different at the 95% confidence interval limit from the 'all routes' average are highlighted.

Due to small numbers and data completeness caution must be used when interpreting results for screen detected cervical and breast cancers at Cancer Network level.

[1] Cancer Research UK Cancer Survival Group (2006). *strel* computer program, version 1.2.7 and life tables for cancer survival analysis. Downloaded from www.lshtm.ac.uk/ncde/cancersurvival/tools.htm on 8 Feb 2010. Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, UK

4.0 Data quality issues and limitations

4.1 Data quality issues for London

The matching of HES data to National Cancer Repository data is incomplete for some London Primary Care Trusts (PCTs). Investigations into why this has happened are ongoing, and rather than publish results based on this data, the particular PCTs (see Table 4.1) where the problem was most acute have been omitted from this report. 9,173 patients are resident in these London PCTs and are excluded from the analysis (2.7% of the national cohort).

Table 4.1: London Primary Care Trusts excluded from analysis

Number	PCT
1	Newham
2	Redbridge
3	Richmond and Twickenham
4	Southwark
5	Sutton and Merton
6	Tower hamlets
7	Waltham forest
8	Wandsworth
9	Westminster

4.2 Screening data

An analysis of completeness of screening flags for England provided by cancer registries was undertaken, see Table 4.2. The breakdown by cancer registry shows a variation in the percentage of screen detected records assigned by each cancer registry, and in particular the figures for cervical in situ appear to be lower than expected for the majority of registries, see Table 4.3. This supported the exclusion of in situ codes from the main analysis.

Table 4.2: The number of records for England patients against each of the breast and cervical ICD-10 groupings (C50, C53, D05 and D06) by screen detected flag

Diagnosis Group		Screen Detected Flag			Percentage
		Yes	No	Total	Screen Detected
Breast	Malignant	8,357	31,002	39,359	21.2%
	In-Situ	2,003	2,704	4,707	42.6%
Cervix	Malignant	360	2,007	2,367	15.2%
	In-Situ	1,051	20,504	21,555	4.9%
Total records		**11,771**	**56,217**	**67,988**	**17.3%**

Table 4.3: The percentage of records against each of the breast and cervical ICD-10 groupings (C50, C53, D05 and D06) by screen detected flag, broken down by English cancer registry

Percentage of records that were Screen Detected	Breast		Cervix	
Cancer Registry	**Malignant**	**In-Situ**	**Malignant**	**In-Situ**
Eastern Cancer Registration and Information Centre	29.9%	63.6%	16.2%	0.0%
North West Cancer Intelligence Service	17.2%	38.0%	7.1%	0.2%
Northern and Yorkshire Cancer Registry and Information Service	26.6%	62.0%	23.5%	0.2%
Oxford Cancer Intelligence Unit	29.8%	61.8%	27.7%	0.1%
South West Cancer Intelligence Service	27.0%	54.2%	22.2%	0.1%
Thames Cancer Registry	8.1%	10.3%	0.2%	0.0%
Trent Cancer Registry	31.2%	57.6%	11.6%	8.4%
West Midlands Cancer Intelligence Unit	11.8%	12.3%	27.2%	33.5%
Total records	**21.2%**	**42.6%**	**15.2%**	**4.9%**

4.3 Ethnicity

It would have been desirable to examine the effect of ethnicity on the routes to diagnosis. However, there is significant under reporting of ethnicity data in the 'Routes to Diagnosis' dataset, see Table 4.4.

Table 4.4: Comparison of ethnicity recording in the NCIN ethnicity report against the 'Routes to Diagnosis' dataset

Ethnicity data profile	NCIN Ethnicity report		Routes to Diagnosis	
	People	**Percentage**	**People**	**Percentage**
White	435,168	73%	140,369	40%
Asian	6,685	1%	1,603	0%
Black	6,540	1%	1,619	0%
Chinese	651	0%	264	0%
Mixed	1,058	0%	378	0%
Other	3,194	1%	148,515	42%
Unknown	145,299	24%	58,018	17%
Total	**598,595**	**100%**	**350,766**	**100%**

4.4 Death Certificate Only

Patients who were registered as a DCO on the National Cancer Data Repository and could not be matched to any of the data sources referenced in Section 1.2 above were assigned a DCO route grouping. However, there were patients registered as DCOs where additional information was found in inpatient and/or outpatient HES data which allowed these patients to be assigned a different route grouping, see Tables 4.5 and 4.6. This finding has important incidental implications for reducing the DCO rate for cancer registries, see Table 4.7. All tables below show the number of records as opposed to the number of distinct patients, which includes all records in the analysis (i.e. it does not exclude multiples or in situ tumours).

Table 4.5: Comparison of the number of records assigned to the different routes to diagnosis against the number of records that have been flagged by the cancer registry as being DCO or non-DCO

Count of records		Registry DCO Flag		
Route to Diagnosis		**Yes**	**No**	**Total**
DCO		1,426	-	1,426
Not DCO	Emergency presentation	3,780	49,371	53,151
	GP referral	856	57,168	58,024
	Inpatient elective	60	5,303	5,363
	Other outpatient	557	33,644	34,201
	Screening	6	7,682	7,688
	TWW	67	58,905	58,972
	Unknown	587	22,707	23,294
Not DCO Total		5,913	234,780	240,693
Total records		7,339	234,780	242,119

Table 4.6: Comparison of percentage of records assigned to the DCO and non-DCO routes to diagnosis groupings against the percentage of records that have been flagged by the cancer registry as being DCO or non-DCO

Percentage of records	Registry DCO Flag		
Route to Diagnosis	**Yes**	**No**	**Total**
DCO	0.59%	0.00%	0.59%
Not DCO	2.44%	96.97%	99.41%
Total	3.03%	96.97%	100.00%

Table 4.7: Comparison of the percentage of records assigned to the DCO routes to diagnosis groupings against the percentage of records that have been flagged by the cancer registry as being DCO, broken down by cancer registry

Percentage of Registry records	Registry DCO	Route to
Cancer Registry	**Flag = Yes**	**Diagnosis = DCO**
Eastern Cancer Registration and Information Centre	0.02%	0.00%
North West Cancer Intelligence Service	7.65%	0.79%
Northern and Yorkshire Cancer Registry and Information Service	1.49%	0.46%
Oxford Cancer Intelligence Unit	1.43%	0.26%
South West Cancer Intelligence Service	1.80%	0.43%
Thames Cancer Registry	2.54%	1.05%
Trent Cancer Registry	2.01%	0.37%
West Midlands Cancer Intelligence Unit	7.03%	0.90%
All records	**3.03%**	**0.59%**

5.0 Recommendations for further analysis

It is recommended that the following areas are investigated in more detail:

- The major contributory factors to the unknown routes of diagnosis.
- Characteristics of patients and their cancers who present as emergencies.
- The interaction between cancer type, patient demographics and routes to diagnosis.
- The causes of the discrepancies between the cancer registry and routes to diagnosis classification of DCOs.
- Continued analysis of the distribution of the proportion of cases in each route to diagnosis, by tumour type, over time.

It is also recommended that there is full clinician engagement with these findings to elucidate the best approaches to achieving earlier diagnosis for those cancer types where late presentation presents a particular challenge.

Appendix 1: Routes to diagnosis codes

A list of all 'routes to diagnosis' codes is provided in Table A1.1. The route code is in the form of 'setting of diagnosis – pathway – source of first contact'. For patients diagnosed with multiple cancers on different days, the route code is preceded by an 'M'.

Table A1.1: Routes to diagnosis codes

Number	Route Code	Route Group
1	DC-10-DCO	DCO
2	DC-19-DCO	DCO
3	IP-01-I99	Other outpatient
4	IP-01-O01	Emergency presentation
5	IP-01-O02	Other outpatient
6	IP-01-O03	GP referral
7	IP-01-O04	Emergency presentation
8	IP-01-O05	Other outpatient
9	IP-01-O06	Other outpatient
10	IP-01-O07	Other outpatient
11	IP-01-O08	Other outpatient
12	IP-01-O10	Emergency presentation
13	IP-01-O11	Other outpatient
14	IP-01-O12	GP referral
15	IP-01-O13	Other outpatient
16	IP-01-O17	Screening
17	IP-01-O92	Other outpatient
18	IP-01-O93	Other outpatient
19	IP-01-O97	Other outpatient
20	IP-02-I99	Other outpatient
21	IP-02-O01	Emergency presentation
22	IP-02-O02	Other outpatient
23	IP-02-O03	GP referral
24	IP-02-O04	Emergency presentation
25	IP-02-O06	Other outpatient
26	IP-02-O07	Other outpatient
27	IP-02-O08	Other outpatient
28	IP-02-O10	Emergency presentation
29	IP-02-O11	Other outpatient
30	IP-02-O12	GP referral
31	IP-02-O13	Other outpatient
32	IP-02-O17	Screening
33	IP-02-O92	Other outpatient
34	IP-02-O93	Other outpatient
35	IP-02-O97	Other outpatient
36	IP-03-I11	Inpatient elective
37	IP-03-I12	Inpatient elective
38	IP-03-I13	Inpatient elective
39	IP-03-I21	Emergency presentation
40	IP-03-I22	Emergency presentation
41	IP-03-I23	Emergency presentation
42	IP-03-I31	Inpatient elective
43	IP-03-I82	Inpatient elective
44	IP-03-I99	Unknown
45	IP-04-I99	Other outpatient
46	IP-04-O06	Other outpatient
47	IP-04-O08	Other outpatient
48	IP-04-O10	Emergency presentation
49	IP-04-O11	Other outpatient
50	IP-04-O12	GP referral
51	IP-04-O97	Other outpatient
52	IP-05-O01	Emergency presentation
53	IP-05-O02	Other outpatient
54	IP-05-O03	GP referral
55	IP-05-O04	Emergency presentation
56	IP-05-O08	Other outpatient
57	IP-05-O11	Other outpatient
58	IP-05-O92	Other outpatient
59	IP-06-I11	Inpatient elective
60	IP-06-I12	Inpatient elective
61	IP-06-I13	Inpatient elective
62	IP-06-I21	Emergency presentation
63	IP-06-I22	Emergency presentation
64	IP-06-I23	Emergency presentation
65	IP-06-I82	Inpatient elective
66	IP-07-I99	Emergency presentation
67	IP-07-O01	Emergency presentation
68	IP-07-O02	Other outpatient
69	IP-07-O03	GP referral
70	IP-07-O04	Emergency presentation
71	IP-07-O05	Other outpatient
72	IP-07-O06	Other outpatient
73	IP-07-O07	Other outpatient
74	IP-07-O08	Other outpatient
75	IP-07-O10	Emergency presentation
76	IP-07-O11	Other outpatient
77	IP-07-O12	GP referral
78	IP-07-O13	Other outpatient
79	IP-07-O17	Screening
80	IP-07-O92	Other outpatient
81	IP-07-O97	Other outpatient
82	IP-08-I99	Other outpatient
83	IP-08-O01	Emergency presentation
84	IP-08-O02	Other outpatient
85	IP-08-O03	GP referral
86	IP-08-O04	Emergency presentation
87	IP-08-O06	Other outpatient
88	IP-08-O08	Other outpatient
89	IP-08-O11	Other outpatient
90	IP-08-O12	GP referral

Table A1.1: Route to diagnosis codes continued

Number	Route Code	Route Group
91	IP-04-O01	Emergency presentation
92	IP-04-O02	Other outpatient
93	IP-04-O03	GP referral
94	IP-04-O04	Emergency presentation
95	IP-04-O05	Other outpatient
96	IP-19-I12	Inpatient elective
97	IP-19-I13	Inpatient elective
98	IP-19-I21	Emergency presentation
99	IP-19-I22	Emergency presentation
100	IP-19-I23	Emergency presentation
101	IP-19-I99	Unknown
102	IP-19-O01	Emergency presentation
103	IP-19-O02	Other outpatient
104	IP-19-O03	GP referral
105	IP-19-O04	Emergency presentation
106	IP-19-O05	Other outpatient
107	IP-19-O06	Other outpatient
108	IP-19-O07	Other outpatient
109	IP-19-O08	Other outpatient
110	IP-19-O10	Emergency presentation
111	IP-19-O11	Other outpatient
112	IP-19-O12	GP referral
113	IP-19-O13	Other outpatient
114	IP-19-O92	Other outpatient
115	IP-19-O97	Other outpatient
116	IP-null-I81	Inpatient elective
117	null-14-null	Unknown
118	null-null-null	Unknown
119	OP-10-I99	Other outpatient
120	OP-10-O01	Emergency presentation
121	OP-10-O02	Other outpatient
122	OP-10-O03	GP referral
123	OP-10-O04	Emergency presentation
124	OP-10-O05	Other outpatient
125	OP-10-O06	Other outpatient
126	OP-10-O07	Other outpatient
127	OP-10-O08	Other outpatient
128	OP-10-O10	Emergency presentation
129	OP-10-O11	Other outpatient
130	OP-10-O12	GP referral
131	OP-10-O13	Other outpatient
132	OP-10-O17	Screening
133	OP-10-O92	Other outpatient
134	OP-10-O93	Other outpatient
135	OP-10-O97	Other outpatient
136	OP-11-I99	Other outpatient
137	OP-11-O01	Emergency presentation
138	OP-11-O02	Other outpatient
139	OP-11-O03	GP referral
140	OP-11-O04	Emergency presentation
141	IP-08-O92	Other outpatient
142	IP-08-O93	Other outpatient
143	IP-08-O97	Other outpatient
144	IP-09-I24	Emergency presentation
145	IP-19-I11	Inpatient elective
146	OP-11-O11	Other outpatient
147	OP-11-O12	GP referral
148	OP-11-O13	Other outpatient
149	OP-11-O17	Screening
150	OP-11-O92	Other outpatient
151	OP-11-O93	Other outpatient
152	OP-11-O97	Other outpatient
153	OP-12-I99	Other outpatient
154	OP-12-O01	Emergency presentation
155	OP-12-O02	Other outpatient
156	OP-12-O03	GP referral
157	OP-12-O04	Emergency presentation
158	OP-12-O05	Other outpatient
159	OP-12-O06	Other outpatient
160	OP-12-O07	Other outpatient
161	OP-12-O08	Other outpatient
162	OP-12-O10	Emergency presentation
163	OP-12-O11	Other outpatient
164	OP-12-O12	GP referral
165	OP-12-O13	Other outpatient
166	OP-12-O17	Screening
167	OP-12-O92	Other outpatient
168	OP-12-O93	Other outpatient
169	OP-12-O97	Other outpatient
170	OP-14-O03	GP referral
171	OP-16-I99	Other outpatient
172	OP-16-O01	Emergency presentation
173	OP-16-O02	Other outpatient
174	OP-16-O03	GP referral
175	OP-16-O04	Emergency presentation
176	OP-16-O06	Other outpatient
177	OP-16-O07	Other outpatient
178	OP-16-O08	Other outpatient
179	OP-16-O10	Emergency presentation
180	OP-16-O11	Other outpatient
181	OP-16-O12	GP referral
182	OP-16-O17	Screening
183	OP-16-O92	Other outpatient
184	OP-16-O93	Other outpatient
185	OP-16-O97	Other outpatient
186	OP-17-I99	Other outpatient
187	OP-17-O01	Emergency presentation
188	OP-17-O02	Other outpatient
189	OP-17-O03	GP referral
190	OP-17-O04	Emergency presentation

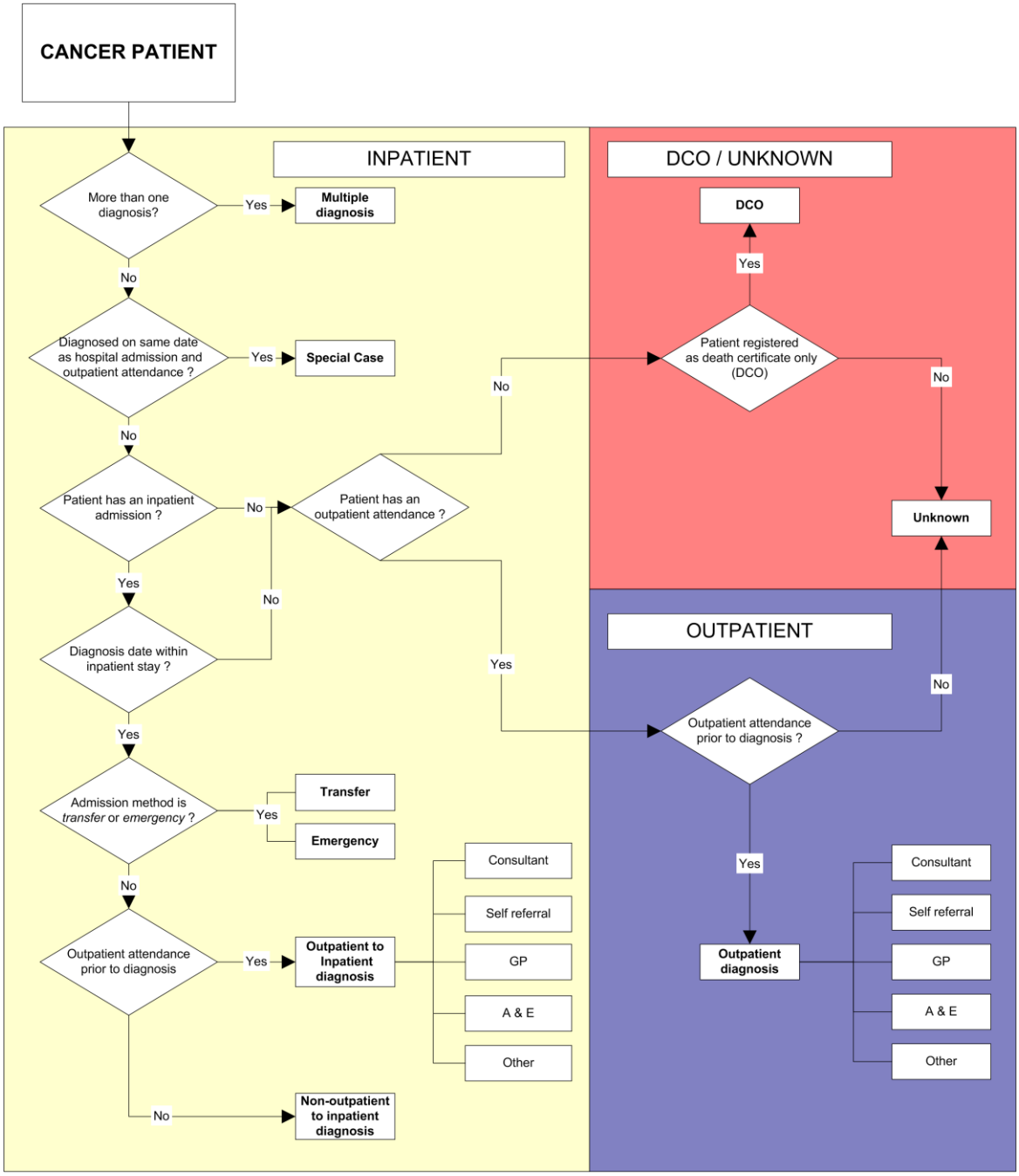
Table A1.1: Route to diagnosis codes continued

Number	Route Code	Route Group
191	OP-11-O05	Other outpatient
192	OP-11-O06	Other outpatient
193	OP-11-O07	Other outpatient
194	OP-11-O08	Other outpatient
195	OP-11-O10	Emergency presentation
196	OP-17-O12	GP referral
197	OP-17-O13	Other outpatient
198	OP-17-O17	Screening
199	OP-17-O92	Other outpatient
200	OP-17-O97	Other outpatient
201	OP-18-I99	Other outpatient
202	OP-18-O01	Emergency presentation
203	OP-18-O02	Other outpatient
204	OP-18-O03	GP referral
205	OP-18-O04	Emergency presentation
206	OP-18-O06	Other outpatient
207	OP-18-O07	Other outpatient
208	OP-18-O08	Other outpatient
209	OP-18-O10	Emergency presentation
210	OP-18-O11	Other outpatient
211	OP-18-O12	GP referral
212	OP-18-O13	Other outpatient
213	OP-18-O17	Screening
214	OP-18-O92	Other outpatient
215	OP-18-O93	Other outpatient
216	OP-18-O97	Other outpatient
217	OP-19-I99	Other outpatient
218	OP-19-O01	Emergency presentation
219	OP-19-O02	Other outpatient
220	OP-19-O03	GP referral
221	OP-19-O04	Emergency presentation
222	OP-19-O05	Other outpatient
223	OP-19-O06	Other outpatient
224	OP-19-O08	Other outpatient
225	OP-19-O10	Emergency presentation
226	OP-19-O11	Other outpatient
227	OP-19-O12	GP referral
228	OP-19-O17	Screening
229	OP-19-O92	Other outpatient
230	OP-19-O97	Other outpatient
231	SC-19-I99	Other outpatient
232	SC-19-O01	Emergency presentation
233	SC-19-O03	GP referral
234	SC-19-O04	Emergency presentation
235	SC-19-O05	Other outpatient
236	SC-19-O06	Other outpatient
237	SC-19-O08	Other outpatient
238	SC-19-O10	Emergency presentation
239	SC-19-O11	Other outpatient
240	SC-19-O92	Other outpatient
241	SC-19-O97	Other outpatient
242	SC-null-I99	Other outpatient
243	SC-null-O01	Emergency presentation
244	SC-null-O02	Other outpatient
245	SC-null-O03	GP referral
246	OP-17-O06	Other outpatient
247	OP-17-O07	Other outpatient
248	OP-17-O08	Other outpatient
249	OP-17-O10	Emergency presentation
250	OP-17-O11	Other outpatient
251	SC-null-O04	Emergency presentation
252	SC-null-O05	Other outpatient
253	SC-null-O06	Other outpatient
254	SC-null-O07	Other outpatient
255	SC-null-O08	Other outpatient
256	SC-null-O10	Emergency presentation
257	SC-null-O11	Other outpatient
258	SC-null-O12	GP referral
259	SC-null-O13	Other outpatient
260	SC-null-O17	Screening
261	SC-null-O92	Other outpatient
262	SC-null-O93	Other outpatient
263	SC-null-O97	Other outpatient
264	UN-10-UNK	Unknown
265	UN-11-UNK	Unknown
266	UN-12-UNK	Unknown
267	UN-13-UNK	Unknown
268	UN-15-UNK	Unknown
269	UN-19-UNK	Unknown

Appendix 2: Algorithm used to identify pathway codes

The diagram in Figure A2.1 illustrates the algorithm used to identify each of the pathway codes.

Figure A2.1: Flow diagram of 'routes to diagnosis'



Appendix 3: Source of first contact codes

A list of all 'source of first contact' codes is provided in Table A3.1. When source of first contact code commences with an 'I' this indicates an inpatient method of admission while an 'O' indicates an outpatient source of referral.

Table A3.1: Source of first contact codes.

Source of First Contact Code	Source of First Contact Name
DCO	DCO
I11	Elective: from waiting list
I12	Elective: booked
I13	Elective: planned
I21	Emergency: via Accident and Emergency (A&E) services, including the casualty department of the provider
I22	Emergency: via general practitioner (GP)
I23	Emergency: via Bed Bureau, including the Central Bureau
I24	Emergency: via consultant outpatient clinic
I28	Emergency: other means, including patients who arrive via the A&E department of another healthcare provider
I31	Maternity: where the baby was delivered after the mother's admission
I32	Maternity: where the baby was delivered before the mother's admission
I81	Transfer of any admitted patient from another hospital provider other than in an emergency; this does not include admissions to high security psychiatric hospitals (HSPH)
I82	Other: babies born in health care provider
I83	Other: babies born outside the health care provider, except when born at home as intended
I84	Admission by the admission panel of an HSPH; patient not entered on the HSPH admissions waiting list (not valid for admissions after 31 March 2002)
I89	From the admissions waiting list of an HSPH (not valid for admissions after 31 March 2002)
I98	Not applicable (eg other maternity event)
I99	Not known
NUL	Null
O01	Following an emergency admission
O02	Following a domiciliary visit
O03	Referral from a general medical practitioner
O04	Referral from an accident and emergency department
O05	Referral from a consultant, other than in an accident and emergency department
O06	Self referral
O07	Referral from prosthetist
O08	Other source of referral
O10	Following an accident and emergency attendance
O11	Other
O12	Referral from GP with special interest
O13	Referral from a specialist nurse (secondary care)
O14	Referral from an allied health professional
O15	Referral from an optometrist
O16	Referral from a orthoptist
O17	Referral from a national screening programme
O92	General dental practitioner
O93	Community dental service
O97	Other – not initiated by the consultant responsible for the consultant outpatient episode
O99	Not known
UNK	Unknown

Appendix 4: ICD-10 cancer site codes

Table A4.1: ICD-10 cancer site codes.

ICD-10 Code	Cancer Site
C00 - C08	Oral
C15	Oesophagus
C16	Stomach
C18 - C20	Colorectal
C25	Pancreas
C32	Larynx
C33 - C34	Lung
C43	Melanoma
C47	Brain & CNS
C50	Breast
C53	Cervix
C54 - C55	Uterus
C56	Ovary
C61	Prostate
C62	Testis
C67	Bladder
C69 - C72	Brain & CNS
C82,C83,C85	Non-Hodgkin's lymphoma
C90	Multiple myeloma
C911, C913, C914, C919, C921, C923, C927, C929, C931, C939, C947, C951, C959	Chronic leukaemia
C910, C920, C924, C925, C930, C950	Acute leukaemia
All other C codes excluding than C44	Others