

EUROCHIP-II
FINAL SCIENTIFIC REPORT
ANNEX 16

REPORT OF
EUROCHIP-2 ACTION IN
THE UNITED KINGDOM

EUROCHIP UK Pilot Study:

**Establishing by qualitative questionnaire the feasibility of
collecting population-based data into "delays in treatment"
and "compliance with guidelines" in the UK**

September 2007

Finian Bannon and Anna Gavin,
N. Ireland Cancer Registry, 2007.
Centre for Clinical & Population Sciences
Queen's University Belfast
Mulhouse Building
Grosvenor Road
BT12 6BJ Belfast
Tel: 028 9063 2571
Fax: 028 9024 8017
email: f.j.bannon@qub.ac.uk

EXECUTIVE SUMMARY

Eurochip undertakes pilot studies to determine the feasibility of improving the availability of health indicators in Europe, if possible from existing data systems. In the UK 'delays in treatment' and 'compliance with guidelines' were chosen as important areas to investigate the capture of new items of information, if possibly electronically.

Data were collected by self-completed qualitative questionnaires that canvassed the opinion of cancer registries on the quality of current and potential data sources for collecting population based data items on 'delays in treatment' and 'compliance with guidelines'. The questionnaires were changed substantially from those used by other European countries which undertook a quantitative approach to the pilot study.

This report summarises data that was compiled from all the UK registries representing a population of 59.6 million (100% of UK population).

The results showed that clinical records were considered the best source in terms of the number of items of information contained therein, and the rate at which data could be captured on the entire population of cancer patients. Clinical records were however, considered labour-intensive to access and process, and therefore resource-intensive. It was reported that *hospital administration records and/or pathological reports* could capture individual items of information as well as clinical records, but do not collect all the required data items. *Multidisciplinary team meeting records*, and *National Cancer Waiting Times Database*, though currently not providing data routinely to all cancer registries, between them could possibly provide data on most items.

The *Multi-disciplinary team* data source holds promise, not on account of its current value, but on the fact that it is set to become the normative setting for decisions to treat cancer patients. Most of the items of information relevant to *Compliance with treatment guidelines* and *Delay in cancer treatment* are supposed to be collected to facilitate joint decision making by groups of clinicians and also monitoring of the patient's progress. Most registries surveyed believed that they could set up systems to automatically receive data electronically from this source and populate their database fields, thus making it a low-expense system. In addition, in time, it is hoped, that the population coverage would rise to the level of clinical records.

The *National Cancer Waiting Times Database* was identified as having the potential to quickly populate date fields in the registry database concerning the delay in cancer treatment; it was recognised that it would be biased towards patients who are urgently referred and go through the acute services. It also includes non-cancer patients.

It was reported that more coding effort and quality control will be needed for indicators of *Compliance with treatment guidelines* than *Delays in cancer treatment* because dates are easier to work with than data items such as morphology, ICD-O sites, and staging.

It is recommended that the data from *multidisciplinary team meetings* and *National Cancer Waiting Times Database* should be further explored as potential data sources for cancer registration to facilitate the monitoring of *compliance with guidelines* and *delays in treatment* on a population basis in the UK.

ACKNOWLEDGEMENTS

Special thanks to Dr. Harry Comber, National Cancer Registry (Rep. of Ireland), who with Dr. Finian Bannon designed and documented the qualitative questionnaire.

Further thanks to all those who were involved with the qualitative questionnaire in the UK cancer registries:

Mrs. Karen Wright, Dr. David Greenberg, Ms. Jane Richardson, and Mr. Jem Rashbass (Eastern Cancer Registration and Information Centre)
Mrs. Kath Chester, Mrs. Jennifer Kennedy, and Dr. Tony Moran (North West Cancer Intelligence Service)
Mr. Michael Walkley, Ms. Sheila Pass, and Prof. David Forman (Northern & Yorkshire Cancer Registry & Information)
Dr. Richard Middleton and Mr. Colin Fox (Northern Ireland Cancer Registry)
Ms. Pat Hall, Mrs. Sandra Edwards, and Dr. Monica Roche (Oxford Cancer Intelligence Unit)
Ms. Alison McDonald, Ms. Karen Smith, and Dr. David Brewster (Scottish Cancer Registry)
Mr. Carlos Rocha, Dr. Sean McPhail, Mr. Leigh Appleby, and Dr. Julia Verne (South West Cancer Intelligence Service)
Ms. Karen Linklater, Mr. Stephen Richards, and Prof. Henrik Møller (Thames Cancer Registry)
Ms. Alexandra Thackeray, Ms. Gillian Percival, and Dr. David Meechan (Trent Cancer Registry)
Ms. Joy McRae and Dr. John Steward (Welsh Cancer Intelligence & Surveillance Unit)
Ms. Natalie Pearce, Ms. Diane Thomas, Ms. Stacey Croft, and Dr. Gill Lawrence (West Midlands Cancer Intelligence Unit)

We would like to thank the following EUROCHIP collaborators:

Dr. Anna Gregor, Eurochip UK representative and Lead Cancer Clinician for Scotland, Scottish Executive Health Department.
Professor Ian Kunkler, University Department of Clinical Oncology, Western General Hospital, Edinburgh, Scotland
Dr. Andrea Micheli, Project Leader of Eurochip, Istituto Nazionale per lo Studio e la Cura e dei Tumori, Milan.

We would also like to thank other participants of the EUROCHIP pilot studies who provided valuable feed back on the qualitative questionnaire:

Dr. Pascale Grosclaude Registre des Cancers du Tarn, Chemin des trois Tarn, F-81000 ALBI
Dr. Vesna Zadnik Epidemiology and Cancer Registry, Institute of Oncology Ljubljana, Zaloska 2, 1000 Ljubljana, Slovenia
Dr. Michael. Schaapveld, Integraal Kankercentrum Noord-Nederland, Groningen, The Netherlands

Finally, our thanks to Mr. Giulio Napolitano (N. Ireland Cancer Registry) who designed the qualitative questionnaire on Microsoft Access.

TABLE OF CONTENTS

EXECUTIVE SUMMARY	1
ACKNOWLEDGEMENTS	2
TABLE OF CONTENTS	3
SECTION 1: INTRODUCTION	4
1.1 Introduction: The scope and outcome of the UK pilot study ...	5
SECTION 2: METHODOLOGY.....	6
2.1 Methodology: Choice of questionnaire.....	6
2.2 Methodology: Questionnaire completion	6
2.3 Methodology: Analysis of results	8
SECTION 3: RESULTS.....	9
3.1 Results: Registry Details.....	9
3.2 Results: Sources of information	11
3.3 Results: Items of information	17
3.4 Results: General.....	27
SECTION 4: CONCLUSIONS.....	29
4.1 Conclusions: Recommendation	30
SECTION 5: APPENDICES.....	31
Appendix 1: Detailed Methodology.....	31
Appendix 2: Information collected on core and additional sources of information in Part 2 of the qualitative questionnaire.....	38
Appendix 3: Information collected on items of information in Part 3 of the qualitative questionnaire	39

SECTION 1: INTRODUCTION

EUROCHIP (European Cancer Health Indicators Project) is a public health programme funded by the Health and Consumer Protection Directorate-General of the European Commission.

To date EUROCHIP has successfully completed its first project (EUROCHIP-1, funded under the Health Monitoring Programme) of developing a list of health indicators designed to provide comparable information about the burden, risk factors, management and outcome of cancer, in order to facilitate cancer control across Europe. The indicators will contribute to a European Health Information system. In addition, the indicators will form the basis of projects or actions in individual countries funded by EUROCHIP-2 (see below).

The second major project (EUROCHIP-2) is to define an organisational network of collaboration that will effectively fight inequalities in cancer in Europe through improved cancer surveillance that leads to a greater understanding in health system deficiencies and the determinants of clinical outcomes. Each of the EUROCHIP National Specialist Groups¹ provided a description of the cancer control situation in their country, indicating specific deficiencies in cancer control that should be addressed. Arising from this analysis, each of the National Specialist Groups put forward a project or action that would address one or two of these specific deficiencies.

Discussions in the UK National Specialist Group identified cancer treatment as a priority area for action. The UK agreed to participate in the EUROCHIP Pilot Study, a feasibility study into collection of population-based data on *delays in cancer treatment* and *compliance with cancer treatment guidelines*.

Several other EU countries carried out similar work, which will allow comparability and help establish whether these indicators are associated with the large variations in cancer survival existing between countries; this also fulfils an important EUROCHIP value, that the actions in individual countries will benefit the entire community of European countries.

¹ National groups of specialists were set up by the members of the “Panel of Experts” and consisted of groups at national level which discussed indicators from a national angle. Panel of experts included one expert for each EU member, and experts from cancer institutions and the major European cancer networks (IARC, EBCN, Cervix Network, EUROCARE, EUROPREVAL, OECD, ENCR, and NCI from US).

The pilot study is a feasibility study to identify relevant data systems that could supply EUROCHIP health indicator information and answer the questions:

1. Is the information already available?
2. If so from what sources, and in what form could it be retrieved?
Would the information be complete and reliable?
3. What steps would be needed in order to collect indicator information for the entire population?
4. How much would this cost?

In summary, there are two important criteria involved in the pilot study

- 1) **where can indicator information be found?**
- 2) **how much resource will it cost to collect it?**

When the data are analysed, a recommendation to the European Union on cost-effective methods of collecting quality information on “cancer waiting times” and “compliance with cancer treatment guidelines” at the population level in the EU will be made. The information is also likely to be useful to the host country, in this case the UK, in locally addressing the particular study questions.

1.1 Introduction: The scope and outcome of the UK pilot study

The study considers some of the health indicators defined in EUROCHIP-1 on *compliance with guidelines* and *delays in cancer treatment* (see Appendix 1: Methodology: Items of information). Furthermore, the study investigates two cancer sites – breast and colorectal. The geographical area considered is the entire UK (Scotland, Northern Ireland, England and Wales), with UK cancer registries – eleven in all – supplying the information.

The outcome of the study would be a report for EUROCHIP that will recommend a cost efficient way of collecting population data in the UK for the health indicators assessed in the report.

SECTION 2: METHODOLOGY

2.1 Methodology: Choice of questionnaire

There was an evolution of thinking on the methodology of the UK pilot study. The pilot project chosen by the UK was shared by other countries in the EU, and a quantitative questionnaire was designed centrally, by EUROCHIP, for use by all participants. It was termed quantitative because it involved the random sampling of registered breast and colorectal cancer patients in each of the National groups, and ascertaining for these patients whether certain of the indicators on *delays in cancer treatment* and *compliance with guidelines*, existed on these patients and in what data sources. The quantitative questionnaire was piloted in the UK in June 2006 among all registries, of whom 9 from 11 responded. However, on processing the results it became apparent that a number of problems had arisen. First of all, for the larger registries, its completion of the data searches required a lot of work. Secondly, there was confusion as to the interpretation of questions. For instance, most registries considered the exercise an audit of current practice with the effect that the feasibility aspect of the project was lost. Finally, in the UK there is a wide diversity in how registries operate and the questionnaire didn't allow enough flexibility for registries to express how they collect data.

The problems encountered by the quantitative questionnaire's piloting in the UK were recognized in a meeting in Milan in September 2006. It was decided that, given the resources available in the pilot study and the wide diversity in the operation of UK registries, it would be prudent to achieve the objectives of the study through the use of a new qualitative questionnaire. The qualitative questionnaire would not involve a random sample of cancer patients as with the quantitative questionnaire, but would concentrate on a description of the sources, both currently and potentially, that contain information on the desired indicators. In addition, for each individual indicator, e.g. date of referral, all sources that carry this information would be assessed for completeness, reliability, and coding effort. The questionnaire would be designed in an Access database setting, with the data entered directly.

The new questionnaire (See Appendix 1 for detail) was developed by Dr Finian Bannon, N. Ireland cancer registry, the host registry for the UK Eurochip project, and Dr Harry comber National Cancer Registry of Ireland with comments from Prof. Ian Kunkler, Dr. Pascale Grosclaude, Dr. Vesna Zadnik and Dr. Michael Schaapveld.

2.2 Methodology: Questionnaire completion

The qualitative questionnaire was completed by registry staff from existing knowledge of current practice. This may have been only one person—a

registry manager or director or someone equivalent— or a number of staff with different areas of expertise, who have a detailed knowledge of the availability and quality of collected data. The “potentially collectable” sources were known to registries from their general knowledge of their current health systems and the developments which were likely to happen in the next few years and also from their experience in collecting data for non-routine audits or studies; the qualitative questionnaire tapped into this knowledge as well as that of data sources routinely used by the registry. For some sources or items of information where uncertainty existed, the registry carried out a sampling exercise to check the quality, completeness or availability of data, but this was not often required.

There were separate questionnaires for the two sites that were studied: breast (ICD10 codes: C50) and colorectal (ICD10 codes: C18-21). The questionnaires were supplied as an Access database, and filled out once per registry for each site.

The questionnaire was in three parts:

1 Registry details. Basic information was requested in this section on the registry, person filling out the questionnaire, relevant cancer incidence figures, and the current annual budget devoted to cancer registrations.

2 Sources of information. The first step in the qualitative questionnaire was to assess all sources of information, readily available or potentially available, independent of the effort or resources needed to access the data or collect it. As an example of “potentially available sources”, England’s National Health Service (NHS) had recently established a *National Cancer Waiting Times Database* which can, in theory, supply information to cancer registries on “cancer waiting times”. Most registries had a good idea of the quality of these sources. For those sources that would require extra resources to collect, the questionnaire quantifies this as a percentage of the current registry budget. There was a core set of sources that the registries were obliged to address, they are listed in Appendix 1.

3 Items of information. A list of key items of information or indicators relevant to “compliance with cancer treatment guidelines” (e.g. type of surgery), and “cancer waiting times” (e.g. date of first request for clinic/hospital appointment), were studied for the individual sources enumerated previously, and the quality of the indicator information in those sources—completeness, reliability, and coding format—was assessed. For items of information for which knowledge was uncertain, the registry could carry out a sampling exercise to check completeness of the data, but this was not obligatory.

Appendix 1 gives a detailed account of the methodology. Appendices 2 & 3 contain two tables which show the conceptual structure of the information sought in Parts 2 & 3, respectively.

2.3 Methodology: Analysis of results

The statistics on sources and items of information are weighted means of those registries that contributed to that statistic; the weights are directly proportional to the population size of each registry, and the number of registries that contributed to a statistic is in brackets next to the statistic. As social insurance records are unavailable in the UK, registries had no experience of them; therefore, it was decided to omit them from the analysis that follows.

The index of availability of a source of information is a measure of the mean reported availability of a source (across N cancer registries) where the following types of availability are scored as follows:

- 1 Easily available and routinely collected
- 2 Easily available and not routinely collected
- 3 Available with extra resources
- 4 Not available at present

The mean score, over the n contributing registries, gives an index between 1 and 4. The ranking of these results is the important output of this analysis, however any quantitative comparison between the indices of two sources is meaningless.

The index of processing ability of a source of information is a measure of how easily the source can be processed, it is calculated as above with the following scores:

- 1 Electronically received and automatically processed
- 2 Electronically received and manually processed
- 3 Paper record received
- 4 Paper record accessed

The index of processing ability has some overlap with index of availability, but focuses more on the practicalities of registration. The ranking of the sources is the important output of this analysis.

SECTION 3: RESULTS

The results, like the questionnaire, are presented in three parts:

Part 1: Registry Details (includes colorectal incidence rate)

Part 2: Analysis of Sources

Part 3: Analysis of Items of Information

3.1 Results: Registry Details

Questionnaires were received from all UK registries. Table 1 summarises the statistics that each had on population served, incidence, and most recent incidence complete year.

Table 1: Information about the UK registries: Population, incidences, and most recent incidence year collected.

Registry name (population in million)	Breast cancer incidence	Colorectal cancer incidence	Most recent incidence year
Eastern Cancer Registration and Information Centre (5.3)	4314	3217	2004
Northern Ireland Cancer Registry (1.7)	1119	954	2004
North West Cancer Intelligence Service (6.5)	4925	4200	2005
Northern & Yorkshire Cancer Registry & Information (6.6)	4900	4200	2004
Oxford Cancer Intelligence Unit (2.8)	2180	1532	2005
South West Cancer Intelligence Service (6.7)	6315	5176	2004
Scottish Cancer Registry (5.0)	3948	3555	2004
Thames Cancer Registry (12.0)	7555	5633	2005
Trent Cancer Registry (4.8)	3818	3156	2005
Welsh Cancer Intelligence & Surveillance Unit (3.0)	2390	2024	2005
West Midlands Cancer Intelligence Unit (5.3)	4424	3233	2005

EUROCHIP-2 Final Report – Annex 16 – EUROCHIP-2 in UK

Table 2 summarises some of the data on the registries. The registration expenditure per million of population in the UK by cancer registries, is estimated by dividing each registry's budget by its population and aggregating these estimates from all registries weighted according to population size. The registration expenditure per million of population is over-estimated as some registries couldn't exclude from their budget figure, expenditure on audit work and/or research. The cost per case is similarly derived, in this case the budget is divided by the incidence cases per annum

Table 2: UK average registration budget in Euro per million, and per case. Crude UK statistics for breast and colorectal and total cancer incidence derived from Table 1

Registration budget per million population in €.	Cost/case in €.	Breast cancer incidence per 100,000*	Colorectal cancer incidence per 100,000*	All cancers (C00-C97) incidence per 100,000
277,000	51	77	62	585

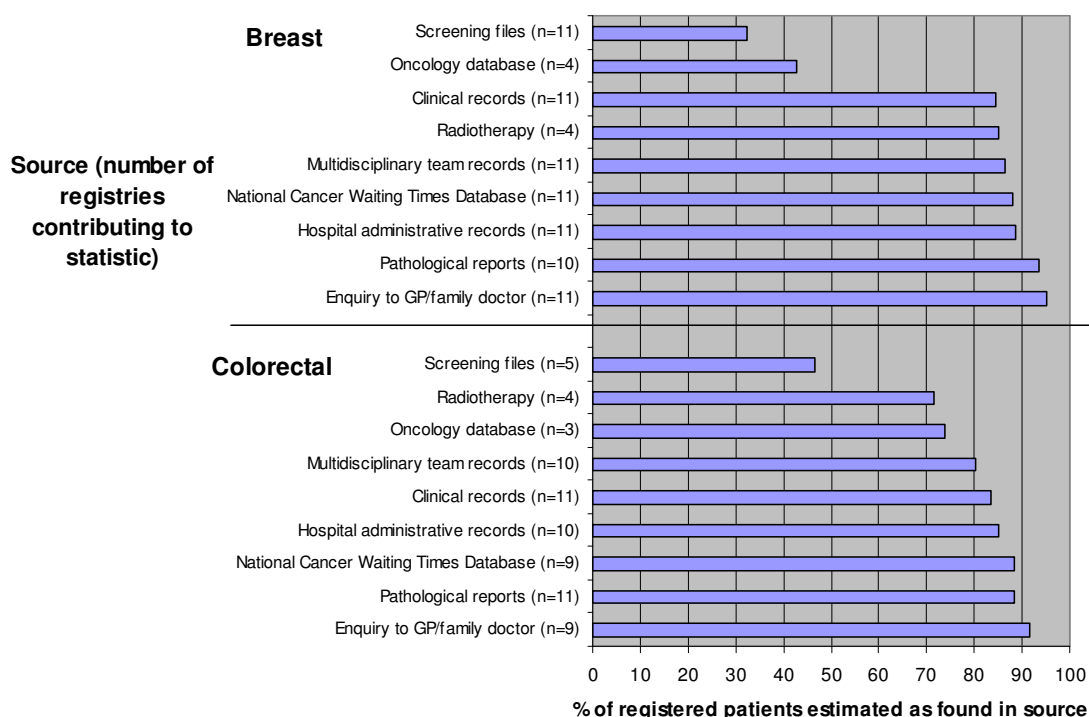
*These statistics are crude as they rely on incidence figure from different 'most recent' incidence years.

3.2 Results: Sources of information

Percentage of registered patients in each source

Figure 1 displays, for each source and cancer site, the percentage of registered patients that registries indicated would be expected to be found in that source. In general, registries considered that most sources have good coverage with only two sources, *screening files* and *oncology databases*, with less than 80% for both cancer sites (Figure 1). Colorectal also has *radiotherapy database* lower than 80%. *Enquiry to GP/family doctor* exceeds 90% in both sites.

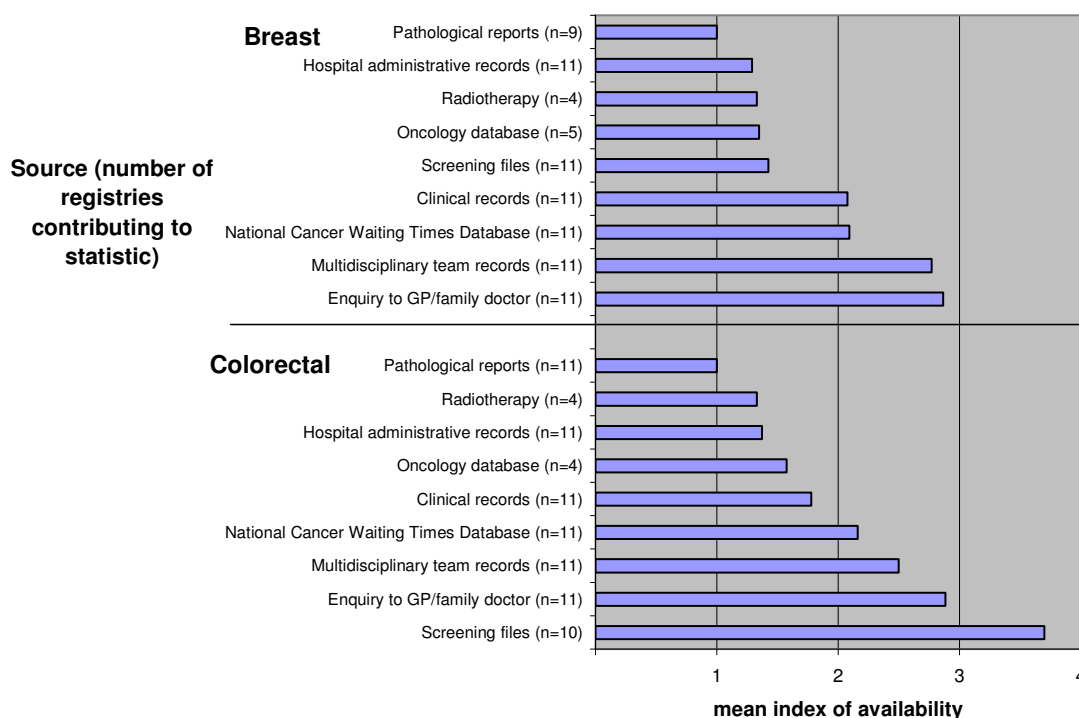
Figure 1: The mean percentage of registered breast and colorectal cancer patients reported to have a record in different sources.



Index of availability

*Pathological reports, oncology database, and radiotherapy sources were ranked either 1st, 2nd, and 3rd in both sites, highlighting the fact that most registries can acquire them without extra resources (Figure 2). Enquiry to GP/family doctor was ranked highly, showing that a number of registries don't routinely collect this source presently in their entirety. Sources that exist but are being currently explored by registries scored higher ranks *multidisciplinary team meeting, National Cancer Waiting Times Database, and screening files (colorectal only)* reflecting their perceived unavailability.*

Figure 2: Reported index of availability of information for breast and colorectal cancer sites from different sources averaged over the number (n) of registries that reported they can/could collect them



Index of processing ability

Figure 3 shows that, for both sites, registries felt that *enquiry to GP/family doctor* and *clinical records* would require the most processing with *pathological reports* and *multidisciplinary team meeting* coming next. *National Cancer Waiting Times Database*, *hospital administrative records*, *oncology database* and *screening files* all were ranked in the top 4, reflecting their potential for automatic input to the cancer registry database. Differences in the ranking of sources between the index of processing ability and the index of availability for sources *National Cancer Waiting Times Database* and *screening files* [colorectal only] reflect the reality that registries are only beginning to use these sources and require additional resources to capture data. In the UK, breast cancer screening is population-based while colorectal cancer screening is only recently available in some areas.

Table 3 below shows how records from different sources are processed, on average, for all the cancer registries. As expected, *clinical records* and *enquiry to GP/family doctor* both are largely registered by accessing paper records. *Pathological reports*, *hospital administrative records*, *screening files* [breast only] and *multidisciplinary team meeting* have significant percentages of their processing not carried out electronically and automatically. *National Cancer Waiting Times Database* and *screening files* [colorectal only] have an estimated 80%+ of their registrations processed electronically and automatically.

Figure 3: Index of processing ability for breast and colorectal cancer sites for different sources averaged over the number (n) of registries that can/could collect them

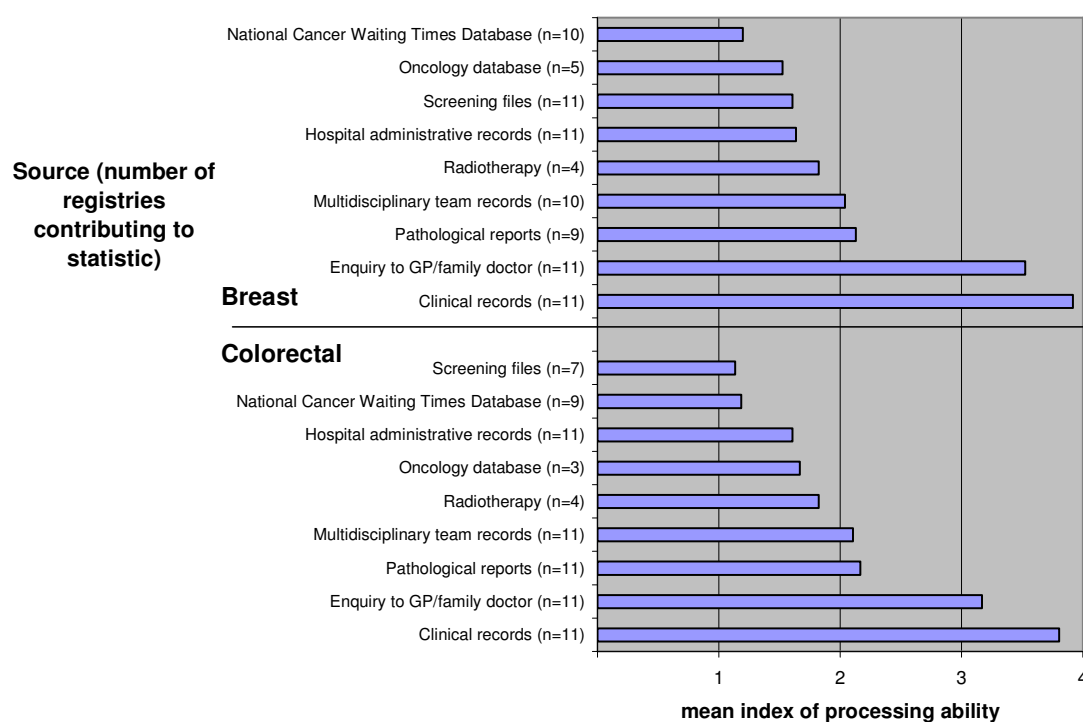


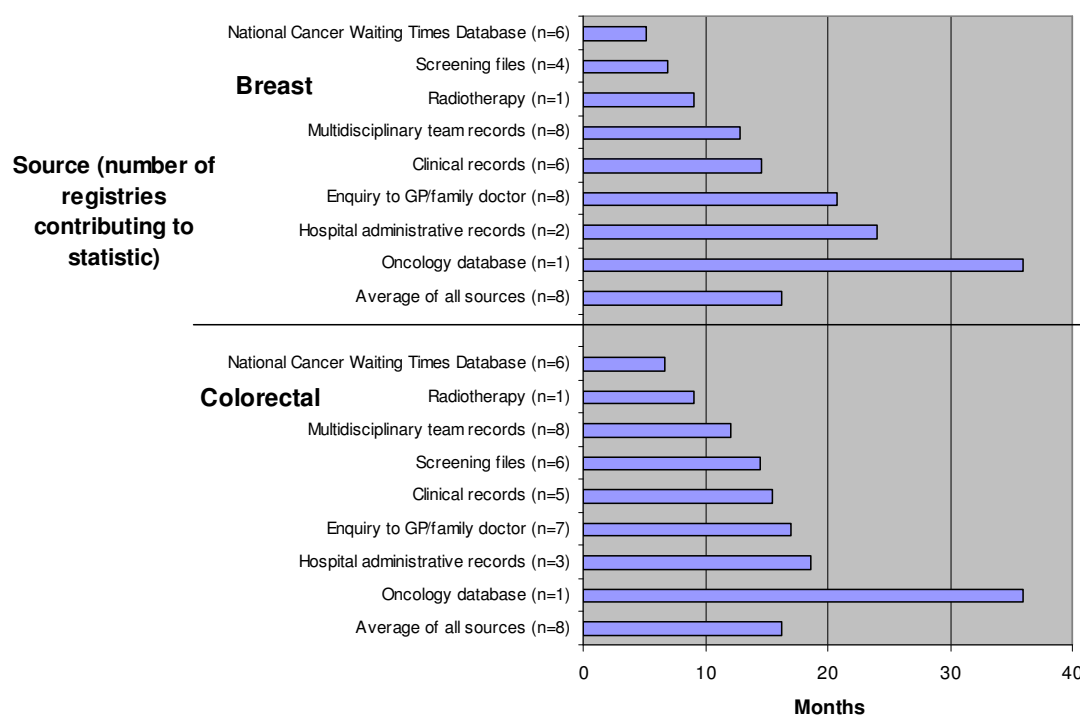
Table 3: Paper versus electronic registration methods for data sources averaged over registries for breast and colorectal cancer sites.

	% of records processed by this method			
	Paper record accessed by registry	Paper record received by registry	Electronically received and manually processed	Electronically received and automatically processed
Source (number of registries contributing)				
<i>Breast</i>				
Clinical records (11)	94.5	2.8	2.7	0.0
Enquiry to GP/family doctor (11)	62.2	19.6	18.2	0.0
Hospital administrative records (11)	2.3	1.8	41.2	54.8
Multidisciplinary team records (10)	21.0	1.0	46.5	31.5
Nat. cancer waiting times database (10)	0.0	0.0	19.0	81.0
Oncology database (4)	0.0	20.0	24.8	55.3
Pathological reports (10)	1.0	18.5	60.0	20.5
Radiotherapy (4)	0.0	27.5	0.5	72.0
Screening files (11)	0.0	9.2	45.4	45.5
<i>Colorectal</i>				
Clinical records (10)	91.0	0.0	9.0	0.0
Enquiry to GP/family doctor (10)	58.6	21.4	20.0	0.0
Hospital administrative records (11)	2.3	0.0	40.7	57.0
Multidisciplinary team records (10)	18.0	10.0	32.0	40.0
National cancer waiting times database (9)	0.0	0.0	20.0	80.0
Oncology database (3)	0.0	26.7	1.7	71.7
Pathological reports (11)	0.5	23.6	57.3	18.6
Radiotherapy (4)	0.0	27.5	0.5	72.0
Screening files (7)	0.0	0.0	14.4	85.6

Time needed to collect new source

Registries that do not routinely collect from a specific source were asked to estimate how long it would take to do so, if they had the sufficient resources. Considerations in this estimate would be recruitment and training of new staff, procurement of new hardware, obtaining permission from data owners. When interpreting Figure 4 give weight to statistics that more registries contributed to.

Figure 4: The estimated number of months required to start collecting data from sources that are not currently collected by registries.



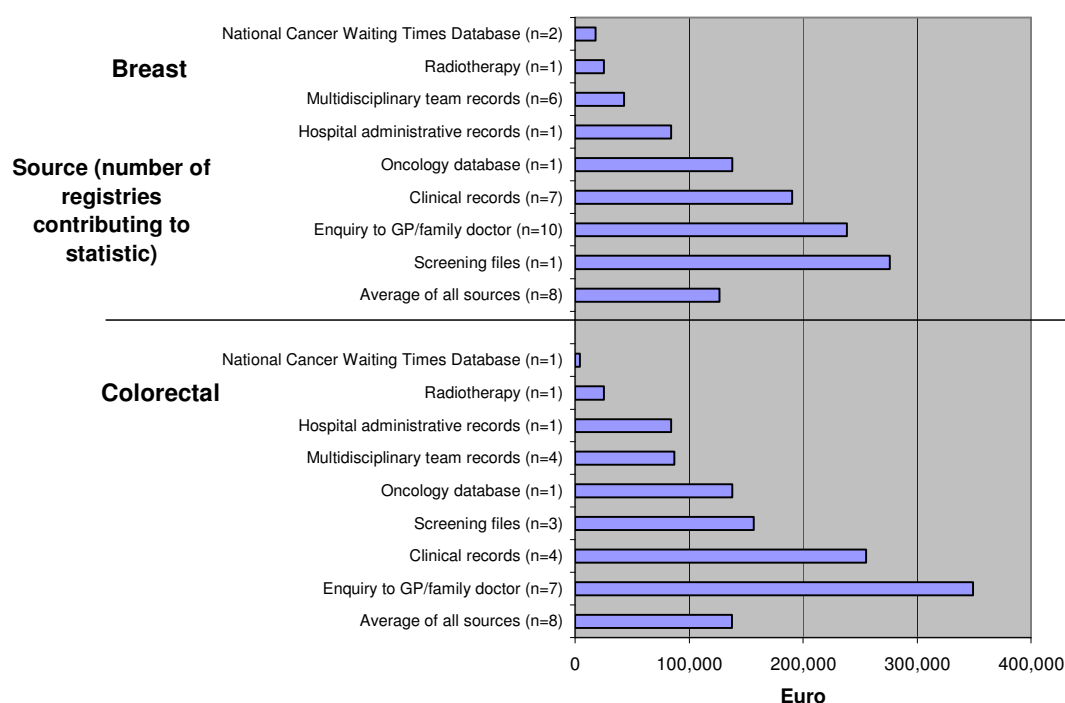
It was reported by the registries that *enquiry to GP/family doctor* would take longer to achieve than *clinical records* (19 versus 15 months approximately), because more staff would need to be trained to cover the wider dispersion of GP practices than acute trusts.

National Cancer Waiting Times Database and *radiotherapy* were sources that could be collected within 10 months, whereas *hospital administrative records*, *multi-disciplinary team meeting* data, and *clinical records* would take between 10-20 months. The overall average time that a registry would need to collect a new source was estimated to be around 16 months.

Extra resources required to collect new sources per million of population

In part 2 of the questionnaire, where sources are assessed, option 3 & 4 of the 'availability' section of source mean that extra resources are required if the registry were to set about collecting that source of information. The registries estimated a figure that was proportional to their current budget (which is a measure of the activity that the registry carries out currently) bearing in mind the source to be collected. The statistic is standardised to extra cost per million of population. When interpreting Figure 5 give weight to statistics that more registries contributed to.

Figure 5: The extra cost to collect a source per million of the population in the UK.



A number of registries felt they would require extra resources to collect routinely *clinical records* and *enquiry to GP/family doctor* estimated as ranging between €190,000 and €350,000 per million population (Figure 5). Smaller sums would be required (<€100,000 per million) for sources that didn't require paper access i.e. *National Cancer Waiting Times Database*, *multidisciplinary team meeting*, *radiotherapy* and *hospital administrative records*.

3.3 Results: Items of information

Each item of information for each source was assessed for ‘% actual record’ (see below), reliability, and coding effort. Table 4 & 5 presents ‘% actual record’ (for breast and colorectal) which combine the % of registered patients and completeness in order to give an insight into the % of patients who will have an item of information in a particular source. Tables 6 & 7 present the reliability scores for breast and colorectal items of information. Likewise Tables 8 & 9 present the coding effort scores of the items of information for the two cancers. The tables are presented first, and comments made afterwards. We will employ the term ‘GUIDELINES’ to signify *compliance with treatment guidelines* and ‘DELAY’ to mean *delay in cancer treatment*.

EUROCHIP-2 Final Report – Annex 16 – EUROCHIP-2 in UK

Percentage actual record

Table 4: Percentage of breast cancer patients whom it was estimated would have an item of information actually recorded in different sources of information (number of registries that contributed to this statistic [†]).										
Items of information	Sources*									Item mean
	CLIN	GP	HAR	MDT	WAITS	ONC	PATH	RAD	SCN	
<i>Compliance with treatment guidelines</i>										
Morphology	79 (11)	63 (9)	66 (4)	80 (9)	84 (2)	37 (4)	91 (10)	90 (1)	32 (7)	69
ICD-O	94 (8)	77 (10)	85 (8)	81 (9)	86 (6)	50 (2)	93 (8)	87 (2)	31 (5)	76
Pathological TNM	60 (9)	41 (6)	20 (1)	69 (8)	0 (0)	32 (2)	66 (9)	90 (1)	37 (1)	52
Clinical TNM	54 (8)	37 (6)	40 (1)	41 (6)	0 (0)	16 (1)	28 (3)	86 (1)	23 (1)	41
Method of diagnosis (if microscopic)	87 (8)	56 (7)	68 (4)	74 (7)	0 (0)	46 (2)	94 (7)	72 (1)	33 (2)	66
Method of diagnosis (if radiological)	87 (8)	61 (6)	59 (3)	65 (6)	68 (1)	34 (2)	5 (1)	85 (2)	20 (1)	54
Method of diagnosis (if clinical)	90 (8)	53 (7)	79 (5)	60 (6)	68 (1)	46 (2)	0 (0)	86 (1)	8 (2)	61
Type of surgery	88 (8)	73 (7)	83 (7)	81 (8)	60 (1)	47 (2)	87 (7)	86 (1)	33 (2)	71
Source mean	80	58	63	69	73	39	66	85	27	
<i>Delays in cancer treatment</i>										
Date of first visit with general practitioner	64 (4)	92 (7)	0 (0)	36 (2)	0 (0)	3 (1)	0 (0)	0 (0)	0 (0)	49
Date of first request for clinic/hospital appointment	85 (7)	92 (8)	70 (2)	60 (6)	87 (4)	3 (1)	0 (0)	0 (0)	0 (0)	66
Date of first clinic/hospital appointment	88 (8)	88 (8)	89 (3)	84 (4)	80 (7)	24 (1)	0 (0)	0 (0)	20 (1)	68
Date of definitive diagnosis	89 (7)	70 (6)	68 (2)	80 (6)	83 (3)	35 (2)	93 (7)	0 (0)	37 (2)	69
Date of surgery	89 (8)	74 (6)	84 (7)	83 (8)	81 (3)	43 (2)	89 (9)	45 (1)	32 (2)	69
Date of Chemotherapy	92 (8)	82 (6)	75 (5)	72 (7)	85 (1)	38 (3)	19 (1)	45 (1)	10 (1)	58
Date of Radiotherapy	93 (8)	82 (6)	80 (4)	67 (7)	85 (1)	53 (3)	0 (0)	85 (4)	23 (1)	71
Date of Endocrine treatment	91 (8)	83 (6)	81 (4)	81 (8)	85 (1)	35 (2)	0 (0)	81 (1)	30 (1)	71
Source mean	86	83	78	70	84	29	67	64	25	
*CLIN = Clinical records, GP = Enquiry to GP/family doctor, HAR = Hospital administrative records, MDT = Multidisciplinary team records, WAITS = National Cancer Waiting Times Database, ONC = Oncology database, PATH = Pathological reports, RAD = Radiotherapy, SCN = Screening files										
[†] Any statistic that has only one registry contributing to it, i.e. (1), represents only one experience and should not be universalised										

EUROCHIP-2 Final Report – Annex 16 – EUROCHIP-2 in UK

Table 5: Percentage of colorectal cancer patients whom it was estimated would have an item of information actually recorded in different sources of information (number of registries that contributed to this statistic [†]).										
Items of information	Sources*									Item mean
	CLIN	GP	HAR	MDT	WAITS	ONC	PATH	RAD	SCN	
<i>Compliance with treatment guidelines</i>										%
Morphology	83 (10)	61 (6)	56 (4)	71 (8)	72 (1)	61 (3)	88 (10)	90 (1)	15 (1)	66
ICD-O	92 (6)	80 (7)	91 (7)	68 (7)	86 (5)	56 (1)	88 (8)	87 (2)	25 (1)	75
pTNM	65 (9)	38 (5)	51 (2)	62 (7)	0 (0)	35 (1)	67 (9)	90 (1)	0 (0)	58
pTNM Duke's stage	75 (7)	36 (5)	23 (1)	61 (7)	0 (0)	35 (1)	69 (9)	86 (1)	0 (0)	55
Clinical TNM	56 (7)	33 (5)	47 (1)	37 (6)	0 (0)	14 (1)	19 (2)	86 (1)	0 (0)	42
Clinical TNM Duke's Stage	40 (5)	32 (3)	0 (0)	46 (3)	0 (0)	0 (0)	19 (1)	86 (1)	0 (0)	45
Method of diagnosis (if microscopic)	86 (7)	45 (6)	71 (3)	64 (7)	0 (0)	56 (1)	89 (8)	72 (1)	0 (0)	69
Method of diagnosis (if radiological)	87 (8)	55 (5)	78 (1)	54 (7)	68 (1)	35 (1)	5 (1)	71 (2)	0 (0)	57
Method of diagnosis (if clinical)	88 (8)	39 (6)	87 (3)	51 (6)	68 (1)	49 (1)	0 (0)	86 (1)	0 (0)	67
Type of surgery	84 (8)	71 (6)	84 (8)	70 (8)	60 (1)	53 (1)	80 (8)	86 (1)	0 (0)	74
Source mean	76	49	65	58	71	44	58	84	20	
<i>Delays in cancer treatment</i>										
Date of first visit with general practitioner	64 (4)	92 (8)	0 (0)	36 (3)	92 (2)	0 (0)	0 (0)	0 (0)	0 (0)	71
Date of first request for clinic/hospital appointment	84 (7)	89 (8)	74 (2)	63 (6)	88 (5)	0 (0)	0 (0)	0 (0)	0 (0)	80
Date of first clinic/hospital appointment	87 (8)	85 (8)	92 (4)	73 (5)	78 (6)	0 (0)	0 (0)	0 (0)	20 (1)	73
Date of definitive diagnosis	89 (8)	58 (6)	78 (1)	72 (7)	80 (3)	35 (1)	87 (7)	0 (0)	0 (0)	71
Date of surgery	89 (8)	73 (6)	87 (7)	76 (8)	83 (4)	49 (1)	84 (7)	45 (1)	0 (0)	73
Date of chemotherapy	92 (8)	77 (6)	78 (5)	66 (8)	85 (1)	40 (2)	19 (1)	45 (1)	0 (0)	63
Date of radiotherapy	92 (8)	78 (6)	83 (4)	64 (8)	85 (1)	61 (2)	0 (0)	71 (4)	0 (0)	76
Source mean	85	79	82	64	84	46	63	54	20	
*CLIN = Clinical records, GP = Enquiry to GP/family doctor, HAR = Hospital administrative records, MDT = Multidisciplinary team records, WAITS = National Cancer Waiting Times Databas, ONC = Oncology database, PATH = Pathological reports, RAD = Radiotherapy, SCN = Screening files										
[†] Any statistic that has only one registry contributing to it, i.e. (1), represents only one experience and should not be universalised										

Tables 4 & 5 presents ‘% actual record’ which estimates the percentage of registered patients who have an item of information as actually recorded in a source; it combines both the % of registered patients that would have a record in a source and, given that a patient has a record, the completeness of that record for a particular item of information. So if 80% of registered breast cancer patient in any one incidence year have an oncology record, and oncology records are 90% complete for a certain item of information, then, the rate of finding that item of information for all registered patients in that source is $0.80 \times 0.90 = 0.72$, or 72%. The number of registries that contribute to a statistic is in brackets next to the statistic; the higher this figure the more reliable the statistic is. The statistic is a weighted mean directly proportional the population of the registry. This statistic is a knowledge-based opinion, not a quantified figure. Lack of discussion on *oncology database*, *radiotherapy*, and *screening* is because very few registries contributed information on these sources, thus they are less reliable than other sources.

Clinical records seems to be the most valuable source for both the % of patients it captures information on and the variety of items of information it captures. As it is shown (Figure 3) to require the most processing and one of the more costly to collect (Figure 5), it is interesting to see whether other [cheaper] sources can substitute effectively in collecting certain items of information.

In general terms, *pathological reports* provided good % actual record on some items of information for GUIDELINES to the standard of *clinical records*: ICD-O, method of diagnosis (if microscopic), date of diagnosis, type of surgery, date of surgery. As *pathological reports* are routinely collected by most registries it holds a lot of promise, even though they require some manual ‘reading’ to resolve cases.

The % actual record for *multidisciplinary team* data was lower than *clinical records*, though better in the breast cancer site than colorectal, perhaps reflecting a better current level of organisation. Currently, however, *multidisciplinary team* is not totally automated (Table 3), and therefore would require some work to collect at a population level. The *multidisciplinary team* data collected a similar breadth of items of information in both GUIDELINES and DELAY as *clinical records*.

The *National Cancer Waiting Times Database* performed quite well in capturing the DELAY, however registry contribution to information on this source was low, possibly because it is a relatively new data source. In fact, the estimates for the *National Cancer Waiting Times Database* and the *multidisciplinary team* data sources could be underestimates of their true worth because there was a lack of knowledge among the database managers.

Enquiry to GP/family doctor proved the best source of information on patient first visit to GP; in general, this source performed better in DELAY than GUIDELINES. The *hospital administrative records* also provide valuable information for both some GUIDELINES and DELAY items of information; most registries are collecting this source already and its potential is

EUROCHIP-2 Final Report – Annex 16 – EUROCHIP-2 in UK

therefore already well realised. *Hospital administration records* were identified as matching % actual record in *clinical records* for data on ICD-O, surgery, date of first clinic/hospital appointment and date of surgery.

Looking at items of information across all sources, in breast cancer *pathological and clinical TNM, Method of diagnosis (if radiological) and Date of first visit with general practitioner* had a % actual record mean of less 55%. In colorectal, *pTNM Duke's stage, clinical TNM, and clinical TNM Duke's Stage* had a % actual record of 55% or lower; the lowest of DELAY was *Date of chemotherapy* at 63%.

Reliability

Table 6: Estimated reliability of items of information from 1 (poor) to 5 (good) in different sources of information for breast cancer (number of registries that contributed to this statistic[†]).

Items of information	Sources*									Item mean
	CLIN	GP	HAR	MDT	WAITS	ONC	PATH	RAD	SCN	
Compliance with treatment guidelines										
Morphology	4.6 (11)	3.6 (9)	4.1 (4)	4.9 (9)	3.6 (2)	3.7 (5)	5.0 (12)	5.0 (1)	4.2 (7)	4.3
ICD-O	4.1 (8)	3.9 (10)	4.2 (8)	4.3 (9)	3.8 (6)	4.5 (2)	4.9 (10)	5.0 (2)	4.4 (6)	4.3
Pathological TNM	3.9 (9)	3.5 (6)	5.0 (1)	4.0 (8)	0.0 (0)	3.5 (2)	4.4 (10)	5.0 (1)	5.0 (1)	4.3
Clinical TNM	3.5 (8)	3.8 (6)	2.0 (1)	4.1 (6)	0.0 (0)	4.0 (1)	4.1 (3)	4.0 (1)	5.0 (1)	3.8
Method of diagnosis (if microscopic)	4.6 (8)	4.1 (7)	4.2 (4)	4.4 (7)	0.0 (0)	4.5 (2)	5.0 (9)	4.0 (1)	4.6 (2)	4.4
Method of diagnosis (if radiological)	4.6 (9)	3.6 (6)	4.7 (3)	4.8 (6)	4.0 (1)	4.5 (2)	4.0 (1)	4.3 (2)	4.0 (1)	4.3
Method of diagnosis (if clinical)	4.5 (9)	3.8 (8)	4.3 (5)	4.7 (6)	4.0 (1)	4.5 (2)	4.0 (1)	4.0 (1)	4.4 (2)	4.2
Type of surgery	4.6 (9)	4.3 (7)	4.1 (7)	4.6 (8)	3.0 (1)	4.5 (2)	4.6 (9)	5.0 (1)	4.6 (2)	4.4
Source mean	4.3	3.8	4.1	4.5	3.7	4.2	4.5	4.5	4.5	
Delays in cancer treatment										
Date of first visit with general practitioner	4.7 (4)	4.7 (7)	0.0 (0)	3.3 (2)	0.0 (0)	4.0 (1)	0.0 (0)	0.0 (0)	0.0 (0)	4.2
Date of first request for clinic/hospital appointment	4.7 (7)	4.8 (8)	5.0 (2)	4.6 (6)	4.7 (4)	4.0 (1)	0.0 (0)	0.0 (0)	0.0 (0)	4.6
Date of first clinic/hospital appointment	4.7 (9)	4.6 (8)	5.0 (3)	4.5 (4)	4.6 (7)	5.0 (1)	0.0 (0)	0.0 (0)	5.0 (1)	4.8
Date of definitive diagnosis	4.6 (8)	4.0 (7)	4.3 (2)	4.6 (6)	4.4 (3)	4.0 (2)	5.0 (8)	0.0 (0)	5.0 (2)	4.5
Date of surgery	4.7 (9)	4.1 (7)	4.6 (7)	4.6 (8)	4.4 (3)	4.0 (2)	4.7 (10)	5.0 (1)	4.6 (2)	4.5
Date of Chemotherapy	4.4 (9)	4.3 (6)	4.2 (5)	4.5 (7)	5.0 (1)	4.6 (3)	5.0 (2)	5.0 (1)	5.0 (1)	4.7
Date of Radiotherapy	4.6 (9)	4.4 (6)	4.5 (4)	4.7 (7)	5.0 (1)	4.6 (3)	0.0 (0)	4.8 (4)	5.0 (1)	4.7
Date of Endocrine treatment	4.4 (9)	4.2 (7)	4.7 (4)	4.6 (8)	5.0 (1)	4.5 (2)	0.0 (0)	5.0 (1)	5.0 (1)	4.7
Source mean	4.6	4.4	4.6	4.4	4.7	4.3	4.9	5.0	4.9	

*CLIN = Clinical records, GP = Enquiry to GP/family doctor, HAR = Hospital administrative records, MDT = Multidisciplinary team records, WAITS = National Cancer Waiting Times Database, ONC = Oncology database, PATH = Pathological reports, RAD = Radiotherapy, SCN = Screening files
[†]Any statistic that has only one registry contributing to it, i.e. (1), represents only one experience and should not be universalised

EUROCHIP-2 Final Report – Annex 16 – EUROCHIP-2 in UK

Table 7: Estimated reliability of items of information from 1 (poor) to 5 (good) in different sources of information for colorectal cancer (number of registries that contributed to this statistic[†]).

Items of information	Sources*									Item mean
	CLIN	GP	HAR	MDT	WAITS	ONC	PATH	RAD	SCN	
<i>Compliance with treatment guidelines</i>										%
Morphology	4.5 (10)	3.3 (7)	4.1 (4)	4.6 (8)	3.0 (1)	4.1 (4)	4.9 (10)	5.0 (1)	3.0 (1)	4.1
ICD-O	4.1 (7)	3.8 (8)	4.2 (7)	4.5 (7)	4.0 (5)	4.5 (2)	4.9 (9)	5.0 (2)	5.0 (1)	4.4
pTNM	4.1 (9)	3.5 (5)	5.0 (2)	4.0 (7)	0.0 (0)	3.5 (2)	4.6 (10)	5.0 (1)	0.0 (0)	4.2
pTNM Duke's stage	4.4 (7)	3.8 (5)	5.0 (1)	4.2 (7)	0.0 (0)	4.0 (2)	4.4 (9)	5.0 (1)	0.0 (0)	4.4
Clinical TNM	3.6 (7)	3.8 (5)	2.0 (1)	4.0 (6)	0.0 (0)	4.0 (1)	4.3 (2)	4.0 (1)	0.0 (0)	3.7
Clinical TNM Duke's Stage	3.5 (5)	4.1 (3)	0.0 (0)	5.0 (3)	0.0 (0)	0.0 (0)	5.0 (1)	4.0 (1)	0.0 (0)	4.3
Method of diagnosis (if microscopic)	4.6 (7)	3.5 (6)	3.9 (3)	4.3 (7)	0.0 (0)	4.5 (2)	4.4 (9)	4.0 (1)	0.0 (0)	4.2
Method of diagnosis (if radiological)	4.5 (9)	3.9 (5)	4.0 (1)	4.7 (7)	4.0 (1)	4.5 (2)	0.0 (0)	3.4 (2)	0.0 (0)	4.1
Method of diagnosis (if clinical)	4.5 (9)	3.9 (7)	4.6 (3)	4.7 (6)	4.0 (1)	4.5 (2)	0.0 (0)	4.0 (1)	0.0 (0)	4.3
Type of surgery	4.4 (9)	4.3 (6)	4.1 (8)	4.3 (8)	4.0 (1)	4.5 (2)	4.3 (9)	5.0 (1)	0.0 (0)	4.4
Source mean	4.2	3.8	4.1	4.4	3.8	4.2	4.6	4.4	4.0	
<i>Delays in cancer treatment</i>										
Date of first visit with general practitioner	4.1 (4)	4.6 (8)	0.0 (0)	3.5 (3)	4.2 (2)	4.0 (1)	0.0 (0)	0.0 (0)	0.0 (0)	4.1
Date of first request for clinic/hospital appointment	4.7 (8)	4.5 (8)	5.0 (2)	4.9 (6)	4.5 (5)	5.0 (1)	0.0 (0)	0.0 (0)	0.0 (0)	4.8
Date of first clinic/hospital appointment	4.5 (8)	4.4 (8)	5.0 (4)	4.6 (5)	4.6 (7)	5.0 (1)	0.0 (0)	0.0 (0)	5.0 (1)	4.7
Date of definitive diagnosis	4.3 (9)	4.0 (7)	5.0 (1)	4.5 (7)	4.7 (3)	4.0 (2)	4.8 (8)	0.0 (0)	0.0 (0)	4.5
Date of surgery	4.7 (9)	4.0 (7)	4.6 (7)	4.6 (8)	4.8 (4)	4.5 (2)	4.6 (8)	5.0 (1)	0.0 (0)	4.6
Date of chemotherapy	4.4 (9)	4.3 (6)	4.2 (5)	4.6 (8)	5.0 (1)	4.6 (3)	5.0 (1)	5.0 (1)	0.0 (0)	4.6
Date of radiotherapy	4.6 (9)	4.4 (6)	4.5 (4)	4.6 (8)	5.0 (1)	4.6 (3)	0.0 (0)	4.8 (4)	0.0 (0)	4.6
Source mean	4.5	4.3	4.7	4.5	4.7	4.5	4.8	4.9	5.0	

*CLIN = Clinical records, GP = Enquiry to GP/family doctor, HAR = Hospital administrative records, MDT = Multidisciplinary team records, WAITS = National Cancer Waiting Times Databas, ONC = Oncology database, PATH = Pathological reports, RAD = Radiotherapy, SCN = Screening files

[†]Any statistic that has only one registry contributing to it, i.e. (1), represents only one experience and should not be universalised

By averaging over the sources or the items of information, the following general comments on reliability of information can be made:

- 1) reliability was estimated to be likely lower in the items of information relating to GUIDELINES than to DELAY.
- 2) In the quest to achieve data on GUIDELINES the source that was reported as likely to be least reliable was *Enquiry to GP/family doctor*. In breast, *National Cancer Waiting Times Database* as a source was also reported as likely having poor reliability, however, the rest of the sources scored, on average, over 4, a high score. There was no similar pattern evident in the DELAY with all sources scoring on average between 4 and 5.
- 3) Averaging reliability of items of information across sources revealed that in GUIDELINES, registries felt that *clinical TNM* was considered the least reliable. For DELAY, *date of first visit with general practitioner* was considered the least reliable.
- 4) The above observations were consistent between the two cancer sites breast and colorectal.

Coding effort

Table 8: Estimated coding effort by registries of items of information from 1 (poor) to 5 (good) in different sources of information for breast cancer (number of registries that contributed to this statistic[†]).

Items of information	Sources*									Item mean
	CLIN	GP	HAR	MDT	WAITS	ONC	PATH	RAD	SCN	
<i>Compliance with treatment guidelines</i>										
Morphology	3.2 (11)	3.7 (9)	2.7 (4)	2.5 (9)	3.4 (2)	3.2 (5)	3.3 (12)	3.0 (1)	2.8 (7)	3.1
ICD-O	3.3 (8)	3.3 (10)	2.8 (8)	3.1 (9)	2.8 (6)	3.0 (2)	3.1 (10)	2.0 (2)	2.3 (6)	2.9
Pathological TNM	3.3 (9)	3.0 (6)	3.0 (1)	2.5 (8)	0.0 (0)	2.5 (2)	3.1 (10)	3.0 (1)	1.0 (1)	2.7
Clinical TNM	3.6 (8)	2.5 (6)	5.0 (1)	2.3 (6)	0.0 (0)	2.0 (1)	2.2 (3)	3.0 (1)	1.0 (1)	2.7
Method of diagnosis (if microscopic)	2.9 (8)	2.5 (7)	1.8 (4)	2.4 (7)	0.0 (0)	3.0 (2)	2.6 (9)	3.0 (1)	1.4 (2)	2.5
Method of diagnosis (if radiological)	2.7 (9)	2.4 (6)	2.7 (3)	2.2 (6)	3.0 (1)	3.0 (2)	2.3 (2)	2.4 (2)	3.0 (1)	2.6
Method of diagnosis (if clinical)	2.8 (9)	2.7 (8)	2.1 (5)	2.3 (6)	3.0 (1)	3.0 (2)	0.0 (0)	3.0 (1)	1.3 (2)	2.5
Type of surgery	3.0 (9)	3.2 (8)	2.1 (7)	2.5 (8)	3.0 (1)	2.0 (2)	3.2 (9)	3.0 (1)	2.0 (2)	2.7
Source mean	3.1	2.9	2.8	2.5	3.0	2.7	2.8	2.8	1.9	
<i>Delays in cancer treatment</i>										
Date of first visit with general practitioner	2.7 (4)	1.7 (7)	0.0 (0)	1.4 (2)	0.0 (0)	1.0 (1)	0.0 (0)	0.0 (0)	0.0 (0)	1.7
Date of first request for clinic/hospital appointment	2.1 (7)	2.2 (8)	2.2 (2)	1.5 (6)	1.0 (4)	1.0 (1)	0.0 (0)	0.0 (0)	0.0 (0)	1.7
Date of first clinic/hospital appointment	2.1 (9)	2.2 (8)	1.7 (3)	2.0 (4)	1.5 (7)	1.0 (1)	0.0 (0)	0.0 (0)	3.0 (1)	1.9
Date of definitive diagnosis	2.2 (7)	2.0 (6)	1.0 (2)	1.6 (6)	2.9 (3)	2.0 (2)	1.0 (8)	0.0 (0)	1.0 (2)	1.7
Date of surgery	2.1 (9)	1.7 (7)	1.6 (7)	1.9 (8)	2.4 (3)	2.0 (2)	2.0 (10)	3.0 (1)	1.0 (2)	2.0
Date of Chemotherapy	2.4 (9)	1.8 (6)	1.6 (5)	1.9 (7)	3.0 (1)	1.2 (3)	1.7 (2)	3.0 (1)	1.0 (1)	2.0
Date of Radiotherapy	2.0 (9)	1.8 (6)	1.6 (4)	1.7 (7)	3.0 (1)	1.0 (3)	0.0 (0)	1.7 (4)	1.0 (1)	1.7
Date of Endocrine treatment	2.2 (9)	2.0 (7)	1.6 (4)	1.8 (8)	3.0 (1)	1.0 (2)	0.0 (0)	3.0 (1)	1.0 (1)	2.0
Source mean	2.2	1.9	1.6	1.7	2.4	1.3	1.6	2.7	1.3	
*CLIN = Clinical records, GP = Enquiry to GP/family doctor, HAR = Hospital administrative records, MDT = Multidisciplinary team records, WAITS = National Cancer Waiting Times Database, ONC = Oncology database, PATH = Pathological reports, RAD = Radiotherapy, SCN = Screening files										
[†] Any statistic that has only one registry contributing to it, i.e. (1), represents only one experience and should not be universalised										

EUROCHIP-2 Final Report – Annex 16 – EUROCHIP-2 in UK

Table 9: Estimated coding effort by registries of items of information from 1 (poor) to 5 (good) in different sources of information for colorectal cancer (number of registries that contributed to this statistic[†]).

Items of information	Sources*									Item mean
	CLIN	GP	HAR	MDT	WAITS	ONC	PATH	RAD	SCN	
<i>Compliance with treatment guidelines</i>										%
Morphology	3.4 (10)	3.6 (7)	2.7 (4)	2.7 (8)	4.0 (1)	3.3 (4)	3.2 (10)	3.0 (1)	4.0 (1)	3.3
ICD-O	3.2 (6)	3.9 (8)	2.5 (7)	2.8 (7)	2.6 (5)	3.0 (2)	3.3 (9)	2.0 (2)	3.0 (1)	2.9
pTNM	2.6 (9)	3.0 (5)	4.2 (2)	2.6 (7)	0.0 (0)	2.5 (2)	2.9 (10)	3.0 (1)	0.0 (0)	3.0
pTNM Duke's stage	2.6 (7)	2.6 (5)	3.0 (1)	2.1 (7)	0.0 (0)	2.0 (2)	2.4 (9)	3.0 (1)	0.0 (0)	2.5
Clinical TNM	3.3 (7)	2.7 (5)	5.0 (1)	2.2 (6)	0.0 (0)	3.0 (1)	2.7 (2)	3.0 (1)	0.0 (0)	3.1
Clinical TNM Duke's Stage	2.6 (5)	2.3 (3)	0.0 (0)	1.9 (3)	0.0 (0)	0.0 (0)	1.0 (1)	3.0 (1)	0.0 (0)	2.2
Method of diagnosis (if microscopic)	2.7 (7)	3.0 (6)	1.8 (3)	2.6 (7)	0.0 (0)	3.0 (2)	2.3 (9)	3.0 (1)	0.0 (0)	2.6
Method of diagnosis (if radiological)	2.8 (9)	2.6 (5)	3.0 (1)	2.3 (7)	3.0 (1)	3.0 (2)	3.0 (1)	2.4 (2)	0.0 (0)	2.8
Method of diagnosis (if clinical)	2.8 (9)	2.4 (7)	1.7 (3)	2.4 (6)	3.0 (1)	3.0 (2)	0.0 (0)	3.0 (1)	0.0 (0)	2.6
Type of surgery	2.9 (9)	3.3 (7)	2.6 (8)	2.5 (8)	3.0 (1)	2.0 (2)	3.0 (9)	3.0 (1)	0.0 (0)	2.8
Source mean	2.9	2.9	2.9	2.4	3.1	2.8	2.6	2.8	3.5	
<i>Delays in cancer treatment</i>										
Date of first visit with general practitioner	2.3 (4)	2.0 (8)	0.0 (0)	1.0 (3)	3.4 (2)	1.0 (1)	0.0 (0)	0.0 (0)	0.0 (0)	1.9
Date of first request for clinic/hospital appointment	1.8 (8)	2.0 (8)	2.2 (2)	1.5 (6)	1.8 (5)	1.0 (1)	0.0 (0)	0.0 (0)	0.0 (0)	1.7
Date of first clinic/hospital appointment	2.0 (8)	2.1 (8)	1.6 (4)	1.8 (5)	1.5 (7)	1.0 (1)	0.0 (0)	0.0 (0)	3.0 (1)	1.9
Date of definitive diagnosis	2.0 (9)	1.8 (7)	2.2 (2)	1.4 (7)	2.6 (3)	2.0 (2)	1.9 (8)	0.0 (0)	0.0 (0)	2.0
Date of surgery	2.0 (9)	1.6 (7)	1.6 (7)	1.9 (8)	3.1 (4)	1.0 (2)	1.4 (8)	3.0 (1)	0.0 (0)	2.0
Date of chemotherapy	2.1 (9)	1.9 (6)	1.6 (5)	1.8 (8)	3.0 (1)	1.2 (3)	3.0 (1)	3.0 (1)	0.0 (0)	2.2
Date of radiotherapy	1.9 (9)	1.8 (6)	1.6 (4)	2.1 (8)	1.0 (1)	1.0 (3)	0.0 (0)	1.9 (4)	0.0 (0)	1.6
Source mean	2.0	1.9	1.8	1.6	2.3	1.2	2.1	2.6	3.0	
*CLIN = Clinical records, GP = Enquiry to GP/family doctor, HAR = Hospital administrative records, MDT = Multidisciplinary team records, WAITS = National Cancer Waiting Times Databas, ONC = Oncology database, PATH = Pathological reports, RAD = Radiotherapy, SCN = Screening files										
[†] Any statistic that has only one registry contributing to it, i.e. (1), represents only one experience and should not be universalised										

By averaging over the sources or the items of information, the following general comments on coding effort of information can be made:

- 1) It was considered that the coding effort would be higher for GUIDELINES than with DELAY.
- 2) In GUIDELINES, the items of information that required most coding effort were: morphology, ICD-0 (or site), followed by clinical, pathological TNM, and type of surgery [breast only]. In DELAY items of information, there was no obvious pattern in coding effort.
- 3) The mean coding effort for source seemed to be stable across GUIDELINES suggesting that effort is largely independent of source but related to the item of information. The source means for DELAY were less stable across the sources, but for no obvious reason, it may be due to the small number of contributions that were made to individual statistics.
- 4) The same general pattern [in 1) and 2) above] was evident for the two cancer sites, colorectal and breast.

3.4 Results: General

Registries felt that sources examined for breast and colorectal cancers would contain a patient record of over 70% of registered patients, with only screening files or oncology databases [breast only] being less than 70%. Only *enquiry to GP/family doctor* source for both cancer sites reached over 90%; so most sources had between 70 and 90% of the registered patients recorded. However, when the completeness of records for particular items of information is factored in, the % actual capture of these items begins to diverge between sources more markedly than % registered patients (see Source mean row in Tables 4 & 5).

The perceived processing requirements of sources followed a consistent and expected pattern for the two cancer sites. Sources that are paper-based or electronic sources that require manual input generally scored as high processing requirement. *Clinical records*, *enquiry to GP/family doctor*, *pathology reports*, or *multidisciplinary team meeting* were ranked top 4 for processing ability, whereas *National Cancer Waiting Times Database* and *screening files* required a low processing (in the lowest three for both cancer sites).

Multidisciplinary team meeting scored as requiring a high level of processing, however a number of registries felt in the future these could be totally electronic and automatic given the general move to electronic 'real time' recording of data at multidisciplinary team meetings. At present it is estimated that less than 50% of *multidisciplinary team meeting* are automatically populating database fields.

Data from the *National Cancer Waiting Times Database* and colorectal cancer *screening* ranked as high for unavailability, despite very low processing requirements suggesting that registries see more difficulties in collection than simply processing these sources to their systems.

The estimated roll-out time to collect new sources was almost identical for breast and colorectal cancers. The estimated roll-out times for both cancer sites for *clinical records* was about 15 months, and for *enquiry to GP/family doctor* 17-21 months, and for *multidisciplinary team meeting* about 12 months. The *National Cancer Waiting Times Database* had the shortest time with 6-8 months. It was felt that the roll-out time of colorectal cancer screening would be twice that of breast.

There was a general correlation between roll-out time and extra-resources required (this is evident if sources with only 1/2 contributing registries are ignored). The perception was that *enquiry to GP/family doctor* or *clinical records* would require 2/3 times more resources than *multidisciplinary team meeting* to roll out.

The rate of registered cancer patients that would have an actual item of information recorded (% actual record) in a source was considered greatest in *clinical records* for almost all of the items of information sought in this study for both GUIDELINES and DELAY. The *multidisciplinary team meeting* was considered similar to the *clinical records* in terms of breadth of items of information captured but was not achieving the same % actual record rate as *clinical records* for individual items of information.

On individual items of information, some electronically captured sources were perceived to match % actual record rate of *clinical notes* for individual items of information. Pathological reports were identified as having a high capture rate for morphology ICD-O, method of diagnosis (if microscopic), date of diagnosis, type of surgery, date of surgery. *Hospital administration records* were identified as matching *clinical records* for data on ICD-O, surgery, date of first clinic/hospital appointment and date of surgery.

National Cancer Waiting Times Database was not identified often as a source for the items of information, suggesting that the people completing the questionnaire were unfamiliar and not confident of this source which is relatively new, however for all DELAY indicators the contributions received for *National Cancer Waiting Times Database* showed a respectable range of % actual record rate between 80-90%. The *National Cancer Waiting Times Database* should be an easy source for registries to upload onto their system and strengthen knowledge on the dates at which patients pass through the system. It will not however, pick up all patients registered in registry, i.e. those not treated in hospitals, or not urgently referred to hospitals. It also contains a large number of non-cancer patients who were referred as suspicious cancers but on further investigation did not have cancer.

In general, reliability was lower in GUIDELINES than DELAY, and coding effort was considered higher in the GUIDELINES, especially items of information such as morphology, ICD-O (or site code), clinical and pathological TNM.

A model for collecting the data on *compliance with treatment guidelines* and *delays in cancer treatment* exists already in Scotland in the form of nationwide prospective audit of cancer services. For every hospital cancer patient a range of data items on treatment and delay are entered on

an ACCESS electronic proforma by audit staff mainly from multidisciplinary team meeting data, but also supplemented with data from chemotherapy diaries, clinical systems, and clinical notes. As the hospital trusts have diagnosis and treatment waiting times targets imposed by government, there has been greater motivation in collecting these data, with less resources available to collect data on the clinical aspects of patient care. However, the Scottish Government Health Department has recently made it clear that it expects equivalent emphasis to be placed on the clinical aspects of patient care. The starting point for measuring waiting times is date of referral from the GP/family doctor, so the audit does not record the date that the patient first visited GP/family doctor in relation to their cancer.

SECTION 4: CONCLUSIONS

Eurochip aims to improve the availability of health indicators in Europe. In the UK *delays in treatment* and *compliance with guidelines* were chosen as important areas to investigate the capture of new items of information.

The results showed that clinical records were considered the best source in terms of the number of items of information contained therein, and the rate at which data could be captured on the entire population of cancer patients. Clinical records were, however, considered labour-intensive to access and process, and therefore resource-intensive. It was reported that *hospital administration* records and/or *pathological reports* could capture certain individual items of information as well as clinical records, but not all items of information. *Multidisciplinary team meeting records*, and *National Cancer Waiting Times Database*, though currently not providing data routinely to all cancer registries, between them could possibly provide data on most items.

The *multidisciplinary team* data source holds promise, not on account of its current value, but on the fact that it is set to become the normative setting for decisions to treat cancer patients, and a move to electronic 'real time' recording of data at multidisciplinary team meetings. Most of the items of information relevant to *Compliance with treatment guidelines* and *Delay in cancer treatment* are supposed to be collected to facilitate joint decision-making by groups of clinicians and also monitoring of the patient's progress. Most registries surveyed believed that they could set up systems to automatically receive data electronically from this source and populate their database fields, thus making it a low-expense system. In addition, in time, it is hoped, that the population coverage would rise to the level of clinical records.

The *National Cancer Waiting Times Database* was identified as having the potential to quickly populate date fields in the registry database concerning the *delay in cancer treatment*; though, it was recognised that it would be biased towards patients who are urgently referred and go through the acute services. It also includes non-cancer patients.

It was reported that more coding effort and quality control will be needed for indicators of *Compliance with treatment guidelines* than *Delays in cancer treatment* because dates are easier to work with than data items such as morphology, ICD-O sites, and staging.

4.1 Conclusions: Recommendation

The data from *multidisciplinary team meetings* and *National Cancer Waiting Times Database* should be further explored as potential data sources for cancer registration to facilitate population-based monitoring of *compliance with guidelines* and *delays in treatment* in the UK because 1) they are relatively low-cost in the long-term to collect 2) they will have good coverage, and 3) cover most of the items of information contained in *compliance with guidelines* and *delays in treatment*.

SECTION 5: APPENDICES

Appendix 1: Detailed Methodology

There was a separate questionnaire for the two sites that were studied: breast (ICD10 codes: C50) and colorectal (ICD10 codes: C18, C19, C20, C21). The questionnaire was supplied as an Access database, and for each site, it was only filled out once per registry.

The questionnaire was in three parts:

1. **Registry details.**
2. **Sources of information.** Assessment of the sources, collected and potentially available, to the registry.
3. **Items of information.** Assessment of the items of information relating to “Compliance with guidelines” and “Delays in Cancer Treatment” which could be accessed from each of the sources described in 2.

Appendices 2 & 3 contains two tables which show the conceptual structure of the information sought in Parts 2 & 3, respectively.

Methodology: Registry details

The main purpose of ‘Registry details’ section was to record the registry’s overall cancer registration budget; this bench mark figure was used to estimate the cost to the registry of collecting new sources of information.

In Part 1, the following information was sought:

1. **Registry name:** This is the usual name given to the registry.
2. **Name and email of person filling out qualitative questionnaire, name and email of [any] assistant:** persons contactable for clarification of any items.
3. **Total cancer incidence (ICD10 code: C00-C97) for most recent complete year of incidence, i.e. number of cases registered.**
4. **Incidence of cancer (breast, colon/rectal): number of cases registered.**
5. **Year of most recent complete year of incidence data: “Complete”** here refers to the fact that all the **data items** usually collected by the registry have been registered for as many cases as possible; not just that all **cases** have been identified.
6. **Population (in millions) of territory covered by the cancer registry.**

7. Registry budget devoted to cancer registration: an estimate of the total amount your registry spends on registration (i.e. exclusive of research and audits).

Methodology: Sources of information

Part 2 of the questionnaire related only to the effort of accessing sources, and did not consider items of information. Core sources (see below) were: hospital administrative records, multidisciplinary team records (MDM), clinical records, pathology reports, enquiries to GP/family doctor, screening files, social insurance records, and National Cancer Waiting Times Database. Other potential sources to a particular registry could be added to this list, when prompted with '*Enter Additional Source*' e.g. a local clinicians' database that is or could be used by the registry.

As this was a population-based feasibility study, only sources of information in their entirety were considered. So for instance, if the registry used a source only to confirm death certificate initiated (DCI) cases, then this was not classified as 'easily available and routinely collected' (see below) because the registry was not collecting information on the non-DCI patients in that source; in other words, the registry was not collecting the entire source.

For each source, the registry recorded:

1. How readily available the source was to the Registry. This was classified as:

- 1.1. Easily available and routinely collected.

This applied to a data source which was (a) easy available within current registry budget and (b) which was routinely collected (in its entirety) and used to provide registration data at that time.

If this was the case, **two** pieces of information were needed:

- i. What % of registered [breast/colorectal] cancer patients (for the last complete incidence year) were recorded in this source? If knowledge on a source was poor, a random sample of registered [breast/colorectal] cancer patients could be matched to the source dataset to estimate this %.
- ii. How the data was received. Four options were available
 - a. **PA** Paper records, accessed: paper records which were actively sought by registry staff
 - b. **PR** Paper records, received: paper records which were sent to registry

- c. **EM** Electronically received, manually processed: electronic records which were entered onto the registration database individually
- d. **EA** Electronically received, automatically processed: electronic records that were merged automatically with the registry database with minimal operator intervention.

The % of records that were captured by each of the four options was entered, which added up to 100%.

1.2. Easily available but not routinely collected.

This applied to data which was (a) easy available without any additional effort but (b) which was not being routinely collected (in its entirety) to provide registration data at present. If this was the case, we needed **three** pieces of information:

i & ii as above

- iii. How soon (in months) the registry could acquire the source for all available cancer information (C00-C97)?

1.3. Available with extra financial resources

This applied to a data source which was potentially available, but would need additional effort to collect in its entirety, and which was not being used to provide registration data at that time. If this was the case, we needed **four** pieces of information:

i, ii and iii as above, but assumed that resources were available to access the data source.

- iv. What % of the current registration budget would be needed to use this source for all cancer cases (C00-C97)?

This was necessarily a rough estimate. For instance, for a registry which got all its data electronically, using medical records might require a doubling of resources, so 200% would have been inserted here.

1.4. Not available at the time

This applies to a data source which was not available for legal (or other non-resource related reasons), but which could give registration information if available. If this was the case, we needed **four** pieces of information:

i, ii, iii and iv as in the previous section.

At the conclusion of the "Sources of information" section, a tick the box "Source information complete" was required.

The Registries could complete this section of the questionnaire largely from their own experience.

The description of the source allowed definition of the type (in terms of availability) of source, without relying on a range of overly subjective

interpretations. There was a generous comment box which allowed for qualifications to any of the descriptive questions, and any other relevant information about the source (its history, or its suspected shortcomings).

The questionnaire came with a core set of sources, to which was added all other sources of information. In addition to the routine sources of the registry, the questionnaire sought to probe all potential sources, which, due to resource constraints, were outside the collecting scope of the registry. The study required that all the core sources were filled out.

The core set of sources were the following, with a abbreviation in brackets that will be used throughout rest of report:

1. Clinical records. These represented the detailed information kept in hospitals, and kept in one place, on patient's history, examination, tests and treatments; also known as case notes or medical charts.
2. Enquiry to GP/family doctor. This was information not contained in the clinical record which could be ascertained from the records of a GP or family doctor. It could be accessed by a phone call, letter or questionnaire to the General Practice, or by directly looking at the GP notes. In some countries, these notes are removed to a central holding area when a patient has died.
3. Hospital Administrative Records. This was information that a hospital kept, usually in electronic form, for the purposes of managing the hospital, for statistical reporting of activity (PAS or HES data), for billing patients for services received, or for other purposes not related to the direct provision of medical care.
4. Multidisciplinary Team Records. This information was recorded at a multidisciplinary team meeting, in which clinicians of various disciplines discussed patients and made joint decisions on the treatment pathway.
5. Pathological reports. These reports came from pathology labs resulting from an examination of a tissue sample.
6. Screening files. This was information that arose from cancer screening services.
7. Social insurance records. This information referred to that which the state held on individuals relating to social welfare benefits, taxes or incomes.
8. National Cancer Waiting Times Database. This information is collected by the NHS to track the progress of suspect cancer patients, identified by their General Practitioner as urgent, through the service from referral to first definitive treatment.

Methodology: Items of information

A list of the items (or indicators defined in EUROCHIP-1) for which information was needed for the pilot project is given below; these were in the Access database. This was the most detailed part of the questionnaire, as we needed to assess the value of all possible sources for a particular item of information.

The definition of most items was identical to those recommended by IARC/ENCR for cancer registration.

Compliance with cancer treatment guidelines (GUIDELINES)

1. Morphology: text or appropriate code (ICDO, SNOMED etc).
2. ICD-O: topography–text, or other sufficient information to allocate a full ICD O code.
3. pTNM: either explicitly recorded or sufficient clinical data for these to be inferred.
4. Pathological Duke's stage: either explicitly recorded or sufficient clinical data for these to be inferred [colorectal only].
5. Clinical TNM either explicitly recorded or sufficient clinical data for these to be inferred.
6. Clinical TNM Duke's Stage: either explicitly recorded or sufficient clinical data for these to be inferred [colorectal only].
7. Method of diagnosis
 - a. Microscopic: if the most valid basis of diagnosis is microscopic (as defined by IARC);
 - b. Radiological: if the most valid basis of diagnosis is imaging (X-ray, scan, CT, MRI etc);
 - c. Clinical: if the most valid basis of diagnosis is physical examination or surgery (without histology)
 - d. Other: the most valid basis of diagnosis by any method; please describe the method.
8. Type of surgery

Delays in cancer treatment(DELAY)

9. Date of first visit with general practitioner: date on which patient first consulted the GP with symptoms suggestive of the cancer.
10. Date of first request for clinic/hospital appointment: date on which GP first requested a hospital appointment or admission for symptoms suggestive of the cancer.
11. Date of first clinic/hospital appointment: date on which the patient attended clinic or hospital; if patient was seen first as an emergency, date on which this took place.
12. Date of definitive diagnosis: as defined by IARC rules for "date of incidence".
13. Date of surgery: date the first surgical procedure is carried out (excluding biopsy without excision).
14. Date of chemotherapy: date of commencement of chemotherapy.
15. Date of radiotherapy: date of commencement of radiotherapy.
16. Date of endocrine treatment: date of commencement of endocrine therapy [breast only] .

For each item of information, the source(s) from which the registry **routinely or potentially** acquired this information were listed provided they were described in 'Sources of information'.

For each item, **four** pieces of information were needed

- a. Was the information usually present in the source (completeness)?
- b. Was the answer to (a) based on a sample taken for this study?
- c. What was the quality of the information on this item from this source (reliability)?
- d. What effort was required to code this item?

Completeness

Having decided that a source carried an item of information, 'completeness' was a % measure of the number of records in that source that actually contained the information (and not a blank or missing value or missing key document). In other words, if one was given 100 records from the source database, or a wad of 100 paper records (e.g. MDTs), what percentage of these actually carried the item of information desired (e.g. date of referral)?

Database managers in cancer registries, drawing on their experience, estimated the completeness (1–100%) of records existing in each source. The database manager could also draw on other feasibility/audit reports that the registry had carried out in the recent past. Database managers could also carry out sampling on the source to estimate completeness (see below).

Sample to estimate completeness

In most situations, the person completing the form was able to estimate completeness from experience or 'guesstimate'; however, if an estimate had been found in a recent audit report, quality control report, or by drawing a sample, then the sampling 'tick box' indicated that the estimate was not a guess.

Reliability

This was an attempt to measure the quality of the data—how likely was it that the information was correct—and was quite subjective. Those who directly manage the source would have considerable experience of the relative quality of data from different sources. The main purpose of this part of the exercise was to identify items which were generally of unacceptable quality from certain sources, even though they were readily

available. Reliability was given as broad bands, representing the percentage of records in which the information was considered to be correct. These were defined as:

- **1:** almost never correct/unusable (information was estimated to be correct in <10% of cases)
- **2:** rarely correct (10%<40%)
- **3:** sometimes correct (40%<60%)
- **4:** usually correct (60%<90%)
- **5:** almost always correct (90%+)

Coding effort

For each item of information, the effort of registering the data not only depended on finding the information, but also on coding or format translation. For instance, if the TNM stage was recorded in multidisciplinary team meetings in the correct format, then the effort to code was small. However, if in the clinical records there was no note of TNM, then it would have to be constructed by the registrar from clinical and pathological data; the effort of coding would be considerable. Coding effort was coded as follows:

- 1:** no additional coding effort—data recorded in source in the format needed
- 2:** slight effort needed; e.g. site given by ICD9 and had to be translated to ICD10
- 3:** some effort needed: morphology was given in words only
- 4:** coding caused some problems: treatment data needs to be coded to ICD9CM
- 5:** coding difficult: TNM not recorded and had to be inferred from clinical and pathological data

Appendix 2: Information collected on core and additional sources of information in Part 2 of the qualitative questionnaire

Sources of information	Source name	Availability	How Captured	If not currently used		Comments
		1: Routinely available and used 2. Routinely available but not used 3. Available with extra resources 4. Not available at present	1. Paper records accessed 2. Paper records received 3. Electronic received, manually processed 4. Electronic received, automatically processed	How many months would be needed start collecting?	What percentage of registration budget would be needed to collect this source? 50%, 400%?	
Core Sources	Hospital administration records e.g. PAS HES					
	Multidisciplinary team meeting					
	Clinical notes					
	Pathology reports					
	Enquiry to GP/family doctor					
	Screening files					
	Social insurance records					
	National Cancer Waiting Times Database					
Additional Sources	Oncology database					
	Etc.....					

EUROCHIP-2 Final Report – Annex 16 – EUROCHIP-2 in UK

Appendix 3: Information collected on items of information in Part 3 of the qualitative questionnaire

(Note the 2-dimensional matrix structure (Items x Sources) with each element of the matrix—item in source—assessed for completeness, reliability, and coding effort.)

Item of information	Core sources of information								Additional sources of information
	Hospital administrative record	Multi-disciplinary team meeting	Clinical records	Pathology reports	Enquiry to GP/Family Doctor	Screening files	Social insurance records	National Cancer Time Waiting	Etc.... →
Morphology	A B D C								
ICD-O									
Pathological TNM									
Clinical TNM									
Date first visit to GP									
Date of first request for hospital appointment									
Date of first hospital appointment									
Date of definitive diagnosis									
Etc....↓									

A Completeness: in %
B Completeness by sample: tick box
C Reliability: 1: almost never correct/unusable (information was estimated to be correct in <10% of cases) 2: rarely correct (10%<40%) 3: sometimes correct (40%<60%) 4: usually correct (60%<90%) 5: almost always correct (90%+)
D Coding effort: 1. no additional coding effort—data recorded in source in the format needed 2. slight effort needed— e.g. site given by ICD9 and has to be translated to ICDO 3. some effort needed— morphology given in words only 4. coding causes some problems—treatment data has to be retrieved from operation notes and coded to ICD9CM 5. coding difficult— TNM not recorded and has to be inferred from clinical and pathological data.