The Completeness of Soft Tissue Sarcoma Data in the National Cancer Data Repository

Tumours Diagnosed Between 2007 and 2009

West Midlands
Cancer Intelligence Unit

CONTENTS

1.0 DATA COMPLETENESS SUMMARY

1.1 Key Findings

In general, there was little change in the completeness of the patient and tumour characteristics data collected across the eight cancer registries in England for soft tissue sarcomas registered in 2006-2008 and in 2007-2009.

The most significant differences between the data for 2006-2008 and 2007-2009 cases are:

- **Improved ethnicity completeness. The completeness for this data item increased from 75% to 79% between the release of the different versions of the NCDR**
- **Laterality completeness decreased for tumours submitted by the NWCIS by 4% to 83%**
- **Only two registries (NYCRIS and WMCIU) are coding GISTs to the allocated morphology code (M8936).**
- **Staging data are incomplete for all registries**

For more information on any of these sections please refer to the relevant section in the main body of the report.

1.2 Executive Summary

Table 3.1: Summary of completeness of soft tissue sarcoma related data items within the NCDR

	% Complete Data item	ECRIC	NWCIS	NYCRIS	OCIU	SWCIS	Thames	Trent	WMCIU
Patient details	Sex	100%	100%	100%	100%	100%	100%	100%	100%
	Date of Birth	100%	96%	100%	100%	100%	100%	100%	100%
	NHS number	100%	99.6%	99.9%	99.8%	99.9%	98.2%	99.9%	99.8%
	Ethnicity	54%	86%	63%	83%	88%	82%	86%	89%
Tumour details	Morphology coding system (ICDM 3)	52%	35%	71%	0%	0%	0%	0%	100%
	Laterality	93%	83%	96%	91%	91%	94%	96%	99%
	Detailed Site Code	86%	80%	85%	79%	84%	71%	88%	85%
Diagnosis Information	Basis of diagnosis (histology)	95%	90%	97%	96%	96%	94%	95%	97%
	Cases registered from more than a death certificate	100%	98.5%	99.8%	99.7%	99.9%	99.2%	98.7%	100%
	Diagnosis dates	99%	97%	100%	100%	100%	98%	100%	96%
Treatment data	Surgery	67%	60%	75%	73%	73%	78%	47%	70%
	Radiotherapy	20%	15%	28%	14%	16%	15%	7%	26%
	Chemotherapy	20%	14%	20%	16%	16%	14%	17%	12%
	Neo-adjuvant therapy	0%	0%	0%	0%	0%	0%	0%	2%
Death data	Cause of death	100%	99%	100%	100%	99%	99%	95%	100%
	Place of death	100%	98%	97%	47%	47%	73%	100%	96%
Staging data	Tumour size	34%	1%	1%	2%	29%	12%	0%	47%
	T component	3%	0%	0%	1%	5%	3%	0%	12%
	Nodes examined	5%	0%	1%	1%	5%	5%	0%	6%
	Nodes positive	1%	1%	0%	0%	1%	1%	0%	2%
	N component	1%	0%	0%	6%	15%	2%	0%	7%
	Metastases ("Yes" or "No")	0%	0%	18%	3%	9%	61%	0%	16%
	M component	2%	1%	0%	4%	10%	1%	0%	8%
	Grade	44%	33%	34%	31%	51%	10%	2%	45%
	TNM stage	1%	1%	3%	3%	9%	0%	0%	8%

Each completeness statistic for each cancer registry was rated as 'red', 'amber' or 'green'. For sections one - patient details, two - tumour details, three - DCO registrations and diagnosis dates and five - death data in Table 3.1, the following cut off points were used:

Key	Description
>95 %	Mostly complete
75% - 94%	Some concerns
<75%	Major concerns

Applicable for sections relating to patient and tumour details, and diagnosis and death data

A rating was not applied for the completeness statistics relating to the basis of diagnosis data in section three and the treatment data in section four, and as it is not clear what the acceptable rates would be.

As staging data were noticeably less complete, a separate rating was used:

Key	Description
>70%	Mostly complete
50%-70%	Some concerns
<50%	Major concerns

Applicable for staging data only

2.0 INTRODUCTION

The West Midlands Cancer Intelligence Unit (WMCIU) is the English lead registry for bone and soft tissue sarcoma. The lead registry analyses national data on the incidence, mortality, survival and treatment of bone and soft tissue sarcomas in England. These analyses are usually conducted using the National Cancer Data Repository (NCDR), a compilation of the eight regional cancer registries which covers all cases diagnosed in England.

In order to understand the robustness of the analyses carried out by the lead registry, it is essential that the limitations of the NCDR are understood, and that the completeness and accuracy of the data items submitted by each registry are evaluated.

This report focuses on the completeness and accuracy of data items collected for soft tissue sarcoma (bone sarcomas are considered in a separate report). The NCDR holds data on all cancers in England, and some fields are site specific. Only fields which relate to soft tissue sarcoma have been analysed in this report. The WMCIU produced a similar report at the beginning of 2011 which focussed on the completeness of data for tumours diagnosed between 2006 and 2008 within the 2010 release of the NCDR (which covers tumour diagnoses between 1990 and 2008). This report investigates the completeness of tumours diagnosed between 2007 and 2009 within the latest version of the NCDR, which covers diagnosis years 1985 to 2009. Comparisons are made with the completeness of data within the previously published report.

3.0 DATA

This data completeness report analyses the completeness of the data items recorded for soft tissue sarcomas registered between 2007 and 2009, as recorded in the most recent edition of the NCDR, which holds all tumours diagnosed between 1985 and 2009.

Soft tissue sarcomas were identified using the International Classification of Diseases – Oncology (ICD-O) coding system for morphology. The list of morphology codes relating to soft tissue sarcomas was agreed by the National Cancer Intelligence Network (NCIN) Sarcoma Site Specific Clinical Reference Group (SSCRG), and includes all International Classification of Diseases version 10 (ICD10) site codes (with the exception of bone; C40 and C41). Only invasive tumours are included in this report. The appendix gives a full list of the invasive morphology codes classified as soft tissue sarcomas.

Extracting only the ICD-O codes for soft tissue sarcomas from the latest version of the NCDR results in 55,887 tumours (55,089 patients) diagnosed between 1985 and 2009. Only sarcomas diagnosed in the most recent three years (2007 to 2009) are included in this data completeness report. This allows the report to focus on current problems of data quality where we can make the largest impact in changing registry practise. There were 8,572 tumours diagnosed in England in this time period, which break down by registry as follows:

Table 2.1 Number of tumours diagnosed within each registry (2007-2009)

Cancer Registry (data source)	Registry (abbreviated)	No. of tumours
Eastern Cancer Registry and Information Centre	ECRIC	875
North West Cancer Intelligence Service	NWCIS	936
Northern and Yorkshire Cancer Registry Information Services	NYCRIS	1,135
Oxford Cancer Intelligence Unit	OCIU	576
South West Cancer Interlligence Service	SWCIS	1,223
Thames Cancer Registry	Thames	2,190
Trent Cancer Registry	Trent	821
West Midlands Cancer Intelligence Unit	WMCIU	816

Each cancer registry submitted all the cases on its local cancer registration database to the NCDR. This produces duplication – for example a patient resident in Bristol but treated in Birmingham

should be registered by the South West as a resident patient, and by the West Midlands as an out-of-region patient treated in region. Only tumours which were flagged as in-region were included within the analysis. Cases flagged as out-of-region were excluded both to focus this initial report on the registries' resident cases (where data quality is most important), and to avoid duplication. Only residents who reside in England are included within the analyses.

4.0 RESULTS: COMPLETENESS OF DATA FIELDS

4.1 Patient Details

4.1.1 Sex

This data item was 100% complete.

4.1.2 Date of birth

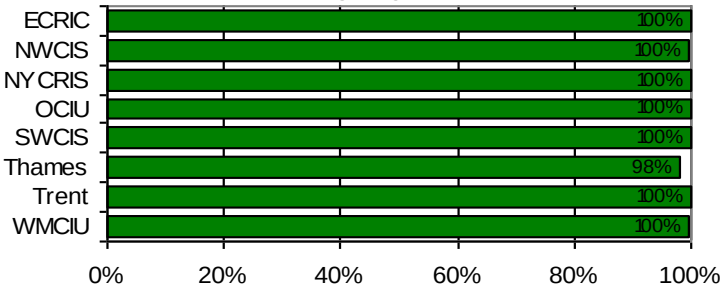
All 8,572 tumours on the NCDR were supplied with a complete date of birth field. A very small proportion of tumours (37 tumours (<1%), 36 from the NWCIS and one from the WMCIU) had a flag set to show that this data item was potentially imputed (e.g. the month and year were known, but the day was imputed).

4.1.3 NHS number

The majority of tumours (8,523, 99%) on the NCDR had an NHS number.

All eight English cancer registries had completeness greater than 99%.

Figure 4.1.3 NHS number completeness by registry



4.1.4 Ethnicity

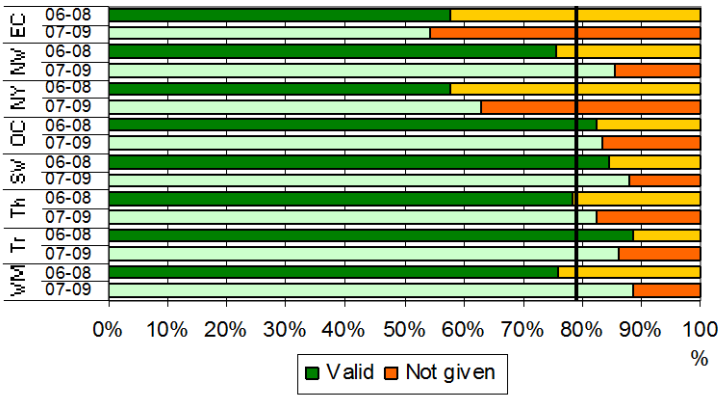
2006 to 2008: 75%

Over three-quarters of 2007-2009 tumours (6,784) had a valid ethnicity code on the NCDR. This is an increase of 4% when compared with the report published previously. Codes were considered valid if they were assigned a specific ethnicity; codes of 'not known' and 'not stated' were excluded from this analysis.

There was wide variation between the English cancer registries, with the WMCIU achieving 89% completeness for ethnicity and the ECRIC achieving only 54%.

Ethnicity data on the NCDR are obtained by linking through to the Hospital Episode Statistics. The data quality issues and regional variation around this linkage will be discussed separately in a follow-on report.

Figure 4.1.4 Ethnicity completeness by registry



4.2 Tumour Details

The tumour details are an important factor when discussing incidence and survival of particular tumour types and the anatomical location of a diagnosis. The completeness of the tumour details are presented in this section.

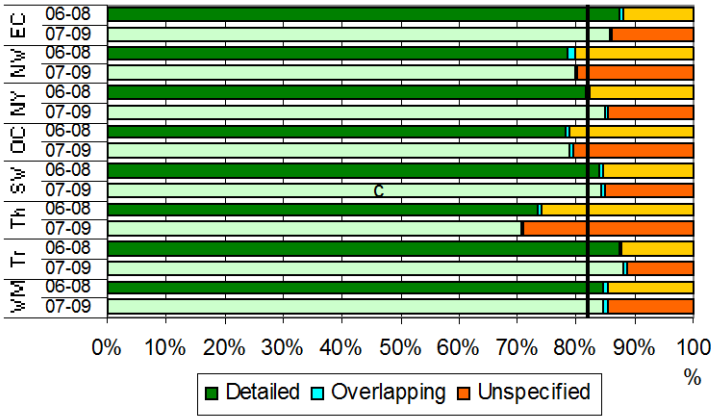
4.2.1 Detailed site code

2006 to 2008: no change

The site of the tumour is coded by registries using ICD-10. The first 3 digits allocate the tumour to a broad site (e.g. "C49" – neoplasm of other connective and soft tissue) and the 4th digit gives a more detailed site code (e.g. "C49.0" – head, face and neck). A 4th digit of '9' means that the detailed site is unspecified.

An accurate site code (where the last digit of the site code does not equal '9') was present for 6,907 (81%) of the tumours. The variation amongst the registries ranged from 71% (Thames) to 88% (Trent).

Figure 4.2.1 Detailed site code completeness by registry



4.2.2 Morphology coding system

All tumours submitted to the NCDR with a morphology were also submitted with a flag to indicate the morphology coding system used.

It was agreed at the UKACR Executive meeting that from diagnosis 2008 onwards, all registries should convert from ICD-O2 to the improved ICD-O3 coding system (library recommendation number Po/08/02). The coding of soft tissue sarcomas has changed noticeably between ICD-O2 and ICD-O3, with over 30 improvements. The most significant difference is the creation of a separate morphology code for GISTs.

Figure 4.2.2 Morphology system usage by registry

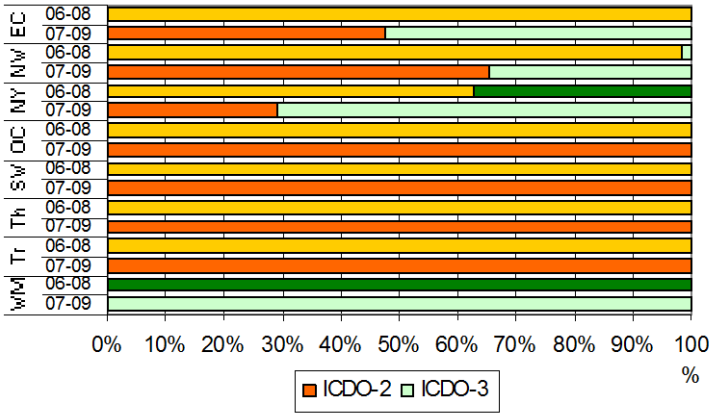


Figure 4.2.2 clearly demonstrates that this change has not yet been fully implemented, as only the WMCIU and the NYCRIS submitted substantial amounts of data using the most up to date coding system. However, the UKACR only recommended the move to ICD-O3 from 2008 onwards, and so registries still submitting in ICD-O2 for 2007 and 2008 are not a cause for concern.

4.2.3 Morphology

There are known variations in registry coding of soft tissue sarcomas across the country, which have implications for the quality of the NCDR. Soft tissue sarcomas are a diverse and complicated tumour type, with over 125 possible morphology codes. As many of these tumours are rare, they are not a priority for registry training sessions. However, accurate coding of morphology is essential to producing reliable statistics.

In the accompanying report, *The Completeness of Bone Sarcoma data in the National Cancer Data Repository,* the completeness of morphology coding for bone cancers is analysed. Only 91% had a valid sarcoma morphology code. Ideally, the current report would contain a similar analysis of morphology coding for soft tissue sarcomas. However, soft tissue sarcomas were identified

using morphology code, not using site code. Therefore any analysis of the completeness of morphology codes is tautologous, with 100% of soft tissue sarcomas having a sarcoma morphology code.

Table 4.2.3 presents the five most common invasive soft tissue sarcoma morphologies coded in each of the cancer registries (highlighted green). The most striking feature of this table is that none of the registries shares the same five most common soft tissue sarcomas, although two morphologies (leiomyosarcoma and sarcoma NOS) are included in the five most common morphologies in all of the registries. These data are difficult to interpret, as differences between the registries may be due to real variation in sarcoma incidence. For example, Kaposi's sarcoma is coded more frequently in the Thames Cancer Registry than elsewhere (12% compared to 5% overall). This disease is linked to AIDS, and London has a greater population of AIDS sufferers than other regions of the UK, so it is likely that these figures reflect true increased incidence. However, other variations are more likely to be driven by coding practices in the various registries.

It can also be seen from Table 4.2.3 that the five most common sarcoma types contribute to only 46% of all tumours registered in WMCIU, compared to 53% within the OCIU and the Thames Cancer Registry. A high percentage of cases registered to common morphology codes may imply that the registry is defaulting to using codes such as 'Not otherwise specified' and missing the opportunities to record more detailed information.

Table 4.2.3: Most common morphology types registered by registry (2007-2009)

Morphology	88903	88003	91403	88503	88323	89303	88013	88113	91203	89903	88513	88303	89363	
Description	Leiomyosarcoma, NOS	Sarcoma, NOS	Kaposi's sarcoma	Liposarcoma, NOS	Dermatofibrosarcoma	Endometrial stromal sarcoma	Spindle cell sarcoma	Fibromyxosarcoma	Haemangiosarcoma	Mesenchymoma, malignant	Liposarcoma, well differentiated	Fibrous histiocytoma, malignant	Gastrointestinal stromal sarcoma	5 Most common types
ECRIC	21%	12%	3%	5%	5%	3%	5%	4%	2%	7%	2%	2%	0%	**50%**
NWCIS	18%	13%	5%	5%	6%	1%	4%	4%	5%	4%	4%	2%	1%	**52%**
NYCRIS	17%	14%	2%	5%	6%	2%	4%	3%	3%	2%	7%	1%	5%	**49%**
OCIU	17%	12%	2%	12%	3%	3%	4%	3%	2%	7%	4%	3%	0%	**53%**
SWCIS	17%	12%	3%	7%	4%	3%	4%	5%	4%	5%	2%	3%	0%	**47%**
Thames	14%	10%	12%	5%	6%	12%	4%	3%	3%	0%	1%	3%	0%	**53%**
Trent	19%	11%	4%	2%	6%	3%	2%	5%	4%	6%	3%	2%	0%	**48%**
WMCIU	17%	6%	3%	3%	4%	1%	8%	2%	6%	0%	5%	8%	8%	**46%**
Grand Total	**17%**	**11%**	**5%**	**5%**	**5%**	**5%**	**4%**	**4%**	**4%**	**3%**	**3%**	**3%**	**1%**	**44%**

4.2.3.1 Sarcoma Not Otherwise Specified

Use of the code M88003, Sarcoma, Not Otherwise Specified (NOS), may be indicative of data quality issues, suggesting that sarcomas are being assigned to a generic code instead of a more specific one. There is a wide variation between registries in the use of this code. Only 6% of cases registered by the WMCIU were coded to this code, compared to 11% overall.

4.2.3.2 Gastrointestinal Stromal Tumours

The most immediate data quality issue identified is the coding of gastrointestinal stromal tumours (GISTs). In ICD-O3 a specific morphology code was allocated to GISTs (M89363). However, the majority of cancer registries did not use ICD-O3 for the time period covered by the NCDR (see Section 4.2.2). Table 4.2.4 summarises the self-reported statements of the registries on how GISTs are coded.

Table 4.2.4: Results from national GIST coding survey

Cancer Registry	Pre ICDO-3 code	ICDO-3	ICDO-3 from:
ECRIC	89903 (Mesenchymoma)	89363 (GIST)	2008
NWCIS	89903 (Mesenchymoma)	89363 (GIST)	2008
NYCRIS	89903 (Mesenchymoma)	89363 (GIST)	2008
OCIU	89903 (Mesenchymoma)	No	---
SWCIS	89903 (Mesenchymoma)	No	---
Thames	89303 (Endometrial stromal sarcoma)	89363 (GIST)	2000
Trent	89903 (Mesenchymoma)	No	---
WMCIU	89903 (Mesenchymoma)	89363 (GIST)	2005

Five of the 8 registries claim to be using the GIST code in ICD-O3. However, only the WMCIU and the NYCRIS submitted a significant number of tumours to the NCDR using ICD-O3, and only the WMCIU (8%) and NYCRIS (5%) had a sufficiently high number of cases coded to the GIST code for this to be included in their most common five morphologies. Although they claim to be using ICD-O3 other registries do not appear to have used the GIST code when registering cases diagnosed from 2008 onwards.

Historically registries have used M89903 (mesenchymoma) to code GISTs. Registries that have not moved to using ICD-O3, or that moved towards the end of the time period covered by this report (NWCIS (4%), Trent (6%), ECRIC (7%), OCIU (7%) and SWCIS (5%)) have an elevated percentage of mesenchymomas.

The Thames Cancer Registry is a clear outlier as it conforms to other international standards in converting registered GIST codes to Endometrial Stromal Sarcomas. Although it self-reports as using ICD-O3 since 2000, all cases were submitted to the NCDR as ICD-O2, and the registry does not follow the registry guidelines for GISTs, being alone in using M89303 (endometrial stromal sarcoma) to code these tumours. This can clearly be seen in Table 4.2.3, where Thames has 12% endometrial stromal sarcoma's in its most common five morphologies.

The NWCIS does not appear to have any code which it is obviously using for GISTs. Even combining cases coded to mesenchymoma (4%) and GIST (1%) does not alter this finding.

4.2.3.3 Dermatofibrosarcoma

Dermatofibrosarcomas on average make up 5% of all soft tissue sarcomas recorded by the registries. However, only 3% of soft tissue sarcomas registered by the OCIU were allocated this code. It is not clear if this indicates poor case ascertainment, variation in coding practise, or true variation in incidence.

4.2.3.4 Liposarcoma

There are 7 separate morphology codes in the ICD-O3 for recording liposarcomas. M88503 (liposarcoma not otherwise specified) is used by all registries, but the OCIU is a clear outlier; coding 12% of all sarcomas to this code. M88513 (liposarcoma, well differentiated, including sclerosing liposarcoma and inflammatory liposarcoma) is much more commonly used by the NYCRIS than by any other cancer registry. Care must be taken when producing statistics by morphology codes not to create regional variation which is only an artefact of coding choices.

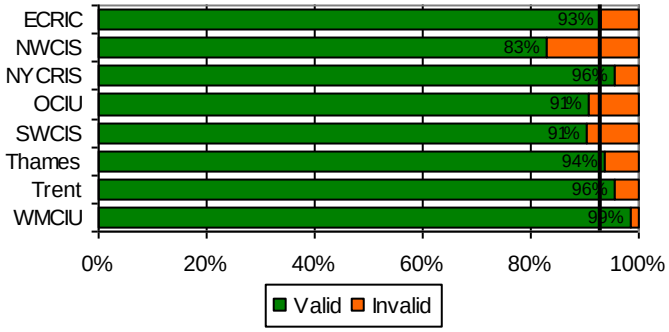
4.2.4 Laterality

2006 to 2008: no change

Overall, 93% of the tumours included within the analysis contained a valid laterality value. Completeness varied between the registries, with the NWCIS submitting values for only 83% compared to 99% from the WMCIU.

Only tumours diagnosed to the limbs (including peripheral nerves and skin), breast, lung and kidney (2,667 tumours) were included in this analysis, of which 2,475 contained a valid laterality. Tumours of sites which only occur once in the body were excluded.

Figure 4.2.4 Laterality completeness by registry

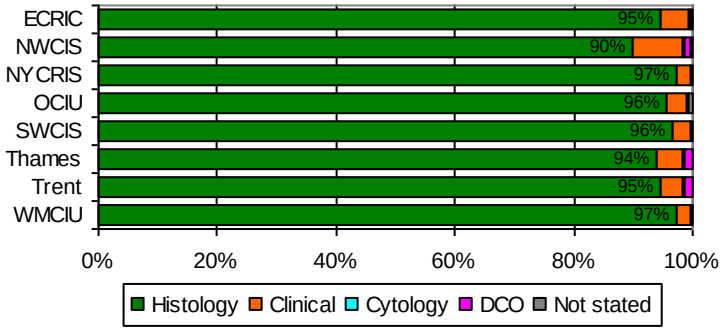


4.3 Diagnosis Data

4.3.1 Basis of diagnosis

A basis of diagnosis was submitted for 8,558 out of the 8,572 tumours. Only 14 tumours had an unknown basis of diagnosis. A histological diagnosis will always provide more reliable information on morphology and behaviour than a clinical diagnosis.

Figure 4.3.1 Basis of diagnosis completeness by registry

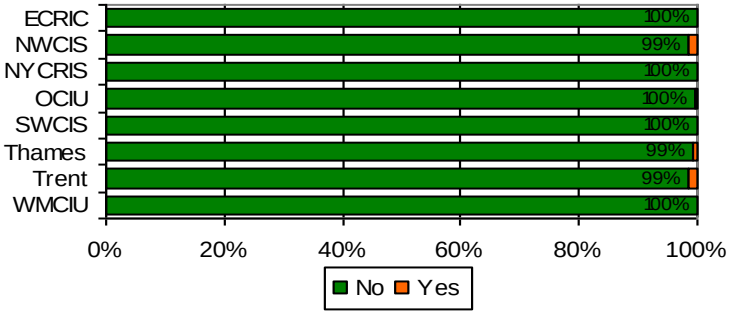


A far higher percentage of cases had a histological diagnosis for soft tissue sarcomas than for bone sarcomas. However, this is not necessarily a sign of better data quality – the soft tissue sarcomas were identified by their morphology, so only cases with reasonable data on morphology (i.e. primarily histologically diagnosed cases) were included in the cohort. It is not known how many soft tissue sarcomas are diagnosed clinically or from death certificates, and coded to codes such as '80003 – neoplasm, malignant' and hence excluded from the cohort.

4.3.2 Tumours possessing more detail than a death certificate

Tumours can be registered from many different sources, although a full pathology report remains the "gold standard". However, there are cases when the only information received by the registry is a death certificate.

Figure 4.3.2 More than death certificate completeness by registry



Death Certificate Only cases are problematic, as they suffer from coding problems, and may indicate that the registry is missing live cases as well as dead cases. Of the 8,572 tumours diagnosed during the period of interest, there are only 48 where the death certificate is the only source of information. While 19

of these (40%) are recorded as “Sarcoma, NOS”, the remaining 29 are coded to detailed specific morphologies. It is surprising that this detailed tumour morphology was available on a death certificate when no further information on the patient’s tumour was available. However, this only relates to a very small proportion of tumours diagnosed within the NWCIS and the Trent Cancer Registry.

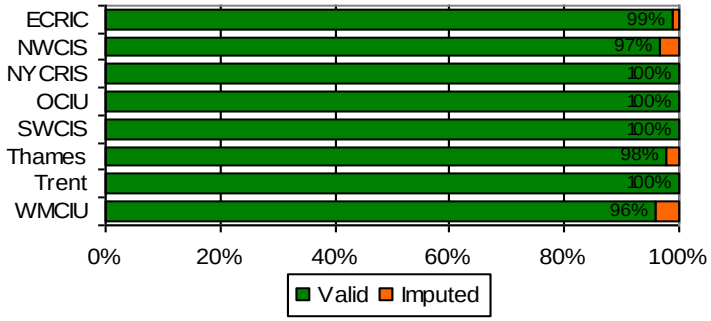
4.3.3 Diagnosis dates

2006 to 2008: 98%

8,449 (99%) of the tumours were supplied with complete diagnosis dates. Of the 123 tumours submitted with imputed elements, 80 had a missing day, 12 had an incomplete day and month, and the remaining 31 were uncertain as to whether imputation had taken place.

The diagnosis date is a vital piece of information required to calculate accurately statistics such as a patient’s age at diagnosis, the number of cases diagnosed in a year, and the patient’s survival time.

Figure 4.3.3 Diagnosis date completeness by registry

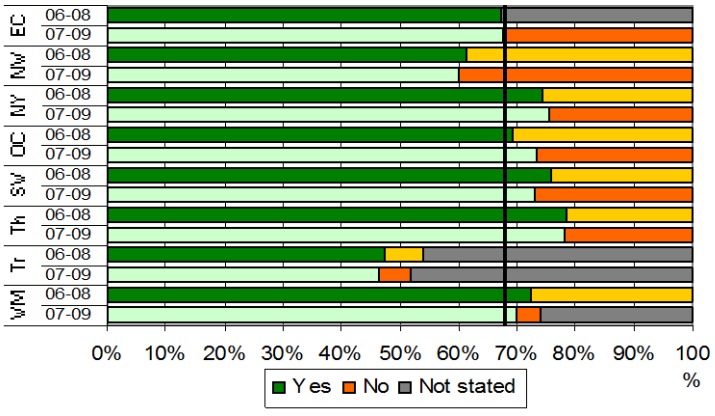


4.4 Treatment Details

4.4.1 Surgery

Overall, 70% of patients were recorded as having surgery. This varied widely between registries, from 47% (Trent) to 78% (Thames). There are many factors which may be driving this variation, but it is more likely that this is due to inconsistent definitions of surgical treatment and problems receiving data on patients treated out of region, than genuine variation in patient care.

Figure 4.4.1 Surgery completeness by registry



There were two different approaches to submitting data on patients who did not have surgery. For 2006-2008 cases in the older version of the NCDR, the ECRIC and the Trent Cancer Registry appear to have taken the analytical approach that, if the registry has not received evidence that the patient was treated surgically and it has not received evidence that the patient wasn’t surgically treated, it has left the surgery field blank. This is also the method adopted by the WMCIU and the Trent Cancer Registry for 2007-2009 cases in the most recent NCDR. The ECRIC on the other hand, appears to have submitted all of its surgical treatment details as either “Yes” or “No” in the latest version of the NCDR. Other cancer registries have consistently submitted ‘no surgery’ for any case where there was no evidence of surgery.

The surgery flag is derived from the corresponding Office of Population, Censuses and Surveys Classification of Surgical Operations and Procedures (OPCS4) codes related to the tumour, and most operations have been performed within the six months following diagnosis. Therefore, the accuracy of surgery “Y” and “N” flags could be further flawed due to any inconsistencies between the cancer registries in the OPCS4 codes classified as “surgical treatment”. This problem could

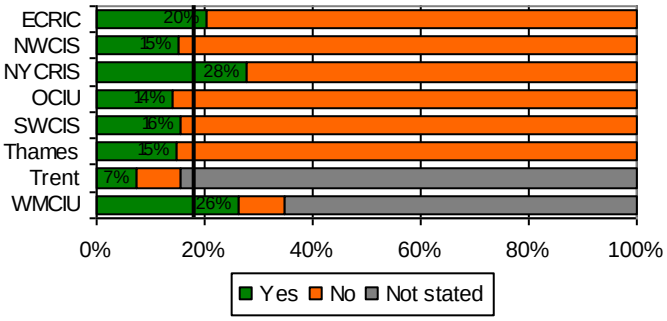
easily be addressed by agreeing at a national level the OPCS4 codes to be classified as surgical treatment.

4.4.2 Radiotherapy

2006 to 2008: no change

Radiotherapy treatment was recorded for 1,502 (17%) tumours. The proportion of tumours receiving radiotherapy is consistently relatively small across all registries.

Figure 4.4.2 Radiotherapy completeness by registry

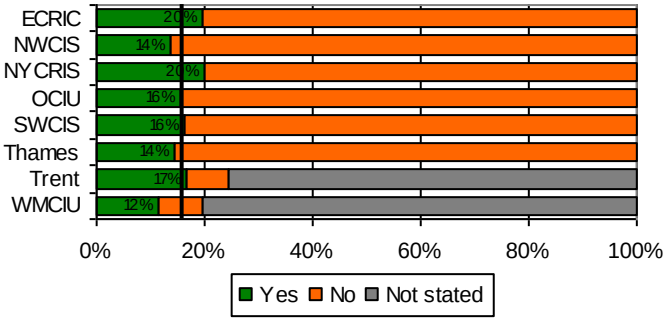


The radiotherapy data submitted by each registry should relate to radiotherapy sessions delivered within six months of the diagnosis. As discussed in Section 4.4.1 for surgery, the value 'no radiotherapy treatment' was approached differently by the registries. It appears that the Trent Cancer Registry and the WMCIU have only supplied positive or negative responses where actually stated for the corresponding tumour. If the radiotherapy status was unknown for the tumour then the field was left blank. Although it is not clear what percentage of cases are expected to receive radiotherapy, and there will be variation in casemix across the regions, the 7% submitted by the Trent Cancer Registry appears very low.

4.4.3 Chemotherapy

Across all registries chemotherapy was recorded for 1,362 (15%) tumours. The proportion of tumours receiving chemotherapy was consistently relatively small across all registries, with the highest levels (20%) being recorded by NYCRIS and ECRIC.

Figure 4.4.3 Chemotherapy completeness by registry

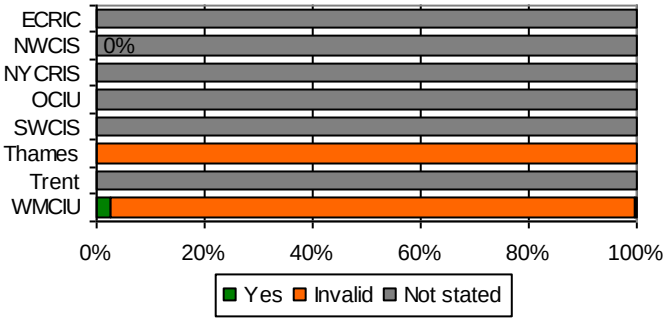


The chemotherapy data submitted by each registry should relate to chemotherapy sessions administered within six months of diagnosis. As discussed in Section 4.4.1 for surgery, the value 'no chemotherapy treatment' was approached differently between the registries.

4.4.4 Neo-adjuvant therapy

Cancer registries as a whole do not collect reliable data on whether the tumour was treated with neo-adjuvant therapy. Only 2 of the 8 registries submitted data in this field to the NCDR, and one of these (Thames) claimed that none of their soft tissue sarcoma patients had received neo-adjuvant therapy.

Figure 4.4.4 Neo-adjuvant therapy completeness by registry



The only cancer registry to identify that patients were receiving neo-adjuvant therapy was the WMCIU, where 20 (2%) tumours were positively identified as receiving neo-adjuvant therapy. This was determined by comparing the dates of surgery and chemotherapy, with the latter being before the former if the patient had neo-adjuvant therapy.

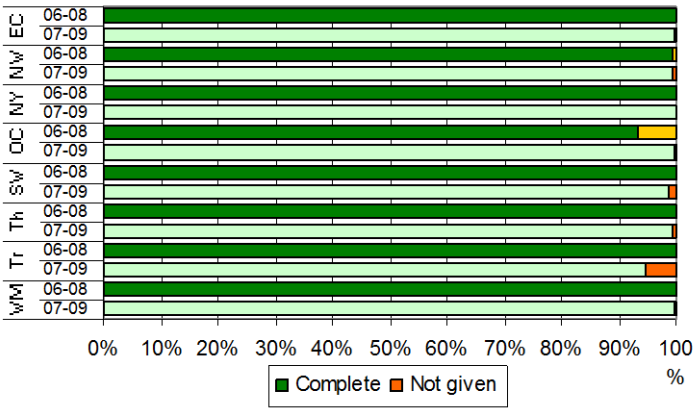
4.5 Death Details

4.5.1 Cause of death

Considering only the tumours where the patient was known to have died (3,166), 3,067 (97%) had a valid cause of death code.

The cause of death information supplied on the death certificates is registered as an ICD10 code. This information has always been provided on death certificates and the NCDR contains four "cause of death" fields. However, not all patients will have four causes of death completed. Therefore, for the purposes of these analyses, only the first cause of death field ("cod_1a") was analysed.

Figure 4.5.1 Cause of death completeness by registry



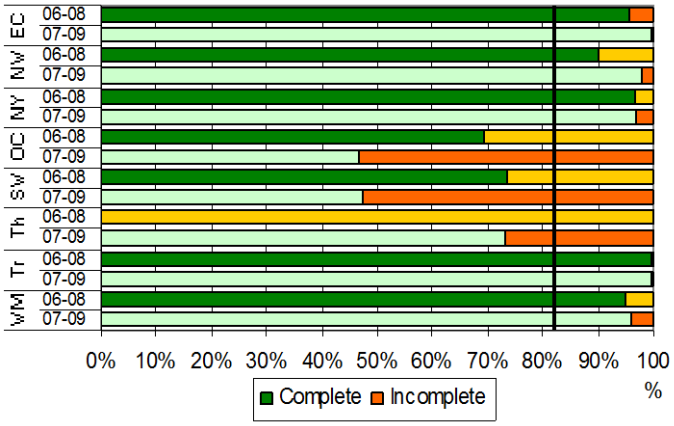
Cause of death was 100% complete for 4 out of the 8 registries and all registries had a cause of death for at least 95% of patients who were known to have died. The Trent Cancer Registry did not submit a cause of death for 44 (5%) of its patients.

4.5.2 Place of death

The completeness of the "place of death" field was calculated for all tumours where the patient was known to have died. This information was present for 2,591 (82%) of the tumours.

There is wide variation in data completeness between cancer registries. The OCIU and the SWCIS have a known place of death for fewer than 47% of their cases. This is similar to the variation found in the bone sarcoma data completeness report.

Figure 4.5.2 Place of death completeness by registry



4.6 Staging Data

Cancer registration staging data have been historically incomplete. While the UKACR Annual Performance Indicators exercise has improved staging for common cancer sites such as colorectal cancer, the recording of staging data for rarer cancer sites such as soft tissue sarcomas has not been a priority.

A new UKACR Performance Indicator, introduced for the 2010 data, will monitor the percentage completeness of all staging data for all cancer sites. It is hoped that data completeness of soft tissue sarcoma stage will improve because of this. However, the data analysed in this report were collected before this new performance indicator was introduced.

4.6.1 Staging systems

The "*Guidelines for the Management of Soft Tissue Sarcomas*" (Grimer et al, 2010, *Sarcoma*) states the most widely accepted staging system is the American Joint Committee on Cancer

(AJCC) TNM staging system. However, the NCDR specified that stages should be submitted using the Union for International Cancer Control (UICC) Tumour, Nodes and Metastases (TNM) staging system. These staging systems are almost identical. In UICC TNM v7 and AJCC TNM v7, T stage, N stage and M stage are defined identically. However, the AJCC TNM staging system uses the French three-grade Federation Nationale des Centres de Lutte Contre le Cancer (FNCLCC) grading system, whereas the UICC TNM staging system uses a two-grade system. The UICC TNM classification book provides a mapping from the three-grade system to the two-grade system. The stage grouping algorithms are nearly identical. Only a T2b, N0, M0, G2 (FNCLCC) tumour will be graded differently between the two algorithms, as a Stage IIB tumour under AJCC TNM and as a Stage III tumour under UICC TNM. Although this is only a single edge case, it does mean that there are potential data quality issues if the stage of the tumour is known, but the staging system is not known. If cancer registries have received a pathology report with a TNM stage on it for a rare soft tissue sarcoma, they are likely to have recorded it assuming it was a UICC stage.

The NCDR contains fields for the collection of the grade of the tumour, T, N and M components and the overall TNM stage (either pathological, clinical and integrated). Both TNM systems utilise tumour characteristics relating to tumour size, nodal spread, grade of the tumour, and distant metastases which may be collected clinically, pathologically, or be recorded as an integrated stage. Initial analyses of these fields indicated that they were incomplete. Therefore, for each component, if information was present in *any* of the fields, the corresponding tumour was presented as having staging information submitted.

There are other staging systems used for soft tissue sarcomas, such as the Surgical Staging System (Enneking). These are not collected in the NCDR, and have not been mandated in the new Cancer Outcomes and Services dataset. It is presumed that no cancer registries will attempt to collect staging data for soft tissue sarcomas using a staging system which is not TNM.

4.6.2 Sites staged

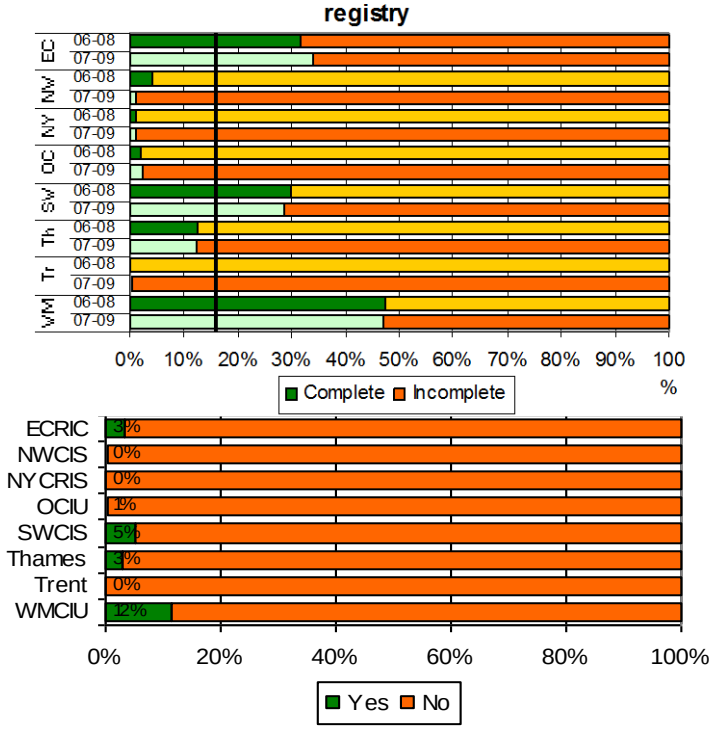
Both the AJCC TNM staging system and the UICC staging system are only appropriate for particular sites and morphologies. Neither system will stage Kaposi's sarcoma. The UICC TNM staging system for soft tissue sarcomas (which the NCDR claims to contain) does not stage dermatofibrosarcomas, angiosarcomas, sarcomas arising from the dura mater, brain, hollow viscera or parenchymatous organs. A UICC TNM staging system for GISTs does exist, but is separate to the UICC TNM staging system for other soft tissue sarcomas. Therefore until site and morphology specific staging systems are agreed and collected for all soft tissue sarcomas, 100% completeness of staging data will remain an impossibility.

4.6.3 Tumour size

Of the 8,572 tumours diagnosed between 2007 and 2009, only 1,339 (16%) had a tumour size recorded in mm. The range of tumour sizes varied between 1 and 500 mm.

There was large variation between registries, with the WMCIU collecting size for almost half its tumours, and the Trent Cancer Registry submitting no size data at all.

Figure 4.6.1 Tumour size completeness by registry



4.6.4 T component

Of the 8,572 tumours diagnosed between 2007 and 2009, just 256 (3%)

possessed a clinical, pathological or integrated T stage. (A value of "TX", primary tumour cannot be assessed, was not included as a valid T stage in this analysis).

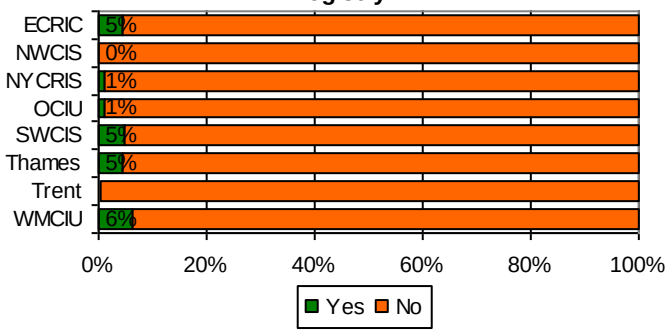
It is evident from Figure 4.6.1 that the tumour size is considerably more complete than the "T" component of stage in Figure 4.6.2. For example, 47% of the tumours registered within the WMCIU have a tumour size, yet only 12% were submitted as having a T component. This was also true for the ECRIC, which submitted a size for 34% of tumours, yet submitted a "T" value for fewer than 3%.

The top-level T-stage (T1 or T2) can be derived directly from the size of the sarcoma, and so could be submitted for any sarcoma with a size. The detailed T-stage (T1a, T1b, T2a or T2b) cannot be derived without knowing the depth of the sarcoma. However, this detailed T-stage is not required for calculating the overall stage, and so failure to submit a T-stage for all sarcomas with a size shows that the full power of the data collected by the registries is not being exploited.

4.6.5 Number of Nodes examined

Cancer registry data on whether or not nodes were examined were incomplete. The number of nodes examined was present for just 275 (3%) of all tumours. Information was considered complete if the field relating to nodes was not blank.

Figure 4.6.3 Nodes examined completeness by registry

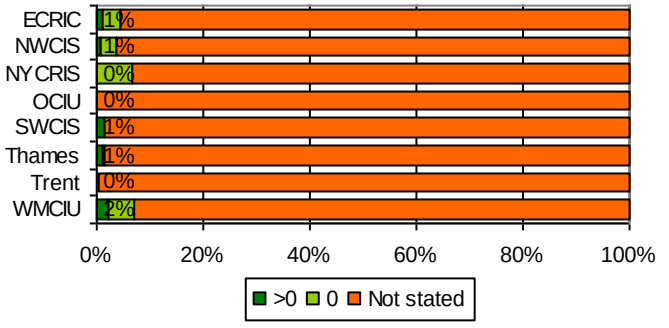


4.6.6 Number of Nodes positive

Cancer registry data on the number of positive nodes were also incomplete. There were 83 (1%) tumours with positive nodes and 173 with no positive nodes. In total, information was available for 256 tumours (3%)

There are some inconsistencies between the data in Figure 4.6.3 and Figure 4.6.4. For example, the NWCIS reported that nodes had not been examined for 100% of their tumours in Figure 4.6.3. But 4 tumours from NWCIS had the number of positive nodes recorded, and 16 had recorded that no nodes were positive.

Figure 4.6.4 Nodes positive completeness by registry

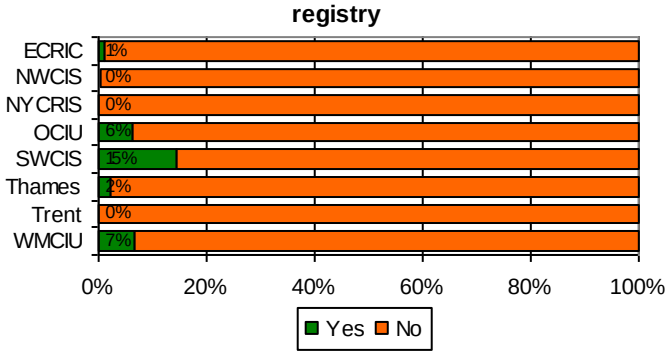


4.6.7 N component

The N component of stage was also incomplete, with only 329 tumours (4%) having either a clinical, pathological or integrated N stage (the presence or absence of metastasis in the regional lymph nodes).

Again, there are discrepancies between Figure 4.6.3 and Figure 4.6.5. For example, the SWCIS reported an N component for

Figure 4.6.5 N component completeness by registry

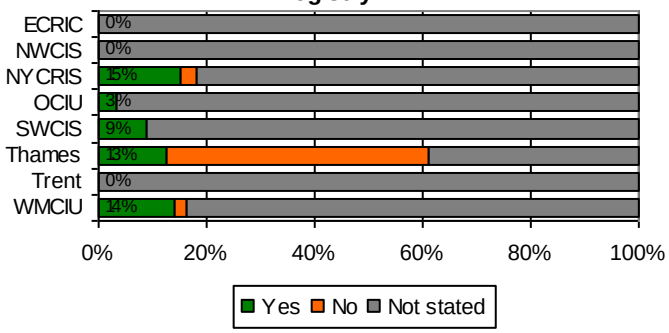


15% of soft tissue sarcomas recorded. However only 5% of its soft tissue sarcomas were reported as having had nodes examined. The Trent Cancer Registry submitted no nodal information on any soft tissue sarcoma in the NCDR.

4.6.8 Metastases

Overall only 1,803 of the 8,572 tumours (21%) had a flag which clearly stated whether metastases were present or not. There was large variation between registries, with the majority of the cases with reported metastases coming from the Thames Cancer Registry.

Figure 4.6.6 Metastases completeness by registry

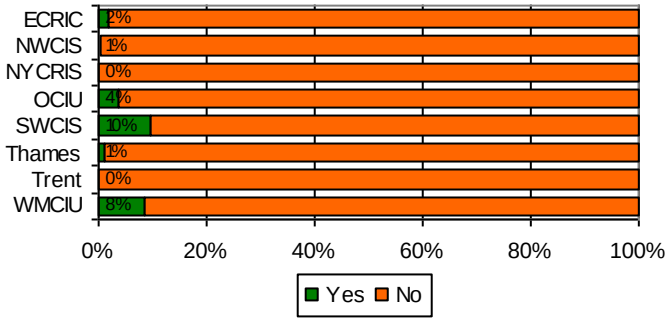


4.6.9 M component

Of the 8,572 soft tissue sarcomas diagnosed, only 252 tumours (2.9%) had either a clinical, pathological or integrated "M" component.

The "M" value relates to the presence or absence of distant metastases. Comparing Figure 4.6.6 to Figure 4.6.7 it is evident that the information supplied relating to metastases is inconsistent across registries. For example in the Thames Cancer Registry in Section 4.6.8 information on metastases was available for 61% of cases, and yet an M component of TNM was submitted for just 1% of cases.

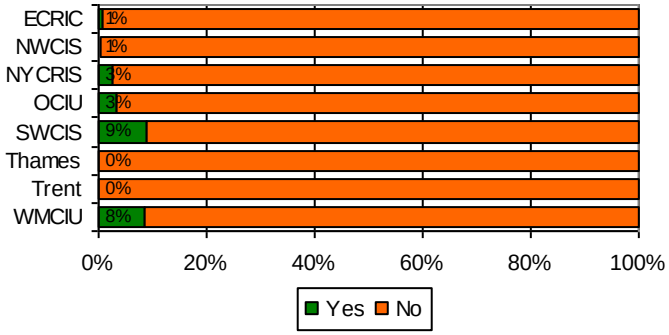
Figure 4.6.7 M component completeness by registry



4.6.10 TNM stage

Figure 4.6.8 clearly demonstrates that the overall TNM staging for soft tissue sarcoma is incomplete across all registries, with just 236 tumours (3%) being submitted with an overall TNM stage (either clinical, pathological or integrated).

Figure 4.6.8 TNM value completeness by registry

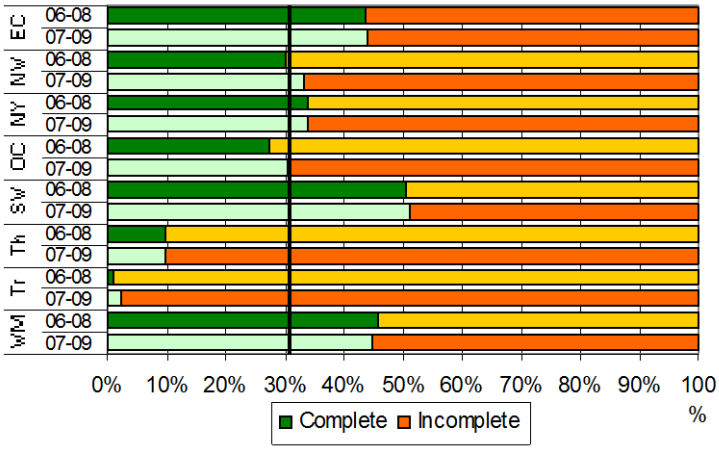


4.6.11 Grade

There were only very limited data available on tumour grade in the NCDR, with only 2,484 tumours (29%) having a grade submitted.

Completeness ranged widely between registries Figure 4.6.9), with even the best performing registry (SWCIS) only submitting a grade for 51% of tumours, and the worst performing registry (Trent) submitting grade information for just 2% of tumours.

Figure 4.6.9 Grade completeness by registry



5.0 DISCREPANCIES BETWEEN 2010 AND 2011 NCDR RELEASES

This section has been added to the report to investigate the differences in the soft tissue sarcomas diagnosed between 2006 and 2008 within the 2010 NCDR data release, and for the same period within the 2011 NCDR data release.

The 2006-2008 completeness report for soft tissue sarcomas stated that 8,279 tumours were diagnosed based on the 2010 release of the NCDR, compared to 8,468 recorded in the same period and submitted for the 2011 release of the NCDR.

Table 5.1: Changes in the tumour details submitted as soft tissue sarcoma's

	Cancer Registry							
	ECRIC	NWCIS	NYCRIS	OCIU	SWCIS	Thames	Trent	WMCIU
--in 2008 NCDR dataset	872	821	1,082	526	1,226	2,172	802	778
tumours in 2010 NCRD and not 2011	1	10	1082	3	16	14	4	16
plus the tumours in 2011 which were not submitted in 2010	18	122	1,112	20	18	30	7	8
Grand total	**889**	**933**	**1,112**	**543**	**1,228**	**2,188**	**805**	**770**

Between the releases of the two iterations of the NCDR, NYCRIS altered its tumour numbering system, making it difficult to compare tumour details recorded in the two datasets. All registries reported additional diagnoses of soft tissue sarcoma tumours in the latest release of the NCDR. These could relate to late registrations with regard to tumours diagnosed in 2009, but they could also relate to changes in diagnosis.

APPENDIX

Morphology codes classified as soft tissue sarcoma

Some morphology codes within the ICD-O3 classification are normally reserved for either benign or uncertain diagnoses. However, there are instances where these have been used for malignant diagnoses. These tumours are highlighted grey in the table below.

Morphology	Description
8710	Glomangiosarcoma: Glomoid sarcoma
8711	Glomus tumour (nad varients), malignant glomus tumour
8713	Myopericytoma
8800	Sarcoma, NOS
8801	Spindle cell sarcoma
8802	Giant cell sarcoma (except of bone M9250/3); pleomorphic cell sarcoma
8803	Small cell sarcoma; round cell sarcoma
8804	Epithelioid sarcoma, epithelioid cell sarcoma
8805	Undifferentiated sarcoma
8806	Desmoplastic small round cell tumour
8810	Fibrosarcoma, NOS, sclerosing epitheliod fibrosarcoma
8811	Fibromyxosarcoma
8812	Periosteal fibrosarcoma (C40._, C41._); periosteal sarcoma, NOS (C40._, C41._)
8813	Fascial fibrosarcoma
8814	Infantile fibrosarcoma; congenital fibrosarcoma
8815	Solitary fibrous tumour, NOS
8821	Aggressive fibromatosis, Desmoid tumour NOS
8822	Abdominal fibromatosis (ICDO-2)
8823	Desmoplastic fibroma (ICD-O-2)
8824	Myofibromatosis (ICD-O3)
8825	Inflammatory myofibroblastic tumour, Myofibroblastic tumour, NOS
8830	Fibrous histiocytoma, malignant; fibroxanthoma, malignant
8832	Dermatofibrosarcoma, NOS (C44._); dermatofibrosarcoma protuberans, NOS (C44._)
8833	Pigmented dermatofibrosarcoma protuberans; Bednar tumour
8834	Giant cell fibroblastoma
8835	Plexiform fibrohistiocytic tumour
8836	Angiomatoid fibrous histiocytoma
8840	Myxosarcoma
8841	Angiomyxoma
8842	Ossifying fibromyxoid tumour, atypical
8850	Liposarcoma, NOS; fibroliposarcoma
8851	Liposarcoma, well differentiated; Liposarcoma, differentiated
8852	Myxoid Liposarcoma; myxoliposarcoma
8853	Round cell liposarcoma
8854	Pleomorphic liposarcoma
8855	Mixed liposarcoma
8857	Fibroblastic liposarcoma
8858	Dedifferentiated liposarcoma
8860	Angiomyoliposarcoma
8890	Leiomyosarcoma, NOS
8891	Epithelioid leiomyosarcoma
8894	Angiomyosarcoma
8895	Myosarcoma
8896	Myxoid leiomyosarcoma
8897	Smooth muscle tumour
8898	Metastising leiomyosarcoma

Morphology	Description
8900	Rhabdomyosarcoma, NOS; rhabdosarcoma
8901	Pleomorphic rhabdomyosarcoma
8902	Mixed type rhabdomyosarcoma
8910	Embryonal rhabdomyosarcoma; sarcoma botryoides; botryoid sarcoma
8912	Spindle cell rhabdomyosarcoma
8920	Alveolar rhabdomyosarcoma
8921	Rhabdomyosarcoma with ganglionic differentiation; Ectomesenchymoma
8930	Endometrial stromal sarcoma (C54.1)
8931	Endometrial stromal sarcoma, low grade
8935	Stromal Sarcoma
8936	Gastrointestinal stromal sarcoma
8940	Ossifying fibromyxoid mixed tumour
8951	Mesodermal mixed tumour
8963	Rhabdoid sarcoma
8964	Clear cell sarcoma of kidney
8982	Myoepithelioma
8990	Mesenchymoma, malignant; mixed mesenchymal sarcoma
8991	Embryonal sarcoma
9020	Phyllodes tumour, malignant (C50.) Cystosarcoma phyllodes, malignant (C50.)
9040	Synovial sarcoma, NOS; synovioma, NOS; synovioma, malignant
9041	Synovial sarcoma, spindle cell
9042	Synovial sarcoma, epithelioid cell
9043	Synovial sarcoma, biphasic
9044	Clear cell sarcoma (except of kidney M8964/3)
9120	Haemangiosarcoma, Angiosarcoma of soft tissue
9130	Haemangioendothelioma, NOS, Kaposiform haemangioepithelioma
9133	Epithelioid haemangioendothelioma, malignant
9135	Endovascular papillary angioendothelioma
9136	Spindle cell hemangioendothelioma
9140	Kaposi sarcoma; Multiple haemorrhagic sarcoma
9150	Haemangiopericytoma, NOS
9170	Lymphangiosarcoma; lymphangioendothelial sarcoma
9174	Lymphangiomyomatosis
9180	Osteosarcoma, NOS (C40._, C41._)
9181	Chondroblastic osteosarcoma (C40._, C41._)
9182	Fibroblastic osteosarcoma (C40._, C41._); osteofibrosarcoma (C40._, C41._)
9183	Telangiectatic osteosarcoma (C40._, C41._)
9184	Osteosarcoma in Paget's disease of bone (C40._, C41._)
9185	Small cell osteosarcoma (C40._, C41._)
9186	Central osteosarcoma (C40._, C41._);
9187	Intraosseous well differentiated osteosarcoma (C40._, C41._)
9190	juxtacortical osteosarcoma ICD-O-2
9192	Parosteal osteosarcoma (C40._, C41._)
9193	Periosteal osteogenic sarcoma (C40._, C41._)
9194	High grade surface osteosarcoma (C40._, C41._)
9195	Intracortical osteosarcoma (C40._, C41._)
9200	Aggressive osteoblastoma
9210	Osteochondroma
9220	Multiple chondromatosis, Chondromatosis NOS
9221	Juxtacortical chondrosarcoma (C40._, C41._)
9230	Chondroblastoma, malignant (C40._, C41._)
9231	Myxoid chondrosarcoma
9240	Mesenchymal chondrosarcoma

Morphology	Description
9242	Clear cell chondrosarcoma, (C40._, C41._)
9243	Dedifferentiated chondrosarcoma (C40._, C41._)
9250	Giant cell tumour of bone, NOS
9251	Giant cell tumour of soft parts, NOS
9252	Malignant tenosynovial giant cell tumour (C49._)
9260	Ewing's sarcoma, Ewing's tumour, Extraskeletal Ewing tumour
9261	Adamantinoma of long bones; tibial adamantinoma (C40.2)
9270	Odontogenic tumour
9290	Ameloblastic odontosarcoma: Ameloblastic fibrodentinosarcoma
9310	Ameloblastoma
9330	Ameloblastic fibrosarcoma: Ameloblastic sarcoma: Odontogenic fibrosarcoma
9341	Clear cell odontogenic tumour
9342	Odontogenic carcinomsarcoma
9364	Peripheral neuroectodermal tumour; neuroectodermal tumour, NOS
9365	Askin tumour
9370	Chordoma
9371	Chondroid chordoma
9372	Dedifferentiated chordoma
9373	Parachondroma
9473	Primitive neuroectodermal tumour
9540	Malignant peripheral nerve sheath tumour MPNST, NOS
9560	Malignant peripheral nerve sheath tumour with thabdomyoblastic differentiation
9561	Perineurioma, malignant; Perineural MPNST
9571	Granular cell tumour, malignant; granular cell myoblastoma, malignant
9580	Granular cell tumour, malignant
9581	Alveolar soft part sarcoma