



## The Completeness of Bone Sarcoma data in the National Cancer Data Repository

Tumours diagnosed between 2006 and 2008

West Midlands Cancer Intelligence Unit

### Completeness of cancer registry data 2008 – Bone Sarcomas

1	DATA COMPLETENESS SUMMARY	. 3
1.1	Key findings	. 3
1.2	Executive Summary	.4
2	INTRODUCTION.	. 5
3	DATA	. 5
4	RESULTS: COMPLETENESS OF DATA FIELDS	. 6
4.1	Patient details	. 6
4.1.1	Sex	. 6
4.1.2	2 Date of birth	. 6
4.1.3	NHS number	. 6
4.1.4	Ethnicity	. 6
4.2	Tumour details	.7
4.2.1	Morphology	.7
4.2.2	2 Morphology coding system	.7
4.2.3	B Laterality	. 8
4.2.4	Detailed Site Code	. 8
4.3	Diagnosis data	. 9
4.3.1	Basis of diagnosis	. 9
4.3.2	2 Cases registered from more than a Death Certificate	. 9
4.3.3	B Diagnosis dates	. 9
4.4	Treatment details1	10
4.4.1	Surgery1	10
4.4.2	2 Radiotherapy 1	10
4.4.3	3 Chemotherapy 1	11
4.4.4	Neo-adjuvant therapy1	11
4.5	Death details1	12
4.5.1	Cause of death1	12
4.5.2	2 Place of death1	12
4.6	Staging data1	13
4.6.1	Tumour size1	13
4.6.2	2 T component1	13
4.6.3	3 Nodes Examined 1	14
4.6.4	Nodes Positive 1	14
4.6.5	5 N component 1	14
4.6.6	6 Metastases 1	15
4.6.7	7 M component1	15
4.6.8	3 Grade 1	15
4.6.9	TNM stage1	16
4.6.1	0 Staging audit1	16

### 1 DATA COMPLETENESS SUMMARY

- Key demographic data (sex, age at diagnosis, name) were well recorded by all registries.
- Staging data are incomplete, with an overall stage submitted for only 1% of tumours.
- Although key components of stage such as tumour size, presence of metasteses and grade were more complete, these did not appear to be well utilised to derive stage.
- All registries with the exception of the WMCIU and NYCRIS used ICD-O2 for the whole period covered by this report.
- The majority of registries provided a valid sarcoma morphology code for over 90% of their cases; OCIU, SWCIS and NWCIS did not.
- Treatment data vary widely across the country. It was not possible to separate out variation in the provision of treatment to bone cancer patients from variation due to poor data quality and completeness, and this must be taken into account when analysing NCDR data
- Death data (cause and place of death) were well submitted for the majority of registries, although one registry submitted no place of death data.

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

#### **1.2 Executive Summary**

Table 1.1 shows the proportion of data items which hold a valid value.

	% Complete	Registry							
	Data item	NYCRIS	Trent	ECRIC	Thames	OCIU	SWCIS	WMCIU	NWCIS
÷ "	Sex	100%	100%	100%	100%	100%	100%	100%	100%
ien ails	Age at diagnosis	100%	100%	100%	100%	100%	100%	100%	97%
oat det	NHS number	99%	99%	99%	97%	100%	98%	88%	99%
· ·	Ethnicity	58%	91%	61%	77%	84%	86%	64%	83%
5 0	Morphology	97%	94%	<mark>91%</mark>	96%	85%	79%	96%	87%
not ail\$	Morphology coding system (ICDM 3)	36%	0%	0%	0%	0%	0%	98%	1%
det	Laterality	97%	<mark>93%</mark>	93%	97%	78%	84%	100%	85%
	Detailed Site Code	82%	<mark>91%</mark>	91%	86%	85%	88%	98%	88%
sis	Basis of diagnosis	95%	87%	90%	93%	84%	78%	94%	84%
iagnos ormati	Cases registered from more than a death certificate	100%	98%	99%	98%	99%	100%	97%	99%
D İri	Diagnosis dates	100%	100%	100%	97%	100%	100%	91%	96%
ut	Surgery	72%	50%	53%	78%	57%	53%	65%	44%
me Ita	Radiotherapy	15%	6%	14%	12%	18%	7%	7%	22%
eat da	Chemotherapy	43%	35%	27%	33%	17%	33%	40%	33%
Tr	Neo-adjuvant therapy	0%	0%	0%	0%	0%	0%	25%	0%
ath ata	Cause of death	100%	70%	96%	95%	85%	97%	93%	96%
De dâ	Place of death	96%	98%	97%	0%	69%	69%	95%	92%
	Tumour size	0%	0%	16%	12%	0%	6%	43%	1%
	T component	0%	0%	1%	1%	0%	1%	0%	4%
ata	Nodes examined	0%	0%	1%	2%	0%	1%	3%	0%
q	Nodes positive	2%	0%	1%	0%	0%	1%	3%	3%
ing	N component	0%	0%	1%	0%	0%	0%	1%	3%
ag	Metastases ("Yes" or "No")	12%	0%	0%	70%	1%	4%	0%	1%
ŭ	M component	0%	0%	1%	0%	1%	4%	0%	6%
	Grade	19%	0%	32%	6%	19%	25%	56%	14%
	TNM stage	0%	0%	0%	0%	1%	4%	2%	4%

Table 1.1: Summary of completeness of bone cancer related data items within the NCDR

Each completeness statistic for each cancer registry was rated as 'red', 'amber' or 'green'. For sections one, two, three and five in Table 3.1, the following cut off points were used:

.

Кеу	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 treatment in section four, and the basis of diagnosis statistics in section three, as it is not clear what an acceptable rate 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 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 National Cancer Data Repository are understood, and that the completeness and accuracy of the data items submitted by each registry is evaluated. This also provides an opportunity to identify and address issues at an early stage.

This report focuses on the completeness and accuracy of data items collected for bone sarcoma (soft tissue sarcomas will be 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 bone sarcoma have been analysed in this report.

#### 3 DATA

This data completeness report analyses the most recent edition of the National Cancer Data Repository, holding all tumours diagnosed between 1990 and 2008.

Bone sarcomas were classified using the ICD10 coding system - all tumours where the first three digits of the site codes are C40 or C41 are included within the report. Cancers with any recorded morphology were included within the analysis whether or not they were a valid primary bone sarcoma morphology. This problem is discussed in more detail in section 4.2.1.

Extracting only these bone sarcomas from the NCDR results in 8,092 tumours (8,070 patients) diagnosed between 1990 and 2008. Only sarcomas diagnosed in the latest three years (2006 – 2008) were 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 1,519 tumours diagnosed in England in this time period, which break down by registry as follows:

Cancor Bogistry (data source)	Registry	No. of	
	(abbreviated)	tumours	
North West Cancer Intelligence Service	NWCIS	189	
West Midlands Cancer Intelligence Unit	WMCIU	142	
South West Cancer Intelligence Service	SWCIS	251	
Oxford Cancer Intelligence Unit	OCIU	96	
Thames Cancer Registry	Thames	361	
Eastern Cancer Registry and Information Centre	ECRIC	159	
Trent Cancer Registry	Trent	150	
Northern and Yorkshire Cancer Registry Information Services	NYCRIS	171	

Table 3.1: number of tun	ours diagnosed within	each registry (	2006 - 2008)
			,

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 patients who reside in England are included within the analyses.

#### 4 RESULTS: COMPLETENESS OF DATA FIELDS

#### 4.1 Patient details

#### 4.1.1 Sex

This data item was very well completed, with only one tumour out of the 1,519 on the NCDR where the patient's sex was unknown.

#### 4.1.2 Date of birth

All 1,519 tumours on the NCDR were supplied with a complete date of birth field. A very small proportion of tumours (6 tumours (<1%), all from the NWCIS) 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 (1,480, 97%) on the NCDR had an NHS number. Seven out of the eight English cancer registries had NHS number completeness of greater than 90%.

The WMCIU had no NHS number for 17 of their 142 sarcomas. As the report was written by the WMCIU, we were able to investigate these cases in detail. The data submitted to the NCDR had cases flagged as in-region if their region was the West Midlands, or if their region



was blank (it was assumed that a blank region defaulted to the West Midlands). However, some of the cases with a blank region were non-English cases from Northern Ireland and the Channel Isles. The Royal Orthopaedic Hospital is one of only five specialist bone sarcoma units in England so will treat patients from outside England. These non-English cases should not have an NHS number, but should also not be flagged as in-region cases. The WMCIU is now aware of the problem and will change the data extraction processes for future iterations of the NCDR.

### 4.1.4 Ethnicity

Approximately three-quarters of tumours (1156, 76%) had a valid ethnicity code on the NCDR. 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 Trent achieving 91% completeness for ethnicity and NYCRIS achieving only 58%.

Figure 4.1.2 Ethnicity completeness by cancer registry



Ethnicity data on the NCDR is 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.

#### 4.2 Tumour details

#### 4.2.1 Morphology

Site codes C40 and C41 are site codes reserved for diagnoses of primary bone cancer. All primary bone cancers are sarcomas.

However, of the 1,519 bone cancers recorded in the NCDR, only 1,380 (91%) were coded as sarcomas. A further 35 tumours (2%) were coded to a non-sarcoma morphology, including carcinomas, melanomas, and teratomas. There is a known



problem with sacrococcygeal teratomas, which some registries are assigning to bone cancer codes. The WMCIU recommends that these cases are coded to C76.3 (overlapping sites of the pelvis) but there is no nationally agreed method of coding these cancers. This will be discussed with the Site Specific Clinical Reference Group. 104 tumours were allocated a non-specific morphology code ("Neoplasm, malignant", or "Tumour cells, malignant"), and 3 tumours possessed no morphology code at all. These 107 tumours were classed as 'unknown' morphology.

Morphology completeness varies between the registries, with the SWCIS failing to submit a valid sarcoma morphology for over 20% of their cases. Use of a carcinoma code for primary bone cancer is particularly worrying, as this could indicate that in these cases the tumour registered was not a primary malignancy. The three tumours with no morphology code were all registered by the WMCIU from death certificates with no supporting information available. (See section 4.3.2). Just eight of the remaining tumours with invalid bone sarcoma morphology codes were registered from a death certificate.

### 4.2.2 Morphology coding system

All tumours submitted to the NCDR with morphology (1,516) were also submitted with a corresponding morphology coding system.

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).



Fig 4.2.2 clearly demonstrates that this has not yet occurred as only the WMCIU and NYCRIS submitted data using the most up to date coding system. However, the UK Association of Cancer Registries only recommended the move to ICD-O3 from 2008 onwards, and so registries still submitting in ICD-O2 for 2006 and 2007 are not a cause for concern. This is not a major problem for bone sarcoma, as the changes between the coding systems are minimal.

#### 4.2.3 Laterality

Only tumours diagnosed to the limbs (770 tumours) were included in this analysis, as laterality is not a relevant data item for other sites.

Overall, 92% of the tumours included within the analysis contained a valid laterality value. There was clear variation between the registries, with over 20% of cancers registered by OCIU having no laterality recorded.

#### 4.2.4 Detailed Site Code

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. "C40" – neoplasm of bone and cartilage of limbs) and the 4th digit gives a more detailed site code (e.g. "C40.1 – short bones of the upper limb). A 4<sup>th</sup> digit of "9" means that the detailed site is unspecified.



Figure 4.2.4 Detailed site completeness by cancer registry



An accurate site code (where the last digit of the site code does not equal "9") was present for 1,338 (88%) of the tumours. The variation amongst the registries ranged from 82% (NYCRIS) to 98% (WMCIU). Just 6 of the tumours with an "unspecified" site code were "Death Certificate Only" cases.

### 4.3 Diagnosis data

#### 4.3.1 Basis of diagnosis

A basis of diagnosis was submitted for 1,509 out of the 1,519 tumours. Thus, there were just 10 tumours where the basis of diagnosis was unknown.

However, the percentage of tumours which had had a histological diagnosis varied widely between registries (78% in SWCIS to 94% within the WMCIU and NYCRIS). A histological diagnosis will



Figure 4.3.1 Basis of diagnosis completeness by

always provide more reliable information on morphology and behaviour than a clinical diagnosis. This is particularly important for bone sarcoma, where there is a possible confusion in coding between primary bone sarcomas and secondary bone cancers. However, registries with a low percentage of clinical diagnosis may be missing cases which were diagnosed clinically but where it was not felt clinically appropriate to perform more invasive investigations.

### 4.3.2 Cases registered from more 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, and these are referred to as "Death Certificate Only" cases. Death Certificate Only cases are problematic, as they suffer from coding problems, contain no detailed information about the tumour or its treatment, and may indicate that the registry is missing live cases as well as dead cases.



Of the 1,519 tumours diagnosed during the period of interest, there are only 21 where the death certificate is the only source of information.

#### 4.3.3 Diagnosis dates

1,488 (98%) of the tumours were supplied with complete diagnosis dates. Of the 31 tumours submitted with imputed elements, 20 had a missing "day", 3 had an incomplete day and month, and the remaining 8 were uncertain as to whether imputation had taken place. Figure 4.3.3 Diagnosis dates completeness by cancer registry



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

#### 4.4 **Treatment details**

#### 4.4.1 Surgery

The percentage of patients known to have had surgery varied widely between registries, from 44% to 78%.

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.



There were two different approaches to submitting data on patients who have not had surgery. ECRIC and Trent appear to have taken the analytical approach that, although the registry has not received evidence that the patient was treated surgically, it has not received evidence that the patient wasn't surgically treated, and so have left the surgery field blank for many cases. Other cancer registries have consistently submitted 'no surgery' for any case where there was no positive evidence of surgery.

The surgery data submitted by each registry should relate only to surgical resections undertaken within six months of the diagnosis.

#### 4.4.2 Radiotherapy

Radiotherapy could be positively identified for 187 (12%) of tumours. The proportion of tumours receiving radiotherapy is consistently low across all registries, although there is wide variation, from only 6% of patients resident in Trent to 22% of patients registered in NWCIS.



Figure 4.4.2 Radiotherapy completeness by cancer registry

The radiotherapy data submitted by each registry should relate to radiotherapy sessions delivered within six months of the diagnosis.

As discussed in 4.4.1 for surgery, the value 'no radiotherapy treatment' was approached differently between the registries.

#### 4.4.3 Chemotherapy

Across all registries, confirmation of tumours receiving chemotherapy was present for 505 (33%) of the tumours. There was a wide variation between registries, with only 17% of cancers registered by the OCIU recorded as receiving chemotherapy.

The chemotherapy data submitted by each registry should relate to chemotherapy sessions administered within six months of diagnosis.

As discussed in 4.4.1 for surgery, the value 'no chemotherapy treatment' was approached differently between the registries

#### 4.4.4 Neo-adjuvant therapy

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

The only cancer registry to identify that patients are receiving neo-adjuvant

therapy was the WMCIU, where 36 (25%) of the tumours were positively identified as receiving neo-adjuvant therapy. This was determined by comparing the dates of surgery and chemotherapy, with the latter occurring before the former if the patient had neo-adjuvant therapy.



Figure 4.4.4 Neo-adjuvant flag completeness by cancer registry



#### 4.5 Death details

#### 4.5.1 Cause of death

Considering only the tumours where the patient was known to have died (573), 529 (92%) had a valid cause of death code.

The cause of death information supplied on the death certificates are registered as ICD10 codes. This information has always been provided on death certificates and the NCDR contains four causes of death fields. However, not all patients will have four causes of death



# Figure 4.5.1 Cause of death completeness by cancer registry

completed. Therefore, for the purposes of these analyses, only the first cause of death field was analysed.

There was wide variation between the registries with only 70% of cases from Trent having a cause of death recorded

#### 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 is present for 396 (69%) of the tumours.

Thames cancer registry appear not to have submitted this information to the NCDR.

Figure 4.5.2 Place of death: completeness by cancer registry NWCIS WMCIU SWCIS OCIU Thames ECRIC Trent NYCRIS 0% 20% 40% 60% 80% 100% Hospital Incomplete

### 4.6 Staging data

Cancer registration staging data have been historically incomplete. While Performance Indicators have improved staging for common cancer sites like colorectal cancer, rarer cancer sites such as bone sarcomas have not been a priority for registry staging data. A new Performance Indicator, introduced for the 2009 data, will monitor the percentage completeness of all staging data for all cancer sites. It is hoped that data completeness of bone cancer stage will improve because of this. However, the data analysed in this report were collected before this PI was introduced.

There is no national consensus on the staging system for bone sarcomas. At the National Clinical Leads meeting (February 2011) some centres reported they were using TNM, and some reported that they were using Enneking.

The NCDR did not request, and had no data field for, storing data on Enneking stage. However, it did have fields for the collection of the grade of the tumour, T, N and M components and the overall TNM stage (both pathological, clinical, and integrated). The TNM system utilises 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, clinical, pathological, or integrated, the corresponding tumour was presented as having staging information submitted.

#### 4.6.1 Tumour size

Of the 1,519 tumours diagnosed between 2006 and 2008, only 147 (10%) had a tumour size recorded in mm. The range of tumour sizes varied from 1mm to 305 mm.

There was large variation between registries, with the WMCIU collecting size for almost half their tumours, and three registries submitting no size at all.

#### 4.6.2 T component

Of the 1,519 tumours diagnosed between 2006 and 2008, just 16 (1%) 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. For tumours registered within





NWCIS 4% contained a valid "T" value, yet only 1% contain a tumour size. Conversely, 43% of the tumours registered within the WMCIU have a tumour size, yet 0% was submitted as having a T component. As the T stage for TNM can be derived directly from the size of the tumour, this shows that the full power of the data collected by registries is not being exploited.

#### 4.6.3 Nodes Examined

Cancer registry data on the number of nodes examined was incomplete. The number of nodes examined was present for just 13 (<1%) of all tumours. Information was considered complete if the field relating to nodes was not blank.

#### 4.6.4 Nodes Positive

Cancer registry data on the number of positive nodes was also drastically incomplete.

There were more tumours (17) submitted with positive nodes, than there were tumours submitted with a number of nodes examined (13). This may be a true limit in the data received by registries (for example 'nodes examined, 2 positive' does not report



Figure 4.6.3 Nodes examined completeness by

## Figure 4.6.4 Nodes positive completeness by registry



the number of nodes examined) or it may be a sign that registries are not fully utilising the data they receive.

#### 4.6.5 N component

The N component of stage was also incomplete, with only 9 tumours (<1%) having either a clinical, pathological or integrated N stage (the presence or absence of metastasis in the regional lymph nodes).

Again, there are inconsistencies between 4.6.4 and 4.6.5 – for example, the WMCIU has cases where it is known the nodes have been examined and are

Figure 4.6.5 'N' component completeness by cancer registry



positive, but they have no N component of stage, and NYCRIS has cases with an N component of stage, but where it is not stated that nodes have been examined,

#### 4.6.6 Metastases

Overall only 295 of the 1,519 tumours (19%) had a flag which clearly stated whether metastases were present or not. There was large variation between registries, with the majority of these cases coming from Thames cancer registry.



#### 4.6.7 M component

Of the 1,519 bone tumours diagnosed just 23 tumours (1.5%) had either a clinical, pathological or integrated "M" component.

The "M" value relates to the presence or absence of distant metastases. Comparing Figure 4.6.7 to Figure 4.6.6 it is evident that the information supplied relating to metastases is inconsistent Figure 4.6.7 'M' component completeness by cancer registry



across registries. For example the Thames registry reported in section 4.6.6 information on metastases was available for 70% of cases, and yet an M component of TNM was not submitted for any tumours.

#### 4.6.8 Grade

There were only very limited data available on tumour grade in the NCDR, with only 295 tumours (19%) having a grade submitted.

Completeness ranged widely between registries, with even the best performing registry (WMCIU) only submitting a grade for 56% of tumours, and the worst performing registry (Trent) submitting no grades at all.



#### 4.6.9 TNM stage

Fig 4.6.9 clearly demonstrates that the overall TNM staging for bone sarcoma is incomplete across all registries, with just 21 tumours (1.4%) having an overall TNM stage recorded in the NCDR (either clinical, pathological or integrated).



#### 4.6.10 Staging audit

Each of the cancer registries in England was contacted to establish whether bone cancer staging data are routinely collected, and if so, when did the collection of staging data commence and which system is used.

The results of the survey can be seen in Table 4.6.10

Registry	Is staging data collected?	From what period were/will staging data be collected?	What system is used for staging?				
NWCIS	Only if stated on pathology report or received in MDT data	Liverpool from 2008. Manchester always recorded	ТММ				
WMCIU	CIU Historically received Enneking staging from the Royal Orthopaedic Hospital. The earliest staging data within the West Midlands cancer registry dates from 1990		The WMCIU is in a position to record both TNM and Enneking stage. However, this information is not currently supplied on pathology reports.				
SWCIS	Only if stated on pathology report or received in MDT data	Pathology from 2004, MDT data from 2009	MDT would be TNM				
οςιυ	Concentrated on mandatory sites: colorectal, breast, cervix, melanoma. From 2010, any pathology report with staging information will be included						
Thames	Did not respond						
ECRIC	Do not currently collect staging data for bone sarcoma as there is no specialist bone centre in the region and the staging data is not sent through from other registries as extra-regional information						
Trent	Did not respond						
NYCRIS	No. NYCRIS do no plan to prioritise staging data for bone sarcoma in the immediate future						