



BREAST CANCER  
**BREAKTHROUGH**



*Breast Cancer Clinical Outcome Measures*



West Midlands  
Cancer Intelligence Unit

# **Quantifying the Completeness of National Breast Cancer Data (cases diagnosed in 2006)**

## **Executive Summary**

**CONTENTS**

<b>1.0 BACKGROUND</b>	<b>1</b>
<b>2.0 DATA SOURCES</b>	<b>2</b>
2.1 Cancer registry dataset (UK)	2
2.2 Cancer registry survey	2
2.3 Hospital Episodes Statistics (HES) dataset (England only)	2
2.4 Dataset available at local hospitals/cancer network level	2
<b>3.0 RESULTS</b>	<b>2</b>
3.1 Data completeness measures	2
3.2 Assigning surgeons and hospitals	3
3.3 Completeness of cancer registry data	3
3.4 Supplementing cancer registry data with Hospital Episodes Statistics (HES) data	7
3.5 Validation of cancer registry data by local hospitals and Scottish cancer networks	7
<b>4.0 REGISTRY SURVEY RESULTS</b>	<b>9</b>
<b>5.0 CONCLUSIONS</b>	<b>9</b>

## **1.0 BACKGROUND**

The Breast Cancer Clinical Outcome Measures (BCCOM) Project aims to provide for all symptomatic breast cancers diagnosed in the UK, data to enable the generation of appropriate clinical outcome measures. The BCCOM initial dataset is generated using cancer registration data (Step1) before being sent for validation to breast surgeons who have agreed to participate (Step 2). Surgeons can then either submit their data unchecked or amend and add missing data. 76,500 symptomatic cases diagnosed in 2002-2006 have been submitted to BCCOM in the 5 years the Project has been running. Datasets amended by surgeons are returned to local cancer registries at the end of each audit round.

It was anticipated at the start of the BCCOM Project that, as clinicians became more confident in the completeness and accuracy of the cancer registry data, Step 2 would become unnecessary. However, it became apparent early on in the BCCOM Project that several registries collected only a limited amount of information, on some of the BCCOM data items. In some regions this lack of data meant that eligible cancer cases could not be sent to a surgeon for checking as the name of the treating surgeon had not been recorded. Data completeness was found to be very variable between regions and, whilst in some cases the completeness of the registry data increased in subsequent rounds of BCCOM, for some data item/regions, data completeness was not significantly improved when comparing cancers diagnosed in 2002 with those diagnosed in 2006.

As the National Cancer Intelligence Network (NCIN) lead registry for breast cancer, the West Midlands Cancer Intelligence Unit (WMCIU) has access to a breast cancer extract from the National Office for National Statistics/Hospital Episode Statistics (ONS/HES) database. In BCCOM Year 5 (cases diagnosed in 2006), to improve data completeness, data on surgery and chemotherapy were obtained from the ONS/HES extract and used to supplement the dataset for two cancer registries before sending the data to surgeons for validation. Preliminary results showed that the added HES data were accepted by participating surgeons as accurate.

In June 2009, the National Cancer Intelligence Network (NCIN) Breast Clinical Reference Group (CRG) commissioned the WMCIU to report on the accuracy of the data extracted from regional cancer registries for use in the BCCOM audit. In particular, to report on whether the cancer registration data were fit for the following purposes:

- to use to generate surrogate clinical outcome measures in order to audit the management of breast cancer
- to use on the NHS choices website to provide information for breast cancer patients and their carers on the treatment provided by different hospitals
- to use as part of the professional validation process for surgeons

A full report describing the results of this work was circulated to stakeholders in October 2009. Copies of this report can be obtained from Catherine Lagord at [Catherine.lagord@nhs.net](mailto:Catherine.lagord@nhs.net).

The purpose of this Executive Summary report is to summarise the main findings of the main data completeness report focussing in particular on:

- the completeness of the data collected by UK cancer registries for
  - data items included in the National Cancer Registration Dataset
  - additional data items included in the Royal College of Pathologists breast cancer dataset
  - additional data items included in the BCCOM dataset
- the improvements in data completeness that can be achieved by matching cancer registration data to the data items in Hospital Episode Statistics (HES) data
- the improvements in data completeness that can be achieved when data items added/changed by responsible surgeons are added to cancer registration data.

## 2.0 DATA SOURCES

### 2.1 Cancer registry dataset (UK)

For the BCCOM Project all UK cancer registries were asked to transfer to the BCCOM project team at the WMCIU data for all male and female breast cancer patients who were treated or diagnosed in hospitals within their catchment area. For a description of the BCCOM dataset, see the specification in Appendix 2 of the main report. This Executive Summary report summarises the completeness of the data provided by cancer registries for BCCOM Year 5 (latest audit round, cases diagnosed in 2006).

### 2.2 Cancer registry survey

Each cancer registry was sent a list of the data items for which they had not provided any data for BCCOM Year 5. For each data item not collected, cancer registries were asked if they had plan to collect the item in the future. Results of the survey are included in Section 4 of the main report.

### 2.3 Hospital Episodes Statistics (HES) dataset (England only)

For BCCOM Year 5 data from two cancer registries (the North West Cancer Intelligence Service and the Trent Cancer Registry) were matched against a breast cancer extract from the National ONS/HES database. Data on surgery and chemotherapy were obtained from HES and used to supplement the cancer registry dataset. This combined dataset was sent to surgeons for validation; surgeons could either submit their data to BCCOM unchecked or amend and add missing data.

### 2.4 Dataset available at local hospitals/cancer network level

For England, Northern Ireland and Wales, data generated using cancer registration data with or without additional HES data were sent for validation to breast surgeons who had agreed to participate. Surgeons could either submit their data unchecked or amend and add missing data. After exclusion of screen-detected breast cancers, cases for which a surgeon could not be identified and cases treated by surgeons with a low caseload (less than 6 symptomatic cases), two-thirds of symptomatic cancers remained eligible for inclusion in BCCOM. For Scotland, the data used in BCCOM were the data that had been contributed to the Scottish breast cancer audit (2006). Only data from consenting surgeons were included in the BCCOM dataset.

## 3.0 RESULTS

### 3.1 Data completeness measures

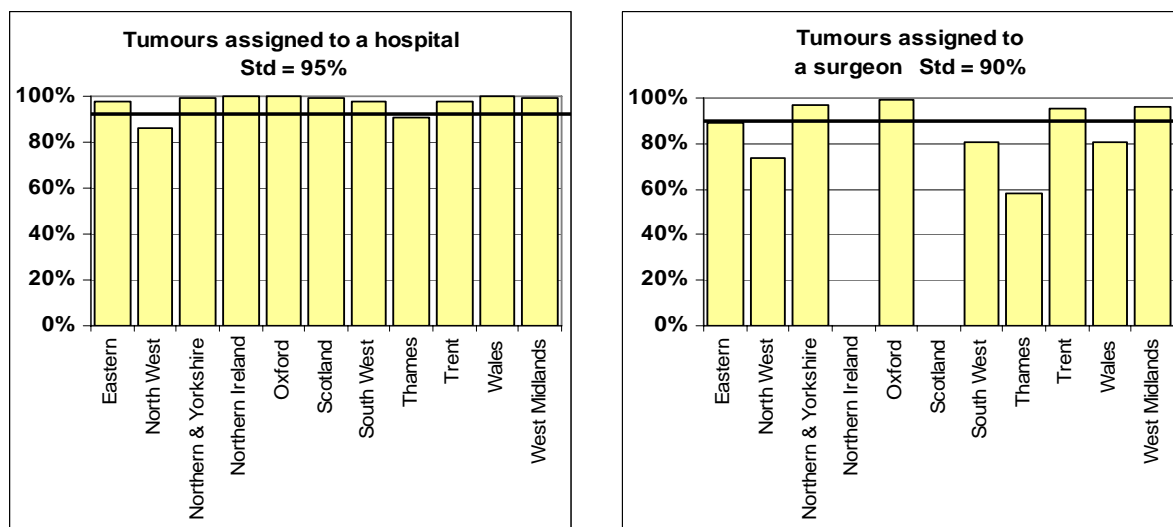
**Measure of completeness:** a tumour was generally deemed to have 'complete' information for a specific data item if the information provided was not 'unknown'. For data items involving information on treatment, cancer registries were not always able to distinguish between tumours which received 'no treatment' and tumours for which it was 'unknown if treatment was given'. Therefore, to allow comparison of results between registries, the proportions of tumours known to have had a particular type of treatment were measured rather than the proportion of tumours with known treatment

**Expected minimum completeness:** when the standard for an equivalent measure of completeness had already been set by cancer registries or other sources such as the NHSBSP audit of screen-detected breast cancers, the expected level of completeness was taken from these sources. For example, when a performance indicator (PI) had been set up by the UKACR as part of the Cancer Registry Peer Review process or the annual UKACR Performance Indicators exercise, the Cancer Peer Review Target or the UK average rate for the UKACR PI was used as the expected completeness in this report. When no relevant standard existed, a standard was suggested. Details of the sources of the standards used for each data item are in given the main report.

### 3.2 Assigning surgeons and hospitals

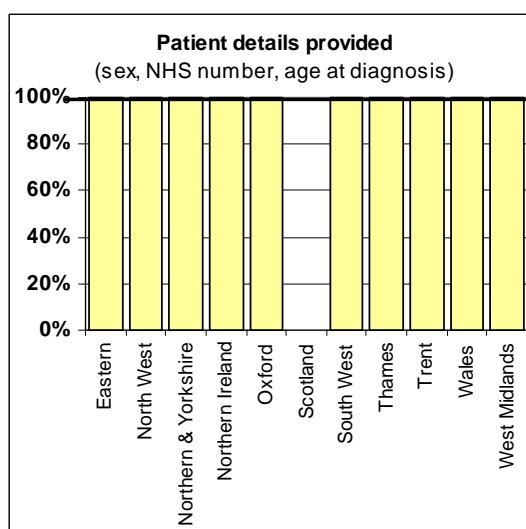
In order to send out the cancer registration data submitted to the BCCOM audit to surgeons for validation, the hospital of diagnosis and the GMC code of the 'responsible' or treating surgeon were required. Figure 3.1 shows the completeness of these data items in each registry.

Figure 3.1 Completeness of hospital and surgeon



In order to match cancer registration data to HES data and to NHSBSP data, patient identifiers were required. Figure 3.2 shows the completeness of these data for each registry.

Figure 3.2 Completeness of patient details



### 3.3 Completeness of cancer registry data

Figures 3.3-3.5 summarise the completeness of the data submitted to BCCOM Year 5 (2006 cases) by each UK cancer registry. Details of the cohorts included for each data item are given in the main report. Each figure shows the % completeness achieved for each data item by each registry relative to an expected minimum completeness standard. The data items are grouped into the following categories:

- data items included in the National Cancer Registration Dataset
- additional data items included in the Royal College of Pathologists breast cancer dataset
- additional data items included in the BCCOM dataset

Figure 3.3 Data items included in the National Cancer Registration Dataset

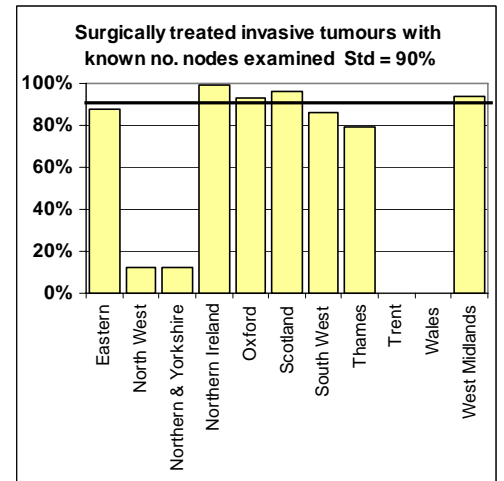
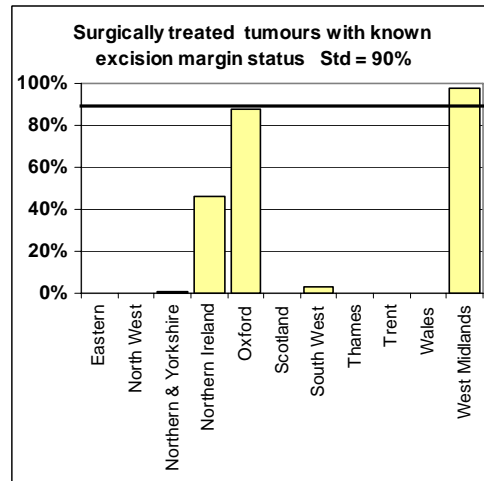
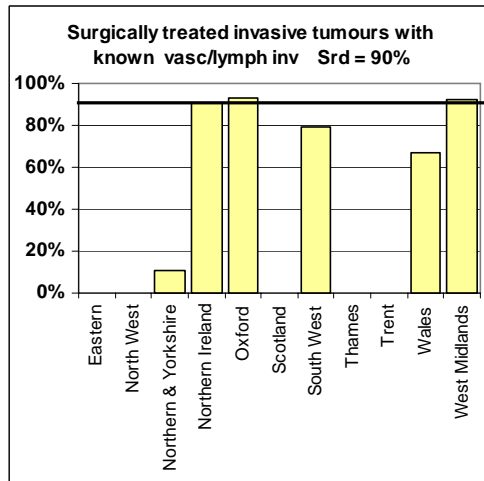
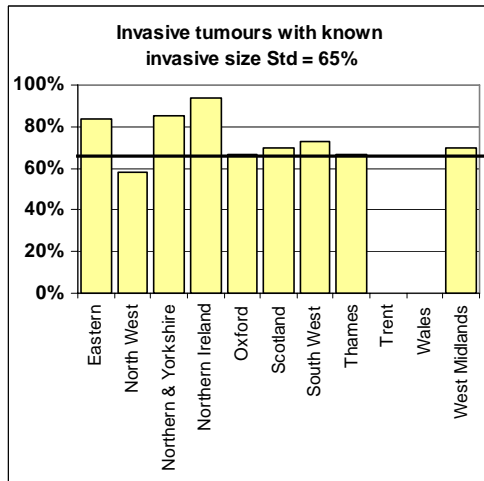
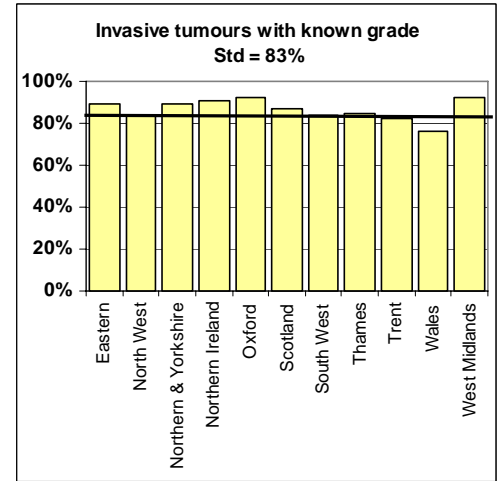
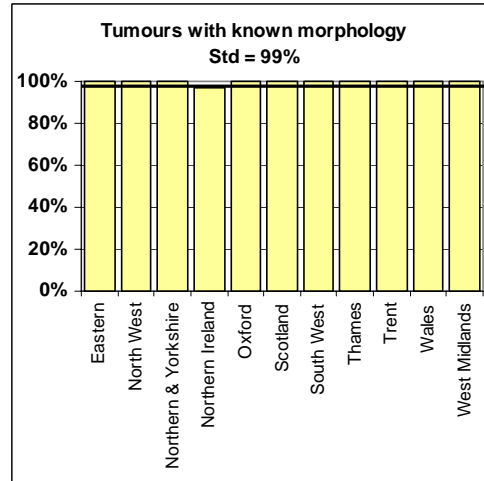
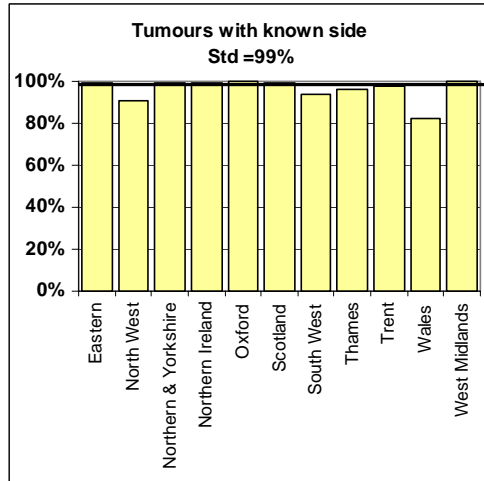
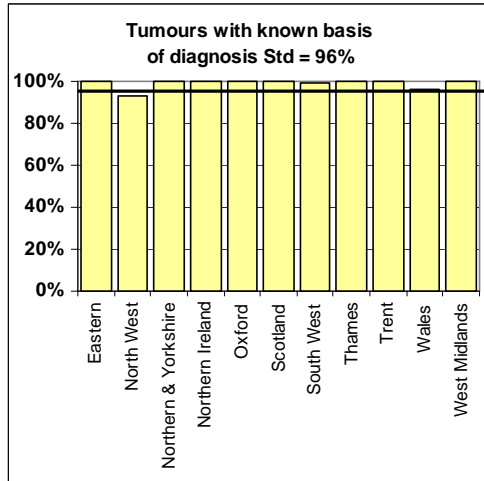


Figure 3.3 Data items included in the National Cancer Registration Dataset (cont)

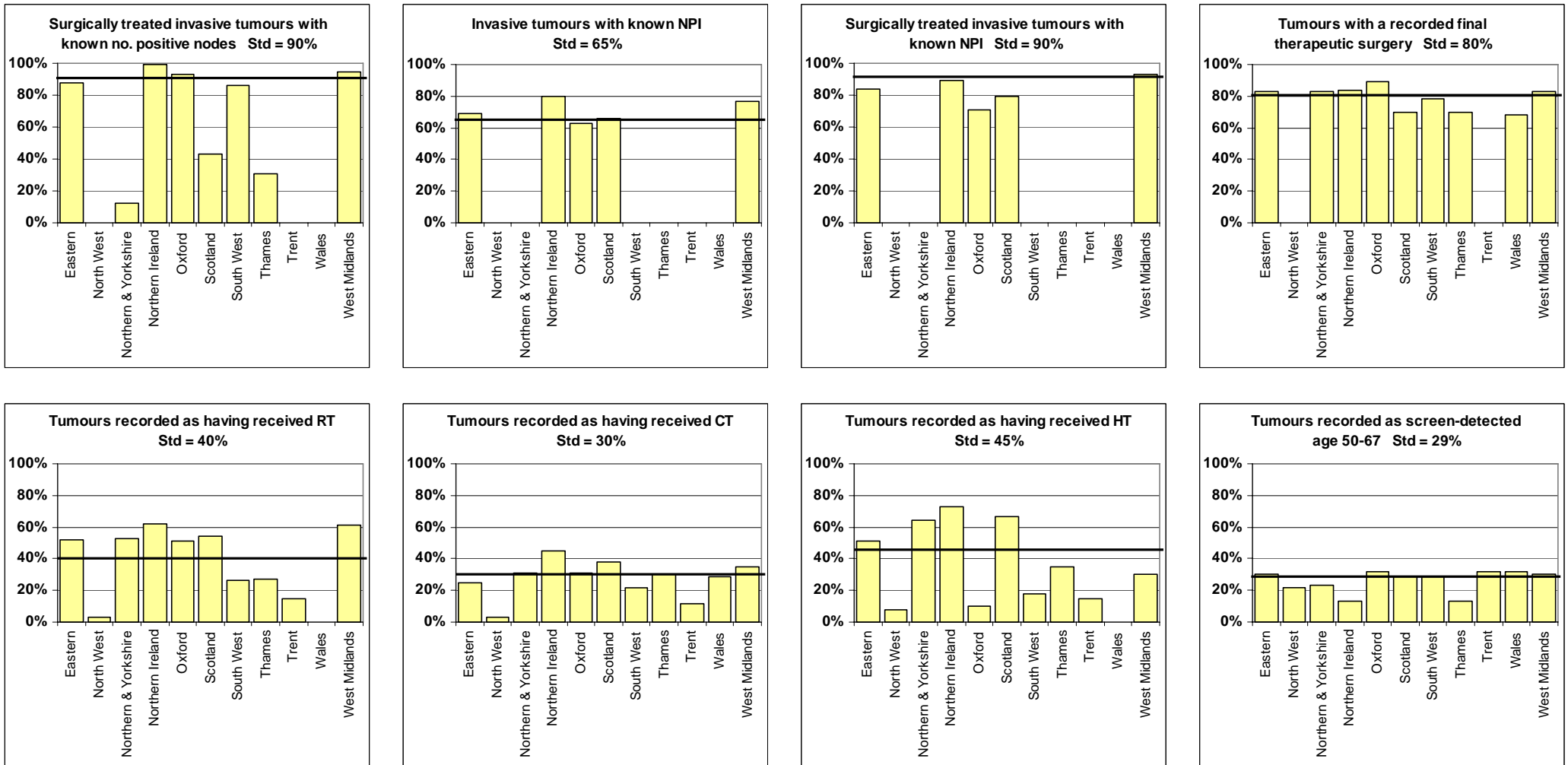


Figure 3.4 Additional data items included in the Royal College of Pathologists breast cancer dataset

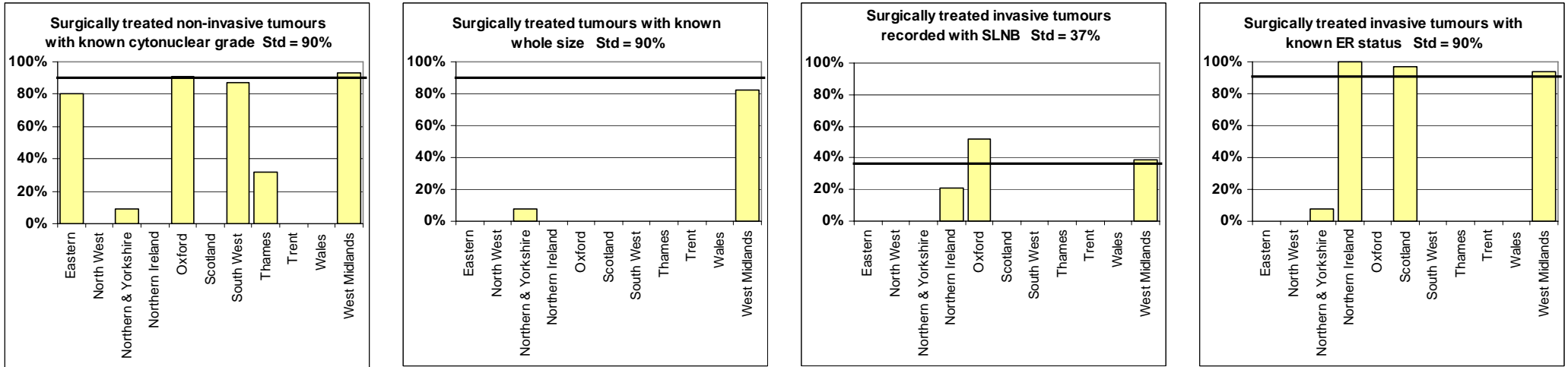
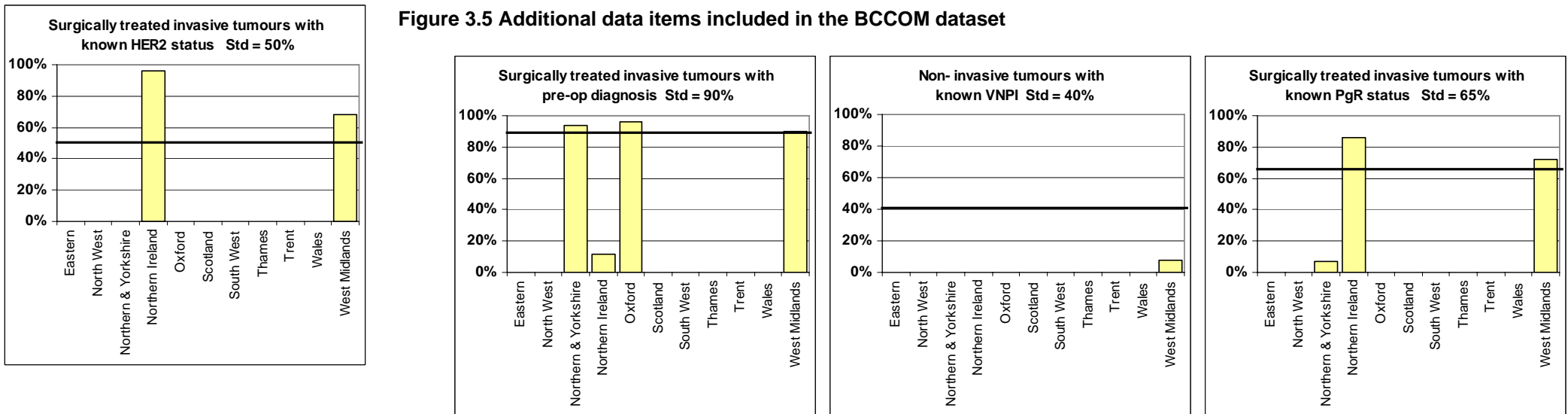


Figure 3.5 Additional data items included in the BCCOM dataset





### 3.4 Supplementing cancer registry data with Hospital Episodes Statistics (HES) data

HES data for patients registered at the Trent Cancer Registry and the North West Cancer Intelligence Service were extracted from the ONS/HES dataset and used to supplement the BCCOM/registry dataset. The rules used for the process were as follows:

- HES episodes were considered if they were within a period four days before the diagnosis date recorded by the registry (4 days limit set to account for registries defining a 'diagnosis date' on the basis of a post-surgery pathological report) to six months after the diagnosis date (six month limit set to exclude treatment to recurrences)
- HES information was used to supplement BCCOM/registry dataset only if the BCCOM/registry value was 'null' or 'unknown'.

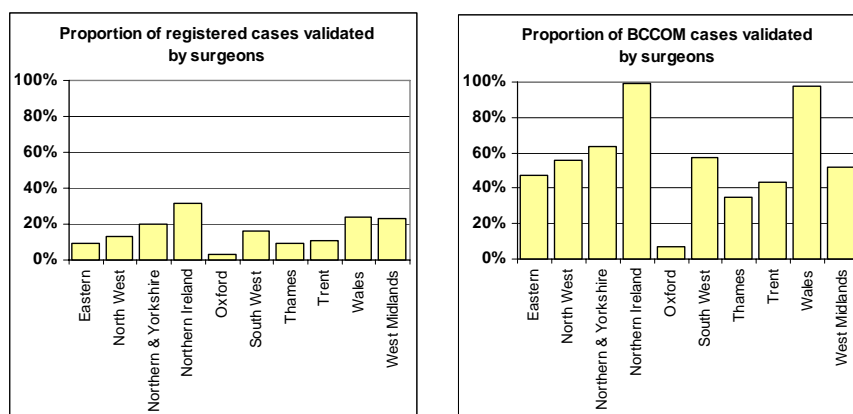
An additional 7,743 data items were added using this methodology. To evaluate the accuracy of the methodology, the combined dataset was sent to surgeons for validation. A sample of 1,014 validated cases submitted to BCCOM Year 5 by surgeons from the North West (419 cases) and Trent (595 cases) was analysed.

The type of final therapeutic surgery to the breast was identified from HES data for 70% of patients registered at the Trent Cancer Registry and the North West Cancer Intelligence Service (3,443 patients having breast conserving surgery (BCS) and 3,250 patients having a mastectomy). Comparison of the registry/HES data with data returned by validating surgeons showed that surgeons were mostly in agreement with the surgical information obtained from HES. For 37 (6%) tumours, Trent surgeons were able to supply information not recorded on HES and for one tumour, the final surgery was converted from breast conserving surgery (BCS) to a mastectomy. For 53 (13%) tumours, North West surgeons were able to supply information not recorded on HES and the final surgery was modified for four tumours.

From HES, a chemotherapy treatment episode was identified for an additional 7% of patients registered at the North West Cancer Intelligence Service and an additional 16% of patients registered at the Trent Cancer Registry. The lower level of chemotherapy information obtained from HES for North West patients needs further investigation. It is possible that consultant preference, local policy and clinical trials could affect the balance of out-patient/in-patient/day case treatment and this may contribute to the low proportion of chemotherapy events identified. Comparison of the registry/HES data with data returned by validating surgeons showed that Trent surgeons were mostly in agreement with the data, although they did identify additional chemotherapy events for 23 (4%) tumours and disagreed on four accounts with the information obtained from the registry/HES data. Overall, 227 (38%) cases submitted by validating Trent surgeons were flagged as having received chemotherapy.

### 3.5 Validation of cancer registry data by local hospitals and Scottish cancer networks

**Figure 3.6 Proportion of registered cases and cases submitted to BCCOM Year 5 that were submitted as having been 'validated' by surgeons (excluding Scotland)**



*West Midlands Cancer Intelligence Unit*

Data were submitted to BCCOM Year 5 by surgeons as either 'unchecked' or 'validated'. The number and proportion of cases validated by surgeons in each region are shown in Figure 3.6 and summarised in the following table. 6,645 cases were submitted as 'validated' by a total of 117 surgeons in England, Wales and Northern Ireland. Overall 15% of registered cases and 51% of cases submitted to BCCOM Year 5 were submitted as having been 'validated'. The proportion of registered cases that were 'validated' varied between 3% (Oxford) to 31% (Northern Ireland).

Cancer registry	Cases registered in 2006			Cases submitted to BCCOM Year 5	
	No. cases	Validated cases	% checked	No. cases	% checked
Eastern	4,774	451	9%	947	48%
North West	5,275	676	13%	1,205	56%
Northern & Yorkshire	5,466	1,109	20%	1,746	64%
Northern Ireland	1,019	320	31%	323	99%
Oxford	2,147	59	3%	860	7%
South West	6,870	1,126	16%	1,952	58%
Thames	8,341	772	9%	2,220	35%
Trent	4,179	464	11%	1,074	43%
Wales	2,614	614	23%	631	97%
West Midlands	4,614	1,054	23%	2,022	52%
<b>Total</b>	<b>45,299</b>	<b>6,645</b>	<b>15%</b>	<b>12,980</b>	<b>51%</b>

The following table summarises the number of cases where 'validating' surgeons have added to or changed the information submitted to BCCOM Year 5 by cancer registries.

Data item	Figure No.	Added information		Changed information	
		No. cases	% cases	No. cases	% cases
<b>National Cancer Registration Dataset items</b>					
Tumour side	1A	118	2%	8	0%
Morphology code	1B	4	0%	133	2%
Invasive grade (invasive tumours) (a)	1C	387	6%	81	1%
Invasive tumour size (invasive tumours) (b)	1D	1,358	22%	169	3%
Vascular/lymphatic invasion	1E	2,329	35%	22	0%
Excision margins	1F	3,675	55%	21	0%
Nodes examined	1G	2,730	41%	172	3%
Nodes positive	1H	2,332	35%	18	0%
NPI score (invasive tumours)	1I	1,550	25%	7	0%
Surgery	1J	575	9%	140	2%
Radiotherapy	1K	2,854	43%	35	1%
Chemotherapy	1L	2,798	42%	58	1%
Hormone therapy	1M	3,259	49%	34	1%
<b>Additional Royal College of Pathology Breast Cancer Dataset items</b>					
Cytoneuclear grade (non-invasive tumours) (c)	2A	165	37%	3	1%
Whole tumour size (invasive tumours) (d)	2B	3,130	47%	71	1%
Sentinel node procedure	2C	4,511	68%	2	0%
ER status	2D	4,208	63%	2	0%
HER2 status	2E	2,988	45%	6	0%
<b>Other data items in BCCOM dataset</b>					
VNPI score (non-invasive tumours)	3A	19	4%	0	0%
PR status	3B	3,087	46%	8	0%
Pre-operative diagnosis	3C	3,306	50%	75	1%

- (a) where surgeon and cancer registry agree that the tumour is invasive
- (b) where surgeon and cancer registry agree that the tumour is invasive; "change" if difference is >3mm
- (c) where surgeon and cancer registry agree that the tumour is non-invasive
- (d) "change" if difference is >3mm

Few surgeons modified registration data on tumour morphology and invasive grade. However, surgeons provided a large amount of new information on tumour size, receptor and nodal status in regions where these data items were not routinely collected. Validating surgeons also significantly improved the completeness of adjuvant treatment data; although this information remained incomplete for 20% of cases. Conflicting data on the number of nodes sampled and the type of surgery may be due to the fact that not all of the registries that collect these data items collect information on all operations. Addition of invasive tumour size may be because some cancer registries do not record tumour size if the cancer has had neo-adjuvant therapy, if the tumour margins are involved or if the tumour has been removed in more than one operation.

Figures 3.7-3.9 summarise for each cancer registry the results of the comparisons between the cases submitted as 'validated' and the original cancer registration data sent to the surgeons. The data items are again grouped into the following categories:

- data items included in the National Cancer Registration Dataset
- additional data items included in the Royal College of Pathologists breast cancer dataset
- additional data items included in the BCCOM dataset

The original data submitted by the cancer registries are shown in yellow and the data added or changed by surgeons in purple. The % completeness shown in each figure relates only to the 6,645 cases submitted as 'validated' by the 117 surgeons who returned their data as having been 'validated'. Details of all checked data will be returned to each originating cancer registry.

Figure 3.10 shows a comparison between the data for cases diagnosed in 2006 submitted by the Scottish Cancer Registry and data provided by the Scottish Cancer Networks which was collected for their 2008 annual audit. The original data submitted by the cancer registry are again shown in yellow and the data added or changed on the basis of data submitted by the networks in purple.

#### **4.0 REGISTRY SURVEY RESULTS**

Figure 3.11 summarises the results of the survey sent to registries requesting information regarding their future plans to collect data items not collected on their cancer registration databases for cases diagnosed in 2006. The data items collected by each registry are shown in green (Figures 3.3-3.5 show the level of completeness achieved for each of these data items by each registry). Data items not collected by each registry for 2006 cases and for which the registry has no firm date for collection to start are shown in red. Data items not collected for 2006 cases but for which a firm start date has been identified are shown in orange together with the year that data recording will start. Figure 3.11 shows that one registry has no plans to collect invasive tumour size, four registries have no plans to record vascular invasion or excision margin status, three registries have no plans to record NPI score, six registries have no plans to record ER status and eight registries have no plans to record HER2 status. Although some registries have no plans to record surgical treatment, sentinel lymph node biopsy procedures and surgeon, these data items can be obtained from HES or equivalent data in Scotland.

#### **5.0 CONCLUSIONS**

It is clear from this report that:

- a) HES data can be used to supplement and improve the data on surgical treatment and chemotherapy collected by cancer registries
- b) surgeons have access to many of the data items that are currently not well recorded by some cancer registries. These registries should be able to improve data completeness by liaising with surgeons to discuss whether their data sources could be accessed by the registries
- c) some registries currently have no plans to record data items such as invasive tumour size, vascular invasion, excision margin status, NPI score, ER status and HER 2 status, knowledge of which is crucial for the evaluation of patient outcomes and adherence to clinical guidelines.

Figure 3.7 Data added by surgeons to data items included in the National Cancer Registration Dataset

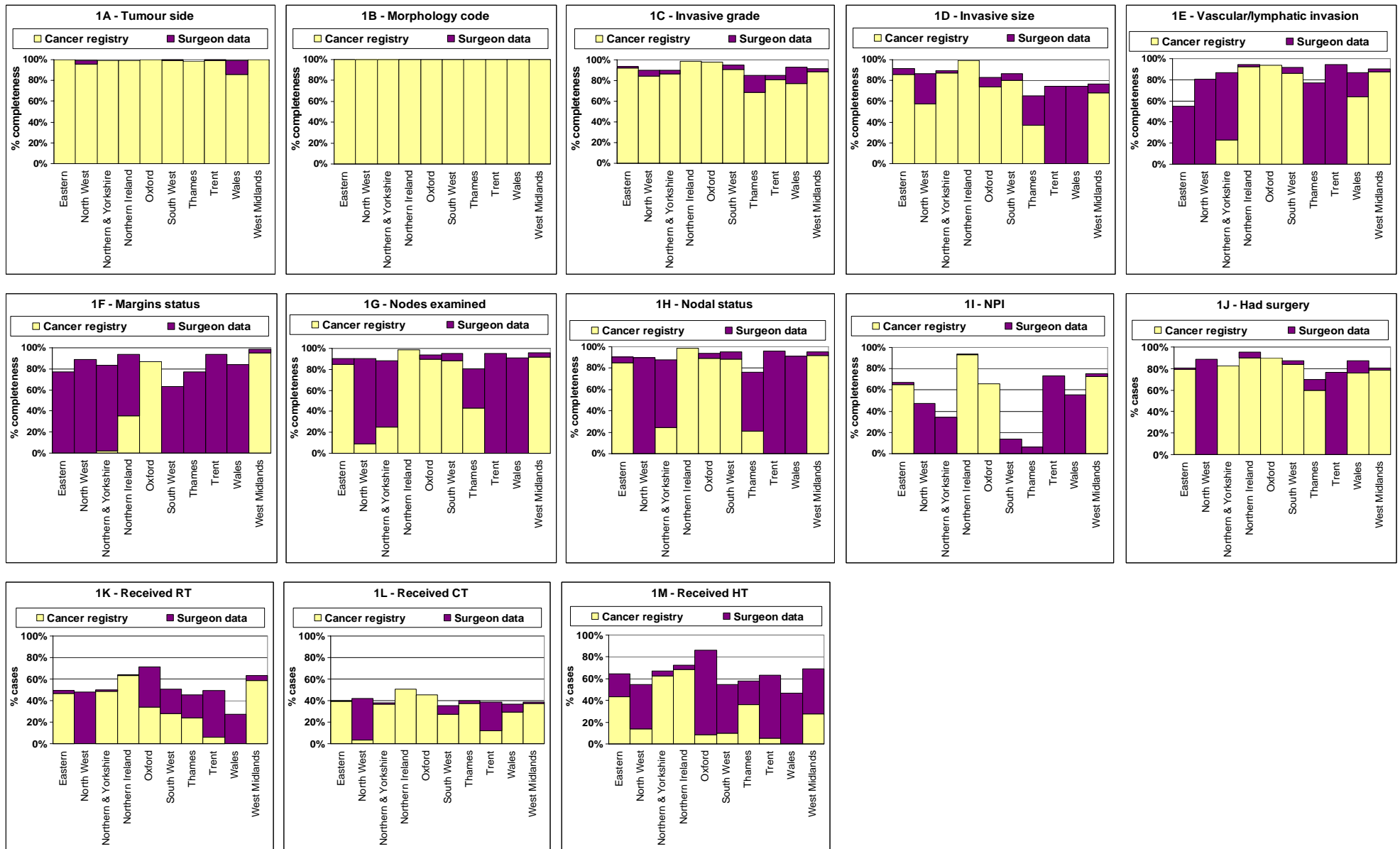


Figure 3.8 Data added by surgeons to data items included in the Royal College of Pathologists breast cancer dataset

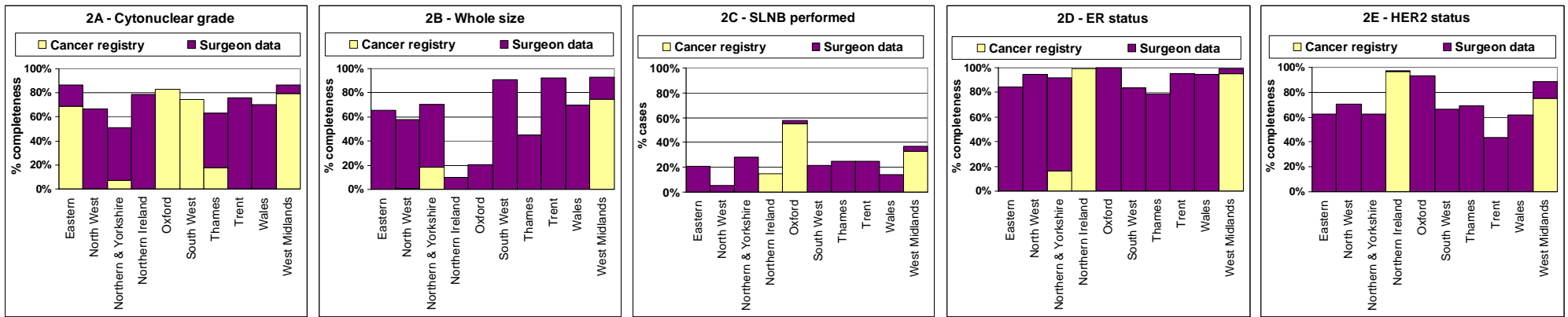


Figure 3.9 Data added by surgeons to data items included in the BCCOM dataset

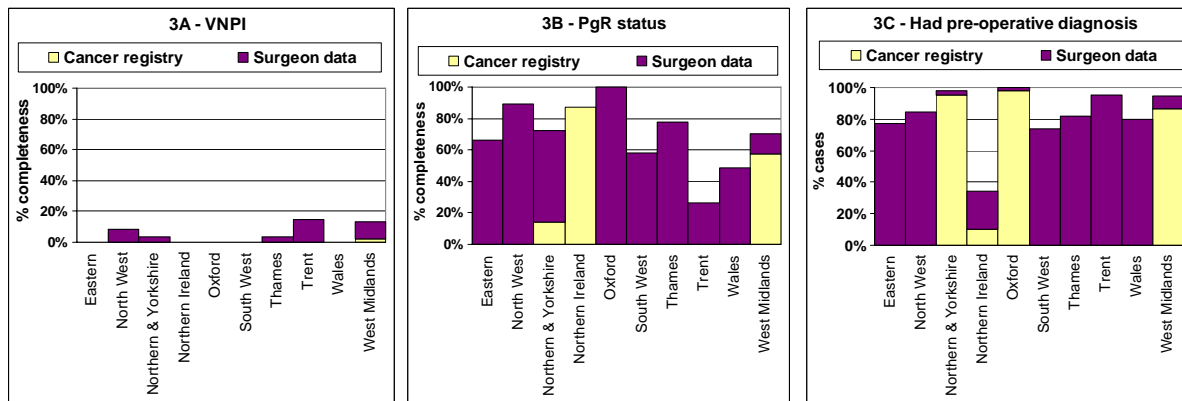


Figure 3.10 Comparison between Scottish breast cancer audit data and data supplied by the Scottish Cancer Registry

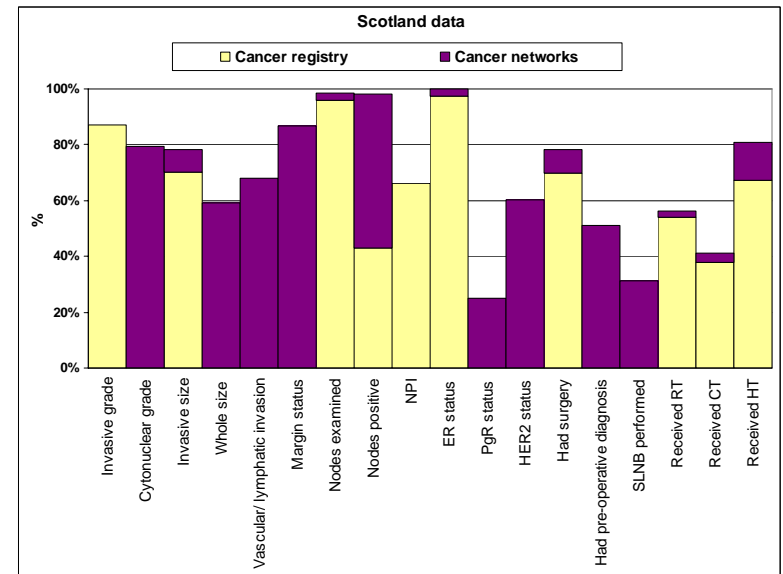


Figure 3.11 Results of the registry survey of future data collection plans

Data item not collected by one or more registries	Cancer Registry										H M S	
	Eastern	North West	Northern & Yorkshire	Northern Ireland	Oxford	Scotland	South West	Thames	Trent	Wales		West Midlands
<b>National Cancer Registration Dataset items</b>												
Invasive tumour size									2008	X		
Vascular/lymphatic invasion	X	2009				X		X	X			
Excision margins	X	2009				X		X	X	2009		
Nodes examined									X	X		
Nodes positive		2009							X	2008		
NPI score (invasive tumours)		X	2008				2006	X	X	2008		
Surgery		X							X			✓
Radiotherapy										2008		
Hormone therapy										X		
<b>Royal College of Pathologists Breast Cancer Dataset items</b>												
Cytoneuclear grade		X		X		X			X	X		
Whole tumour size (invasive tumours)	X	X		X	X	X	X	X	X	2008		
Sentinel lymph node procedure	2009	X	X			X	2007	X	X	2007		✓
ER status	2010	X			X		X	X	X	X		
HER2 status	2010	X	X		X	X	X	X	X	X		
<b>Other data items in BCCOM dataset</b>												
Breast surgeon				2006		X						✓
VNPI score	X	X	X	X	X	X	X	X	X	X		
PR status	X	X			X	X	X	X	X	X		
Pre-operative diagnosis	X	X				X	X	X	X	2010		

**X** = data item not collected for 2006 cases, no firm plans to collect

**date** = data item not collected for 2006 cases but firm start date for collection implemented or planned

**Green** = data item collected and provided for 2006 cases