

CAVEAT ON USE OF DAY OF DIAGNOSIS

Data Item: **Day of Diagnosis**

Date Updated: 11 April 2016

Definition: Day of diagnosis typically reflects the earliest definitive diagnostic or treatment episode received by the Cancer Institute NSW and may not be the exact day of diagnosis.

BACKGROUND

Prior to the integration of the Clinical Cancer Registry (ClinCR) with the Central Cancer Registry (CCR), and the subsequent creation of the NSW Cancer Registry (NSWCR), the Cancer Institute NSW only captured and stored month and year of diagnosis on its system databases for cancer incidence reporting. During the creation of the NSWCR, a series of business rules were developed in order to populate the day of diagnosis field from available historical episodes and to capture this information on an ongoing basis.

DAY OF DIAGNOSIS CREATION

For the majority of cases, day of diagnosis was able to be retrieved from the earliest stored episode that corresponded to the existing case month and year of diagnosis. Where day of diagnosis could not be determined from historical system records, day was estimated as the 15th. This occurred when the original dates of diagnosis came from information that was not stored in the system database (e.g. from doctor's letters or images) and was therefore not able to be easily extracted to populate day of diagnosis historically.

As a result of integrating the registries and populating day of diagnosis, a negligible number of original dates of diagnosis were updated or revised with supplementary information (0.55%).

DATE OF DIAGNOSIS VALIDITY CODE

A new date of diagnosis validity code was created at this time, in order to identify which dates of diagnosis contain *actual dates*, *estimated days and/or months*, and which dates were *unknown*.

1. Actual date

The day of diagnosis was able to be determined from historical system records.

2. Date unknown

Date of diagnosis was unknown. These cases are excluded from cancer incidence reporting.

3. Day estimated

Day of diagnosis was set to 15. Day was estimated for cases in which day of diagnosis could not be easily identified or determined from stored system records.

4. Day and month estimated

Day of diagnosis was set to 15, and month of diagnosis was set to July. Day and month of diagnosis

are estimated for cases where only year of diagnosis was available.

DAY OF DIAGNOSIS ARTEFACT

There is an artefact in the date of diagnosis between 1983 and 1993 indicated by a large increase in *day estimated* cases, mirrored by a decrease in *actual date* cases. This artefact has been linked to pathology notifications that were not stored electronically on the system during a period of operational change. For this reason, the day of diagnosis was unable to be verified for these cases however the month and year of diagnosis were still captured in the system.

Note: any analyses performed using date of diagnosis during this time frame will need to account for this issue.

PROPORTION OF CASES BY DATE OF DIAGNOSIS VALIDITY CODE

The following table provides a breakdown of the proportion of cases by validity code for the diagnosis periods that occur before, during and after the artefact described above.

Code	Description	Proportion of cases		
		1972-1982	1983-1993*	1994-2011
1	Actual date (<i>day, month, year known</i>)	82.2%	74.3%	93.6%
2	Date unknown	0.0%	0.0%	0.0%
3	Day estimated (<i>month, year known</i>)	13.8%	24.2%	6.2%
4	Day and month estimated (<i>year known</i>)	4.0%	1.5%	0.3%

*Includes artefact.

Note 1: This table shows a fall in the proportion of *actual date* cases during the artefact and a notable increase after 1993.

Note 2: Figures are rounded to the nearest decimal place.

USE OF DATA ITEM

Data users should consider this information when using the complete date of diagnosis in any analyses and/or calculations, in particular for survival analyses involving cases with an estimated date of diagnosis.