COMPLETENESS OF CASE ASCERTAINMENT AT THE NATIONAL CANCER REGISTRY, IRELAND

National Cancer Registry Ireland

ABBREVIATIONS

95% CI	95% confidence interval
APC	Annual percentage change
ASR	Age-standardised rate (European standard population)
CIN	Cervical intraepithelial neoplasia
CLL	Chronic lymphocytic leukaemia
CNS	Central nervous system
CSO	Central Statistics Office
ESP	European Standard Population
HD	Hakulinen-Dyba (projection models)
HSE	Health Service Executive
IARC	International Agency for Research on Cancer
ICBP	International Cancer Benchmarking Partnership
ICD	International Statistical Classification of Diseases and Related Health Problems
LOLE	Loss of life expectation
NCCP	National Cancer Control Programme
NCRI	National Cancer Registry, Ireland
NMSC	Non-melanoma skin cancer
NOS	Not otherwise specified
RS	Relative survival
TNM	Tumour, node, metastasis (staging)
WHO	World Health Organisation

Published by:
National Cancer Registry
Building 6800,
Cork Airport Business Park,
Kinsale Road, Cork, Ireland.
T12 CDF7

Telephone: +353 21 4318014
Fax: +353 21 4318016
Email: info@ncri.ie
Website: www.ncri.ie

This report should be cited as:
National Cancer Registry Ireland (2020) Completeness of case ascertainment at the National Cancer Registry, Ireland. NCRI, Cork.

Acknowledgements:
Mr Eamonn O'Leary was responsible for statistical analysis, report compilation, and overall interpretation of results. Dr Paul Walsh was responsible for data collation, as well as report feedback. This work uses data provided by patients and collected by the health service as part of their care and support. The core work of NCRI is funded by the Department of Health.

TABLE OF CONTENTS

REPORT AT A GLANCE

Completeness of case ascertainment at the National Cancer Registry, Ireland

Who are we, and what do we do?

The National Cancer Registry of Ireland (NCRI) works on behalf of the Department of Health and collects information from all hospitals in Ireland on the number of persons diagnosed with cancer and the types of cancer they have.

NCRI also follows up the numbers dying from their cancer or from other causes. All the patient's personal and private details are removed before summaries of this information are made available to the public and health professionals through our annual cancer report and other reports on our website.

How are the numbers reported?

The process of collecting and checking all of this information is done largely by hand and hence is time-consuming, even with increasing use of electronic data sources. Our staff collect cancer diagnosis information and then use an agreed system of coding (The International Classification of Diseases) to group the cancers into different types.

After a process of collating diverse information from Irish hospitals and carefully assigning it to the correct person, it may take up to two years before the annual cancer report can be finalised.

What have we found?

Completeness for 2010 incidence of all invasive cancers excluding NMSC was estimated at 97.2% within five years of incidence. For the four cancers with the highest incidence, colorectal cancer, lung cancer, female breast cancer and prostate cancer, five-year completeness was estimated at 99.0%, 98.7%, 99.3% and 96.2% respectively. This indicates that NCRI has achieved very high levels of case ascertainment, which is essential when providing analysis in respect of the current level of cancer in the country and when providing input in to future planning of cancer services.

Site	ICD10 codes	Flow method completeness % [95% Confidence Interval]		
		one year	three years	five years
All invasive cancers ex NMSC	C00-C43, C45-C96	86.7% [85.6%, 87.7%]	96.1% [95.4%, 96.7%]	97.2% [96.6%, 97.7%]
Colorectal	C18-C21	92.7% [90.1%, 94.9%]	98.9% [97.9%, 99.6%]	99.0% [98.1%, 99.6%]
Lung	C33-C34	93.0% [90.3%, 95.3%]	97.8% [95.8%, 99.2%]	98.7% [96.6%, 99.8%]
Female Breast	C50	93.0% [90.9%, 94.8%]	98.7% [97.7%, 99.4%]	99.3% [98.5%, 99.8%]
Prostate	C61	83.8% [80.8%, 86.6%]	94.1% [91.8%, 96.0%]	96.2% [93.9%, 98.0%]

1. TECHNICAL SUMMARY

Completeness of NCRI data based on incidence in 2010

National Cancer Registry Ireland

- The National Cancer Registry Ireland (NCRI) is a publicly appointed body, established in 1991, to collect and classify information on all cancer cases which occur in Ireland. It has been collecting such data since 1994.
- While the reporting of cancer is not mandatory in Ireland, the NCRI makes considerable efforts to register all cancers diagnosed in Ireland, including active ascertainment and follow-up of cases.
- NCRI receives or compiles data from multiple sources, including from pathology laboratories, screening services, the Hospital In-Patient Enquiry system (HIPE), hospital databases and charts, and radiotherapy centres, as well as from death certificates. NCRI's Data Integration team uses data received from these sources to register new cases, and to add additional data to existing cases.
- NCRI employs Cancer Data Registrars (CDRs), who work in hospitals throughout Ireland, to ensure all relevant data is added to registered cases.

Completeness

- Completeness at the NCRI measures "the extent to which all of the incident cancers occurring in the population are included in the registry database" [1].
- Only with maximum completeness in case-finding procedures will calculated incidence rates and survival statistics be close to their true values [2, 3].
- As missing data is not, by its nature, observable, we can only estimate completeness.
- In this report, a number of different methods are used to assess the completeness of data at NCRI. The methods used fall into two categories:

  (i) Quantitative methods [1] that provide a numerical evaluation of the extent to which all eligible cases have been registered.

  (ii) Semi-quantitative methods [1] that provide an indication of the degree of completeness over time.

Quantitative methods

- The primary quantitative method used in this report was the flow method [4, 5]. This method uses information on survival, on the time between cancer diagnosis and registration, on the likelihood of cancer being listed on a death certificate if a patient dies, and whether a patient is still alive when their cancer is registered, in order to estimate completeness at given times after diagnosis.
- The flow method estimates that completeness of registration of all invasive cancers, excluding NMSC, occurring in Ireland in 2010, at 1 year, 3 years and 5 years after diagnosis is, 86.7%, 96.1% and 97.2% respectively. We estimate that 2.8% of cancers diagnosed in 2010 were unregistered five years later.
- Five year completeness estimates for the four most common cancers were: colorectal cancer (99.0%), lung cancer (98.7%), female breast cancer (99.3%), and prostate cancer (96.2%).
- An additional quantitative measure of completeness used was the DCI/M:I method, or Ajiki method [6]. To use this method we first calculate the mortality incidence ratio by dividing the number of cancer deaths in a year by the

number of cancer cases, and the proportion of Death Certificate Initiated (DCI) cases, i.e. cases first notified to the registry via a death certificate and subsequently followed up by the registry to try and add additional information to the registered case.

- In this study, the Ajiki method used incidence, mortality and DCI data for 2010, and estimated completeness of 92.2% for all invasive cancers combined, excluding NMSC.
- For the four most common cancers, the Ajiki method estimates completeness at 93.2% for colorectal cancer, 95.5% for lung cancer, 98.2% for breast cancer and 92.5% for prostate cancer.
- The Ajiki method has been shown in simulation studies [7] to underestimate the level of completeness in registry data, whereas the same study found that the flow method produced quite accurate results.
- This is consistent with what we have observed in this study.

Semi-quantitative methods

- While quantitative methods provide numerical estimates of completeness, semi-quantitative methods provide indications of problems with the levels of completeness.
- The semi-quantitative [1] estimates used in this study were stability of incidence over time, incidence rates of childhood cancer, the number of sources of notification per case and the proportion of cases that are histologically verified.
- Examining the stability of incidence over time can highlight any unusual trends in incidence which may be the results of changes in completeness. Age standardised rates were calculated between 1994 and 2015. Incidence rates tended to increase between 1994 and 2010, before levelling out and, in the case of males, decreasing. While the changes in rates are not entirely smooth from year to year, there is no changes sufficiently large to indicate problems with completeness in any particular year.
- Incidence rates for all invasive cancers combined, excluding NMSCs, in the childhood age groups (0-4, 5-9 and 10-14) tend to show less variation than rates in adults. Lack of completeness can show up by comparing rates in Irish children with an "expected" range of values. The expected rates are taken from volume XI of Cancer in Five Continents [8]. We found that the incidence rates in Irish children between 2010 and 2014 were within the expected range, broadly supporting the assumption of high levels of completeness in Irish data.
- Using as many sources of notification as possible tends to reduce the likelihood of cancer cases going unregistered, as we are not relying on one or two sources of information to capture all cases. For cancers diagnosed between 2010 and 2014 there were on average 4.0 sources per case which indicates that we can have confidence that we have registered a high percentage of cancers.
- While we expect to have most cases histologically verified, a very high proportion may suggest an over-reliance on the pathology laboratory as a source of information, and cases diagnosed by other means may be missed. By comparing rates of microscopically verified cases in Ireland with other European cancer registries, we can determine if there may be an issue with Irish completeness stemming from an over reliance on pathology data. It was found that while Ireland's proportion of morphologically verified cases was very high, close to 92% on average, this was only slightly above average when compared with the other cancer registries considered.

Conclusions

- Completeness in NCRI incidence data for cases diagnosed in 2010 was at a very high level. The flow method gave an estimate of five year completeness at 97.2% for all cancers, excluding NMSC. The other quantitative method also showed a high levels of completeness, and semi-quantitative methods did not highlight any problems with completeness of NCRI data.
- Completeness for more recent years cannot be estimated at this time, as sufficient years of data collection have not yet elapsed.

2. METHODOLOGY

- Completeness of cancer registry data is defined as "the extent to which all of the incident cancers occurring in the population are included in the registry database" [1].
- In this report, a variety of methods are used to assess completeness. These can be divided into two categories; quantitative methods and semi-quantitative methods.
- Quantitative methods are those which provide a numerical estimate of the extent to which all eligible cases have been registered. In this report we have used two quantitative methods.
- Semi-quantitative methods do not provide specific numerical estimations of completeness, but they do provide an indication of the degree of completeness over time, or in comparison with other registries, and may highlight issues at particular points in time, or in particular cancer sites.

Quantitative methods: Death Certificate Methods

- Access to death certificates is crucial for cancer registries as a means of capturing cases that were not registered while the patient was alive.
- A 'Death Certificate Notification', or DCN, is a case that is first notified to the registry via a death certificate.
- Technically, this differs from a 'Death Certificate Initiated', or DCI, case. DCI cases are a subset of DCN cases, where registry staff have begun attempting to collect further information on the case before any other non-Death Certificate source of notification has been received. If another source of information is received after the Death certificate, but before any trace-back has begun, this should be recorded as a DCN, but not a DCI.
- DCI registrations will also exclude cases that subsequently turn out not to be cancers.
- A 'Death Certificate Only' (DCO) is a case where no further information has been found for the particular cancer case after registry staff have attempted to trace it. DCOs are a subset of DCIs, where attempts to trace-back the case have been unsuccessful (either to confirm or disprove the cancer diagnosis).
- DCI cases represent cases that were not registered while alive, and the proportion of these cases can be used to provide a quantitative estimate of cancer registry completeness.
- Because the distinction between DCN cases and the (likely slightly smaller) DCI subset is difficult to make in retrospective analyses, DCN status has been used to approximate DCI status.
- Two death certificate methods were used in this report, namely
    1. The DCI/M:I, or Ajiki, method
    2. The Flow method.

The DCI/M:I, or Ajiki, method

For this method we define cancers as falling in to one of four categories:

a) Registered while alive and has died

b) Registered while alive and still alive

c) Unregistered while alive, died and registered based on death certificate mentions

d) Unregistered while alive and still alive

- Cancer registrations that are classified as category c are known as Death Certificate Initiated (DCI) cases. After including DCI cases in the registry database,

the undetected cases remaining to be registered, and presumed to be still alive, are those in group d.

- Category d represents the missing cases not registered to date.
- If we can calculate the number of cases in each group then completeness of registration can be calculated as:

$$\frac{a+b+c}{a+b+c+d}$$

- Ajiki et al. [6] provide a formula for estimating completeness using the proportion of DCIs and the M:I ratio, namely:

$$completeness = \frac{1 - DCI \times \left(\frac{1}{M:I}\right)}{1 - DCI}$$

- In order to calculate this estimate, we need the proportion of DCIs and the mortality/incidence (M:I) ratio.
- The M:I ratio is the ratio of cancer deaths to cancer incidence in a particular period, where the number of cancer deaths is obtained independently of follow up of individual cases.
- In this case, mortality data was obtained from the Central Statistics Office (CSO), independent of individual follow-up of registered cases.
- Incidence and mortality data, as well as the proportion of DCI cases recorded, for the year 2010 were used to calculate the estimates of completeness using the Ajiki method.
- This method assumes that the incidence and mortality rates, as well as the proportion of DCIs, remains stable over time.
- It also assumes that the case fatality is the same for registered and unregistered cases. This assumption is unlikely to hold true as unregistered cases are generally in older patients, less likely to be investigated and less intensively treated, and therefore will have higher fatality than the cases detected by the usual case-finding procedures of the registry.

The Flow method

- The flow method [4, 5] estimates completeness of case ascertainment as a function of three time-dependent probabilities. These probabilities can be calculated using routine cancer registry data.
- The unregistered cases previously defined as d, are now split into two different groups:

  M = missing cases. These are unregistered cases that have not died to date.

  L = lost cases. These are unregistered cases that have died, but for whom cancer was not mentioned on the death certificate.
- To estimate these two fractions, three probabilities must be calculated:

  i) $s(t_i)$ = the probability of surviving until time t_i after diagnosis.

  ii) $m(t_i)$ = the probability that cancer is mentioned on the death certificate for a patient who dies between times t_i and t_{i+1}. The estimate of $m(t_i)$ is obtained for the cancer patients who die in the survival analysis.

  iii) $u(t_i)$ = the probability that a patient that has survived to time t_i has not been registered.
- Missing cases at time i are then calculated as:

$$M_i = s(t_i) \times u(t_i)$$

- Lost cases at time i are given by:

$$L_i = [s(t_i) - s(t_{i+1})] \times [1 - m(t_i)] \times [u(t_i)]$$

- Completeness at time T is given by:

$$C_T = 1 - M_T - L_T$$

- Software is provided by the authors [4, 5] for carrying out the calculations and allows for the level of completeness to be estimated at specific times from diagnosis, as well as estimates of missing cases and lost cases.
- Confidence intervals can also be calculated using bootstrapping methods [5].

- For this analysis, completeness was estimated at 1, 3 and 5 years post diagnosis.

Sensitivity Analyses for the Flow Method

- One limitation of our analysis is that cancers listed on death certificates have only been collated by NCRI if the cancer involved was the official cause of death. Other cancers mentioned on a death certificate are not currently collated by NCRI. This will affect the Flow Method estimates of completeness in a number of ways.
- The number of cancer cases where cancer is mentioned on the death certificate is likely to be underestimated.
- It is also likely that there are some cancers that have not been registered, which would have been captured if all mentions of cancer were recorded. These cases would be recorded as DCOs if no further information is found, or DCIs if the trace-back is successful in finding additional sources of notification.
- Some of these potential (missed) DCIs may have been registered from other sources subsequently, but they are not recorded as DCIs. Therefore:

1. The estimate of m(t) is likely to underestimate the true probability that cancer is mentioned on the death certificate.
2. The level of DCOs is likely to be lower than it would be if all cancers listed on Death Certificates were recorded.
3. The proportion of DCIs is likely to be lower than it would be if all cancers listed on Death Certificates were recorded.
4. Finally, the total number of cases in the dataset is likely to be lower than it would be if all cancers listed on Death Certificates were recorded.

- While we do not have information that will allow us to increase the number of cases in the dataset to try and take account of point 4, we can run sensitivity analyses to address points 1, 2 and 3. This will indicate how sensitive our results are to changes in our inputs, which is particularly important here as it is highly likely that the level of DCIs and the probability of the cancer being mentioned on the death certificate are underestimated in our dataset.
- Another limitation with our data is that NCRI records whether a case is a Death Certificate Notification (DCN) rather than DCI, and (as noted earlier) DCN status is used as a proxy for DCI status in analyses here. DCNs are cases that are first notified by a death certificate. If additional sources for these cases are processed before a trace-back process is initiated, then these cases are not considered to be DCIs, however NCRI does not specifically record this information. This may result in the number of DCIs in our data being overstated. We can test how sensitive the results are to this potential issue by reducing the number of DCIs in the data.
- The following sensitivity analyses were performed in order to address points 1-3 and the final sensitivity analysis addresses the potential overestimation in the level of DCI cases:

  i) Set all patients who died (regardless of whether from cancer or otherwise) as having cancer mentioned on their death certs.

  ii) Set half of patients who have died to being DCIs, without changing the portion of DCO patients.

  iii) Set half of patients who have died to being DCIs and change the same patients to being DCOs.

  iv) Set half of the DCIs in the data as not being DCIs.

Semi-quantitative Methods

- While quantitative methods provide numerical estimates of completeness, semi-quantitative methods [1] can also be helpful in providing indications of completeness of registry data. In this report we considered the following semi-quantitative methods:

  1) Stability of incidence rates over time

Examining incidence rates over time can highlight any unusual trends in incidence, or years where incidence rates change markedly, which may be the result of changes in completeness of registration.

2) Incidence rates of childhood cancer

The incidence rates for all cancers combined, excluding NMSC, in the childhood age groups (0-4, 5-9 and 10-14) tend to show less variation than rates in adults. Possible under (lack of completeness) or over (due to duplicates) registration can be investigated by comparing the rates in childhood cancer in Ireland with an "expected" range of values taken from volume XI of Cancer in Five Continents [8].

3) Number of sources of notifications per case

Using as many sources of notification as possible reduces the likelihood of a cancer cases going unregistered, as there are more ways of the cancer being reported to the registry. Therefore the greater the number of notifications, the higher the likely completeness of data at the registry. In some cases NCRI receives multiple reports from the same source (such as from pathology laboratories). As we wish to determine the number of distinct sources per case, we only include one notification from each type of source. The sources considered are as follows: A: Chart; B: Hospital Database/E-Chart; C: Central Sources; D: Death Certificate; E: Death Register; G: GP; H: Hipe; O: Other Outpatient; P: Pathology; R: Radiotherapy; T: Other Inpatient.

4) Histological verification of diagnosis

While the main use of this indicator is as a measure of validity of registry data, it can also be used as an indication of completeness as a very high proportion of cases microscopically verified may suggest an over-reliance on pathology laboratories as a source of information, and cases diagnosed by other means may be missed. Rates of microscopic verification for Ireland are compared with a number of other European registries to see whether the rates in Ireland are out of line with these comparators.

Data

- Cases flagged as multiple primaries of similar morphology of same organ or tissue, based on IARC/IARC rules [9], were excluded. Non-melanoma skin cancers (NMSCs) and non-invasive tumours were also excluded.
- For the semi-quantitative analysis, data from different time periods was considered. Data for the period 1994 to 2015 was included examining incidence rates over. For childhood cancer rates, sources of notification and proportion microscopically verified by site, data for the period 2010-2014 was included. For the comparison of rates of microscopic verification by registry, data for 2008-2012 was examined.
- For the Ajiki method, cases with a year of incidence of 2010 were included. Mortality figures for 2010 were extracted from tables provided by the Central Statistics Office (CSO). NMSCs and non-invasive tumours were excluded from incidence and mortality figures.
- For the flow method, only a patient's first primary invasive cancer (other than NMSC) was included in analysis, meaning that the incidence figures included for the flow method were lower than those for the Ajiki method, as any patient diagnosed with a primary tumour in 2010 who had previously been diagnosed with an invasive cancer (excluding NMSC) was excluded from the flow method analysis.
- Two files were prepared to perform the flow method analysis. Firstly a file was created with all relevant incident cases in 2010. The second file included all cancer patients that died in 2014.
- As well as examining completeness in all invasive cancers, excluding NMSC, 20 specific cancer sites or groups of sites were also examined for estimating completeness.
- The sites included in the analysis are shown in Table 2-1 below:

Table 2-1. Cancer sites and groups of cancer for which completeness estimates are presented in this report

Cancer site	ICD10* codes
All invasive cancers, excl. NMSC	C00-43,C45-96
Head and neck	C01-C14, C30-32
Oesophagus	C15
Stomach	C16
Colorectal	C18-21
Liver, gallbladder and biliary tract	C22-24
Pancreas	C25
Lung	C33-34
Melanoma of skin	C43
Female breast	C50
Cervix uteri	C53
Corpus uteri	C54
Ovary	C56
Prostate	C61
Kidney and renal pelvis	C64-65
Bladder	C67
Brain & central nervous system (CNS)	C70-72
Hodgkin lymphoma	C81
Non-Hodgkin lymphoma	C82-85
Multiple myeloma	C90
Leukaemia	C91-95

*ICD10 but applying ICD-O-3.1 rules regarding tumour behaviour.

3. RESULTS

Completeness: quantitative methods

- The estimates for completeness using the flow method and the Ajiki method are presented in Table 3-1 for cancers incident in 2010.
- For all invasive cancers, excluding NMSCs, the flow method estimated that completeness after one year was at 86.7% (with 95% confidence interval of 85.6% to 87.7%); after three years, 96.1% (95.4%, 96.7%) and after five years, 97.2% (96.6%, 97.7%). This means that 2.8% of all invasive cancers, excluding NMSC, are estimated to be either missing (patient alive but not yet registered) or lost (patient has died without being registered and cancer was not mentioned on their death certificate).
- The Ajiki method estimated completeness of 92.2% for all invasive cancers combined, excluding NMSC.
- For the four sites with the highest incidence of cancer, the flow method five-year completeness estimates, and 95% confidence intervals, were:
  - Colorectal cancer 99.0% (98.1%, 99.6%);
  - Lung cancer 98.7% (96.6%, 99.8%);
  - Female breast cancer 99.3% (98.5%, 99.8%);
  - Prostate cancer 96.2% (93.9%, 98.0%).
- The Ajiki method estimates for these four sites were 93.2%, 95.5%, 98.2% and 92.5% respectively.
- For other sites, five-year completeness estimated using the flow method ranged from 90.0% for leukaemia to 99.8% for cervical cancer. Using the Ajiki method, the completeness estimates ranged from 85.4% for leukaemia to 99.7% for pancreatic cancer.
- Confidence intervals could not be calculated for some sites, where the number of cases, and the levels of DCIs and DCOs in the data were close to zero. This occurred for melanoma, cervix uteri and corpus uteri.

Sensitivity analyses

- Sensitivity analyses were run to test how the results are affected by changes in the portion of patients who have died where cancer was mentioned on their death certificate, the portion of patients who were recorded as DCIs and the portion of patients that were recorded as DCOs. The results of the sensitivity analysis on 5-year completeness (and 95% confidence interval) estimates for all invasive cancers excluding NMSC are as follows:

  i) If all patients who died have cancer mentioned on their death certs: The flow method estimate of completeness at 5 years increases to 98.5% (98.1%, 98.9%), compared with 97.2% estimated for unadjusted data.

  ii) If half of patients who have died are set to DCIs while leaving the portion of DCO patients unchanged: Estimated completeness at 5 years falls slightly to 96.7% (96.1%, 97.3%)

  iii) If half of patients who have died are set to DCOs and the same patients are set to being DCIs: Estimated completeness at 5 years falls slightly to 96.3% (94.3%, 97.9%)

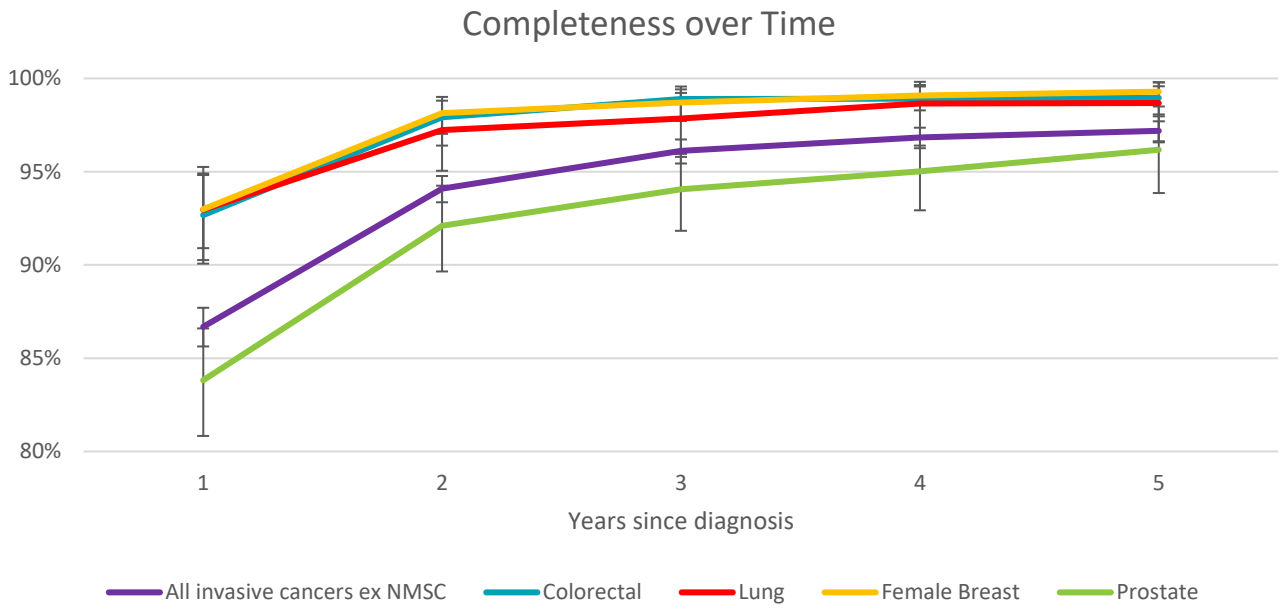
  iv) If half of the DCIs in the data are set as not DCIs: Estimated completeness at five years remains at 97.2% (96.7%, 97.7%).
- These findings suggest that the actual estimate of 97.2% completeness is unlikely to change if the proportion of mentions, DCIs or DCOs changes slightly, and even where there are substantial changes in the inputs, the estimated completeness does not change greatly.

Table 3-1: Estimated completeness of case ascertainment using the flow method and Ajiki method, for all cancers excluding NMSC, and by site or group of sites, for cases incident in 2010.

Site	ICD10 codes	Flow method			Ajiki Method
		Completeness % [95% Confidence Interval]			Completeness %
		one year	three years	five years	
All invasive cancers ex NMSC	C00-C43, C45-C96	86.7% [85.6%, 87.7%]	96.1% [95.4%, 96.7%]	97.2% [96.6%, 97.7%]	92.2%
Head & neck	C01-C14, C30-C32	92.1% [87.0%, 96.0%]	98.8% [97.9%, 99.5%]	98.8% [98.0%, 99.4%]	95.5%
Oesophagus	C15	97.7% [91.9%, 99.9%]	99.0% [96.6%, 100%]	99.1% [97.2%, 100%]	99.6%
Stomach	C16	94.0% [88.3%, 97.9%]	99.0% [96.8%, 99.9%]	99.1% [97.2%, 99.9%]	96.6%
Colorectal	C18-C21	92.7% [90.1%, 94.9%]	98.9% [97.9%, 99.6%]	99.0% [98.1%, 99.6%]	93.2%
Liver, gallbladder and biliary tract	C22-C24	87.5% [81.6%, 92.4%]	96.6% [92.7%, 99.1%]	97.5% [94.5%, 99.4%]	97.2%
Pancreas	C25	93.3% [88.7%, 96.7%]	99.2% [97.1%, 100%]	99.3% [97.7%, 100%]	99.7%
Lung	C33-C34	93.0% [90.3%, 95.3%]	97.8% [95.8%, 99.2%]	98.7% [96.6%, 99.8%]	95.5%
Melanoma	C43	91.9% -	98.1% -	98.8% -	99.4%
Female breast	C50	93.0% [90.9%, 94.8%]	98.7% [97.7%, 99.4%]	99.3% [98.5%, 99.8%]	98.2%
Cervix uteri	C53	88.1% -	99.8% -	99.8% -	98.3%
Corpus uteri	C54	93.5% -	99.2% -	99.2% -	94.4%
Ovary	C56	92.9% [86.2%, 97.5%]	98.3% [95%, 99.8%]	99.0% [97.2%, 99.9%]	96.9%
Prostate	C61	83.8% [80.8%, 86.6%]	94.1% [91.8%, 96.0%]	96.2% [93.9%, 98.0%]	92.5%
Kidney and renal pelvis	C64-C65	80.4% [73.0%, 86.9%]	90.3% [84.5%, 94.8%]	93.2% [85.3%, 98.2%]	88.1%
Bladder	C67	87.7% [83.3%, 91.5%]	96.6% [94.7%, 98.1%]	97.0% [95.3%, 98.4%]	96.2%
Brain & central nervous system (CNS)	C70-C72	86.3% [78.7%, 92.5%]	95.9% [93.0%, 98.0%]	97.2% [94.2%, 99.1%]	92.4%
Hodgkin lymphoma	C81	90.2% -	96.1% -	96.4% -	95.1%
Non-Hodgkin lymphoma	C82-C85	85.5% [80.2%, 90.1%]	97.7% [94.6%, 99.5%]	97.8% [94.9%, 99.5%]	93.4%
Multiple myeloma	C90	69.6% [60.7%, 77.8%]	93.9% [88.0%, 97.8%]	96.4% [92.6%, 98.9%]	96.7%
Leukaemia	C91-C95	58.1% [50.2%, 65.8%]	84.5% [78.2%, 89.9%]	90.0% [85.1%, 94.1%]	85.4%

- Figure 3-1 shows how completeness increases over time for all invasive cancers, excluding NMSC, as well as for the four most common cancers, namely colorectal, lung, female breast and prostate cancer.
- In each case the shape of the curve is similar with the vast majority of cases being captured within one year of diagnosis, and the majority of the outstanding cases being captured in the second year of follow up.
- Gains in completeness decline over time with only a small percentage of outstanding cases being registered in year five.

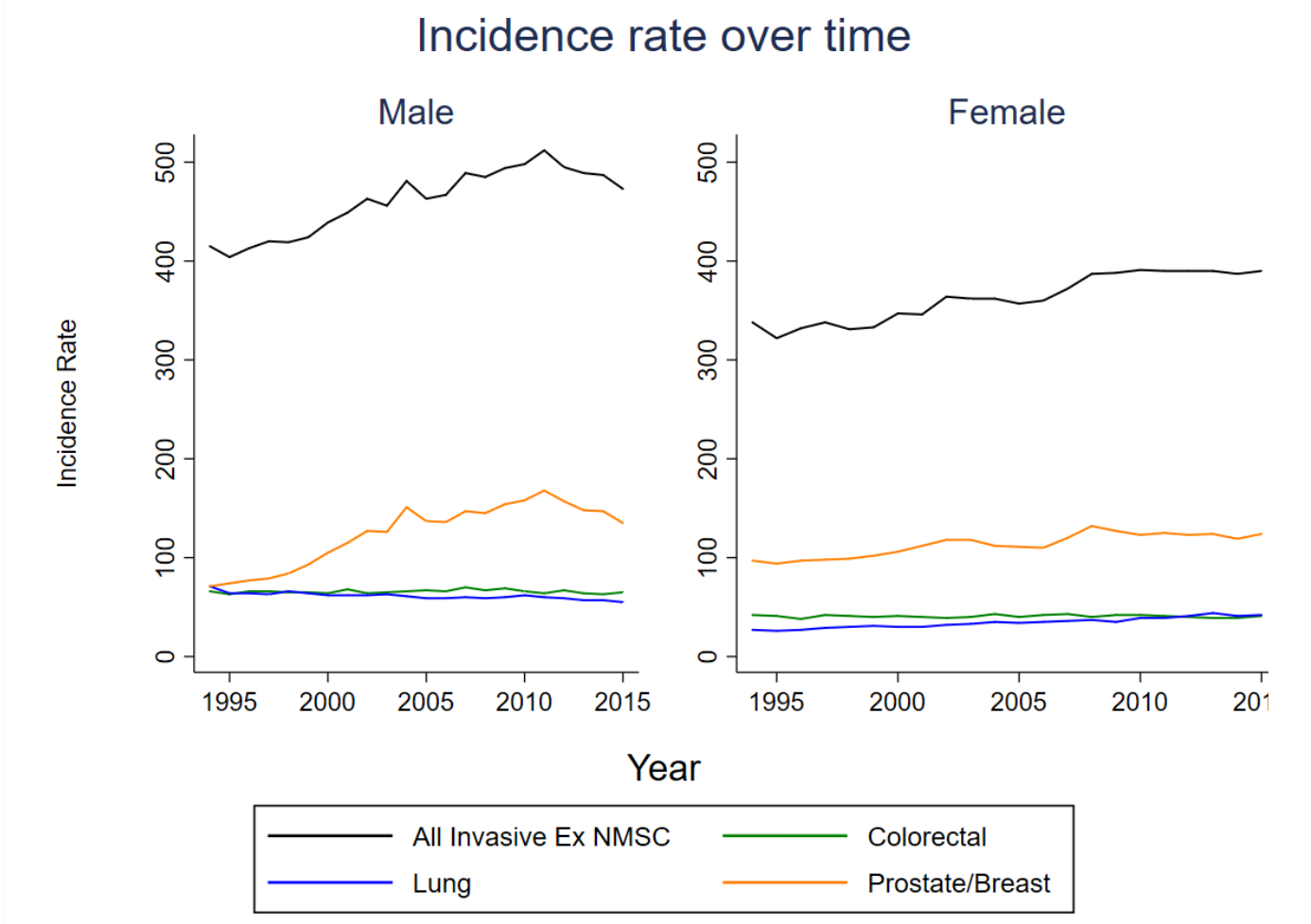
FIGURE 3-1: ESTIMATED COMPLETENESS FOR ALL INVASIVE CANCERS EXCLUDING NMSC, AND THE FOUR MOST COMMON CANCERS



Completeness: semi-quantitative methods

- Figure 3-2 shows the historical incidence rates for males and females for all invasive cancers excluding NMSC, and the top four cancers (colorectal, lung, female breast and prostate cancer). European age standardised incidence rates are presented.
- Incidence rates for all invasive cancers, excluding NMSC, increases steadily over time for males and females until approximately 2010, at which point the rates flatten out for females and decrease for males. These trends are driven to some extent by trends in breast cancer and prostate cancer, which in both cases are influenced by cancer screening.

FIGURE 3-2: INCIDENCE RATES FOR ALL INVASIVE CANCERS EXCLUDING NMSC, AND THE FOUR MOST COMMON CANCERS



- Childhood incidence: Age specific incidence rates for all invasive cancers (excluding NMSC) for children aged 0-4, 5-9 and 10-14 are provided in table 3-2 below. Also included in the table are reference figures which represent an expected range of values for boys and girls. These are taken from Table 5.3 of Cancer in Five Continents volume XI [8].
- As can be seen, the rates in Ireland are within the expected values for boys and girls in each age group.

Table 3-2. Childhood incidence rates per 100,000 for all cancers, excluding NMSC, by sex, Ireland 2010-2014

Age	girls	reference	boys	reference
0-4	22.1	12.1, 23.7	22.0	12.6, 26.4
5-9	10.2	7.0, 13.0	11.0	8.9, 17.9
10-14	12.3	8.2, 16.0	12.1	9.0, 17.2

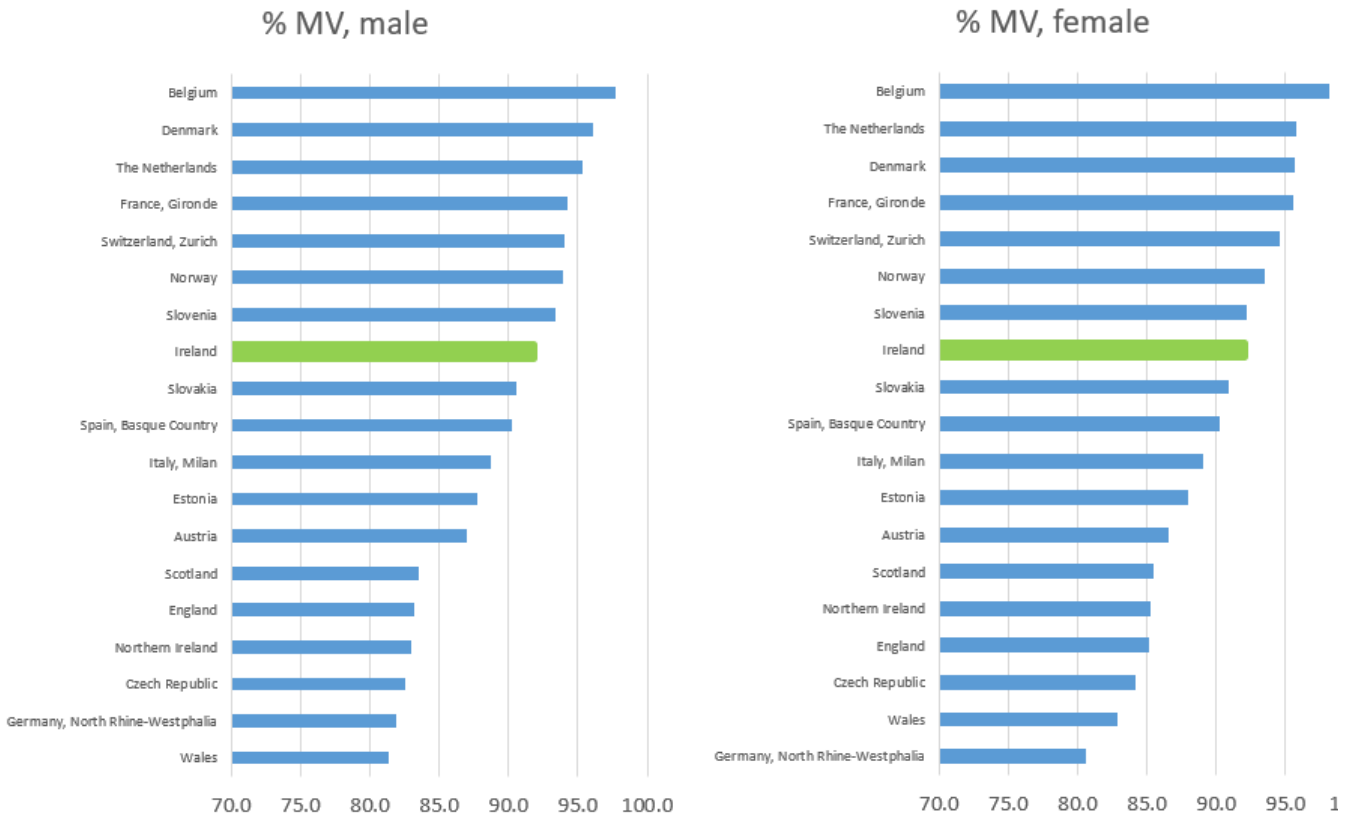
- Table 3-3 shows the average number of distinct sources per case and the percentage of cases where the morphology has been microscopically verified.
- For all invasive cancers, excluding NMSC, there were on average four distinct sources per case. Oesophageal cancer had the highest number of sources, with an average of five per case, and melanoma had the lowest number with just over three per case.
- For all cancers combined, excluding NMSC, the percentage of cases microscopically verified was 92%. Of the sites examined, the percentage of cases microscopically verified was lowest for liver, gallbladder and biliary tract (66.5%), pancreatic cancer (72.3%), brain & CNS (77.5%) and kidney and renal pelvis cancer (78.9%). For many cancers the level was close to 100%, including melanoma (99.9%), Hodgkin lymphoma (99.7%), cervical cancer (99.4%) and female breast cancer (99.3%).

Table 3-3. Average annual incidence, number of distinct sources of notification per case, and the percentage of cases microscopically verified, by site, for 2010-2014

Site	ICD10 codes	Annual Average Incidence	Number of sources per case	MV %*
All invasive cancers ex NMSC	C00-C43, C45-C96	20775	4.0	92.0
Head & neck	C01-C14, C30-C32	601	4.5	97.8
Oesophagus	C15	382	5.0	96.2
Stomach	C16	560	4.6	96.3
Colorectal	C18-C21	2520	4.0	94.9
Liver, gallbladder and biliary tract	C22-C24	417	4.0	66.5
Pancreas	C25	513	4.3	72.3
Lung	C33-C34	2368	4.4	84.0
Melanoma	C43	945	3.2	99.9
Female breast	C50	2887	4.5	99.3
Cervix uteri	C53	310	4.7	99.4
Corpus uteri	C54	442	4.2	98.1
Ovary	C56	377	4.3	90.9
Prostate	C61	3433	3.3	94.7
Kidney and renal pelvis	C64-C65	602	3.3	78.9
Bladder	C67	405	3.8	91.1
Brain & central nervous system (CNS)	C70-C72	374	4.5	77.5
Hodgkin lymphoma	C81	142	3.7	99.7
Non-Hodgkin lymphoma	C82-C85	763	4.0	98.4
Multiple myeloma	C90	268	3.9	92.4
Leukaemia	C91-C95	537	3.6	96.3

- *Figure 3-3* below shows the percentage of cases microscopically verified, for all cancers combined, for a range of European countries or regions. As can be seen, Ireland ranked 8th out of 19 registries when it came to percentage microscopically verified. This shows that Ireland's figures are in line with what we would expect to see when compared with other European cancer registries, and therefore this measure does not indicate any issues in respect of completeness.

FIGURE 3-3: PERCENTAGE OF CASES MICROSCOPICALLY VERIFIED BY COUNTRY, OR REGION, FOR ALL CANCERS COMBINED, EXCLUDING NMSC (CANCERS IN FIVE CONTINENTS, VOLUME XI)



4. DISCUSSION

- In this report we considered a number of different measures of case completeness, both quantitative methods that provided numerical estimates of completeness, and semi-quantitative measures that gave indications of completeness, for cancer registration in Ireland.
- The flow method gave a high estimates of completeness of NCRI data, estimating that for cases diagnosed in 2010, incidence data was 97.2% complete after five years for all invasive cancers combined, excluding NMSC. This closely matches the estimate of 97.0% in the previously reported by NCRI [11, 12] for cases diagnosed in 2005. Estimated five-year completeness for colorectal cancer, female breast cancer and prostate cancer were slightly higher than the estimates in the previous report, increasing from 97.4% to 99.0%, 98.0% to 99.3%, and 95.8% to 96.2%, respectively. Estimated five-year completeness of lung cancer was unchanged at 98.7%.
- Completeness of other sites was not reported previously. For most sites the estimated completeness was very high, at 99% or over for cancers of the oesophagus (99.1%), stomach (99.1%), pancreas (99.3%), cervix (99.8%), corpus uteri (99.2%) and ovary (99.0%).
- Leukaemia had the lowest estimated five year completeness at 90.0%, with a 95% confidence interval of 85.1% to 94.1%.
- While most European cancer registries assess completeness, registries often focus on semi-quantitative methods rather than the more complex quantitative methods, such as the flow method. In a survey [13] it was found that of the registries asked, 79% used historical comparison (examining incidence rates over time), whereas only 18% used the flow method to assess completeness.
- The Northern Ireland Cancer Registry (NICR) reported five year completeness at 96% for incidence in 2010-2012 [14], a result similar to that found in NCRI data. In Austria, five year completeness was estimated at 94.2% for cases diagnosed in 2005 [15].
- For cancer cases diagnosed between 2006 and 2011, three year completeness in Swiss registries was estimated at 92.1% [16]. This compares to 96.1% at NCRI.
- From these published results, we can see that completeness at NCRI compares well with other European cancer registries.
- The Ajiki method estimated completeness of 92.2% for all invasive cancers, excluding NMSC. While the Ajiki method provides a high estimate of completeness, it is substantially lower than the estimate produced using the flow method. This is as expected, as a previous simulation study has shown that the Ajiki method was biased and significantly underestimates the true level of completeness [7]. The same simulation study found that where the model assumptions are met, the flow method accurately estimates completeness.
- The simulation study also found that the flow method tends to underestimate completeness in cancers with better survival, such as breast cancer. Completeness was estimated at 99.0% for female breast cancer in this report, so this does not appear to have been an issue in our data.
- In sensitivity analysis, where the proportion of cancers mentioned on death certificates, the proportion of DCIs and the proportion of DCOs in the data were changed, the results from the flow method did not change substantially.
- We can therefore be confident that the high estimates of completeness found in this study, based on the flow method, reflects the fact that NCRI captures nearly all cancers diagnosed in Ireland.
- Ensuring high levels of completeness is extremely important. NCRI's functions, as set out in legislation, are to capture information on all

tumours diagnosed in Ireland. High levels of completeness ensure that estimates of age standardised incidence rates and survival statistics are as accurate as possible [2, 3].

- Data from the NCRI is used to inform public policy, and underestimates of cancer incidence, or biased estimates of survival, may lead to insufficient resources provided for cancer services.
- Accurate incidence data is essential, not only for current decision making, but also for planning into the future. A recent NCRI report on cancer incidence projections [17] found that, if current rates were applied to projected populations, cancer cases would double between 2015 and 2045. It also found that if recent trends in incidence were taken into account, there may be a more modest 50% increase in cases in the same period.
- If the recent levelling off in incidence rates in females, and decrease in males, was due to incomplete incidence data, rather than a true reflection of a decline in risk, the projections in that report could underestimate the likely future incidence of cancer and contribute to inappropriate planning decisions for future services. The high level of registration completeness estimated in the current report provides support for the reliability of NCRI's assessment of recent trends.
- One limitation of the NCRI data when applying the flow method is that NCRI data slightly under represents the proportion of cancers mentioned on death certificates, as only cancers that are officially certified as the main cause of death are currently captured. The flow method requires all mentions of cancer to be recorded. However, the sensitivity analysis noted above showed that a large increase in the proportion of mentions, as well as in the number of DCIs (where a cancer is first registered via a notification from a death certificate, and a trace-back procedure is initiated to attempt to find further information on the cancer) did not lead to a substantial change in the estimate of completeness.
- If the mentioned cancers would have led to a new registration which would have been recorded as a DCI, but that would have been subsequently recorded from information received from other sources, then the overall completeness will not be affected, and the estimates of completeness will also not be affected greatly.
- However, if the cancers mentioned on the death certificates are not subsequently registered from other sources, these cancers will be lost to registration, whereas the flow method assumes that they will be recorded. This may lead to the flow method overestimating completeness.
- Another potential limitation is that NCRI records cases as Death Certificate Notification (DCN) but does not explicitly flag Death Certificate Initiated (DCI). NCRI records where a case is first notified via a death certificate, but not whether another source is received before the trace-back is initiated.
- For this analysis it was assumed that the number of DCNs is the same as the number of DCIs. This is likely to be a reasonable assumption, and unlikely to influence the results as the sensitivity analysis showed that reducing the proportion of DCI substantially had virtually no effect on the completeness estimate.
- Semi-quantitative measures of completeness are intended to highlight situations where completeness of data may change over time or differ in comparison to international experience. In the measures examined in this report there were no indications of issues with the NCRI data. Incidence over time changed in a way that was consistent with what we might expect, with no major changes from year to year. Childhood incidence rates were within reference bands based on Cancer in Five Continents, Volume XI [8]. The level of microscopically verified cases of cancer in Ireland was comparable with other European registries, and the level of microscopic verification in Ireland was very high for the majority of cancer sites, but not so high as to suggest an over-reliance on a single source of notification which may lead to some non-microscopically verified cancers being missed. Compared to the previous completeness report [12], NCRIs MV% has increased, from 85.4% in the period 2003-2007 to 92.1% for cases diagnosed between 2010 and 2014.

5. CONCLUSIONS

- NCRI has very high levels of completeness in its registered data, with completeness of cases in 2010 estimates at 97.2% within five years of diagnosis for all invasive cancers combined, excluding NMSC.
- Completeness has shown little change, or increased slightly, between 2005 and 2010, for all cancers excluding NMSC, with small increases for three of the four most common cancers, namely colorectal, female breast, and prostate cancer.

6. RECOMMENDATIONS

- NCRI should examine the feasibility of recording and attempt to trace-back all mentions of cancer on death certificates (not just those certified as official cause of death) as part of routine case-ascertainment. This would likely lead to a small increase in registrations, and also allow a more accurate estimation of completeness into the future.
- NCRI does not currently differentiate between DCN and DCI cases, and currently only provides information on DCN cases. Typically there is a delay in receiving and processing death certificate data which in practice means that the difference between the two measures is likely to be limited, and the effect on completeness is likely to be minimal. However, it would be preferable to record explicitly whether a DCN case is also a DCI case. This will increase confidence in future estimates of completeness.

7. REFERENCES

1. Parkin DM, Bray F. Evaluation of the data quality in the cancer registry: Principles and methods. Part II. Completeness. 2009, *Eur J Cancer*, Vol 45, pp. 756-764.
2. Brenner H, Hakulinen T. Population-based monitoring of cancer patient survival in situations with imperfect completeness of cancer registration. 2005, Br J Cancer, Vol. 92, pp. 576-579.
3. Møller H et al. Completeness of case ascertainment and survival time error in English cancer registries: impact on 1-year survival estimates. 2011, Br J Cancer, Vol. 92, pp. 576-579
4. Bullard J, Coleman MP, Robinson D, Lutz J-M, Bell J, Peto J. Completeness of cancer registration: a new method for routine use. 2000, *Br J Cancer*, Vol. 82, pp. 1111-1116.
5. Silcocks PB, Robinson D. Completeness of ascertainment by cancer registries: putting bounds on the number of missing cases. 2004, J Public Health; Vol. 26 pp. 161–167.
6. Ajiki W, Tsukuma H, Oshima A. Index for evaluating completeness of registration in population-based cancer registries and estimation of registration rate at the Osaka Cancer Registry between 1966 and 1992 using this index. *Nippon Koshu Eisei Zasshi*, 1988, Vol. 45, pp. 1011-1017. (in Japanese, English abstract).
7. Silcocks PB, Robinson D. Simulation modelling to validate the flow method for estimating for estimating completeness of case ascertainment by cancer registries. 2007, *J Public Health* Vol. 29, pp. 455-62.
8. Bray F, Colombet M, Mery L, Piñeros M, Znaor A, Zanetti R and Ferlay J, editors (2017) *Cancer Incidence in Five Continents, Vol. XI* (electronic version). Lyon: International Agency for Research on Cancer. Available from: https://ci5.iarc.fr, accessed 06/11/2020.
9. International classification of diseases for oncology (ICD-O) – 3rd edition, 1st revision. World Health Organization. 2013.
10. Schmidtmann I, Blettner M. How do cancer registries in Europe estimate completeness of registration? 2009, Methods Inf Med, Vol. 48, pp. 267-271.
11. O'Brien K, Comber H, Sharp L. Completeness of case ascertainment at the Irish National Cancer Registry. 2014, *Ir J Med Sci*; Vol. 183, pp. 219–224.
12. Data Quality and Completeness at the Irish National Cancer Registry. 2012.
13. Zanetti R, Schmidtmann I, Sacchetto L et al. Completeness and timeliness: Cancer registries could/should improve their performance. 2015, *Eur J Cancer, Vol.* 51 pp. 1091-1098.
14. Donnelly C, Cairnduff V, Chen JJ, Kearney T, Fitzpatrick D, Fox C, Gavin A. The completeness and timeliness of cancer registration and the implications for measuring cancer burden. 2017, *Cancer Epidemiology,* Vol. 49, pp. 101-107.
15. Hackl M, Waldhoer T. Estimation of completeness of case ascertainment of Austrian cancer incidence data using the flow method. 2013, *Eur J Public Health*, Vol. 23, Pages 889-893.
16. M Lorez et al. Evaluation of completeness of case ascertainment in Swiss cancer registration. 2017, *Eur J Cancer Prevention,* Vol. 26, pp. 139-146.