European Network of Cancer Registries

# The completeness of Cancer Registry data

Carlotta Buzzoni[1,2]
c.buzzoni@ispo.toscana.it

**1 -** Clinical and Descriptive Epidemiology Unit, ISPO-Cancer Research and Prevention Institute, Florence, Italy
**2 -** Italian Association of Cancer Registries (AIRTUM)

# Definition…

## …of completeness

"the quality of being whole or perfect and having nothing missing"

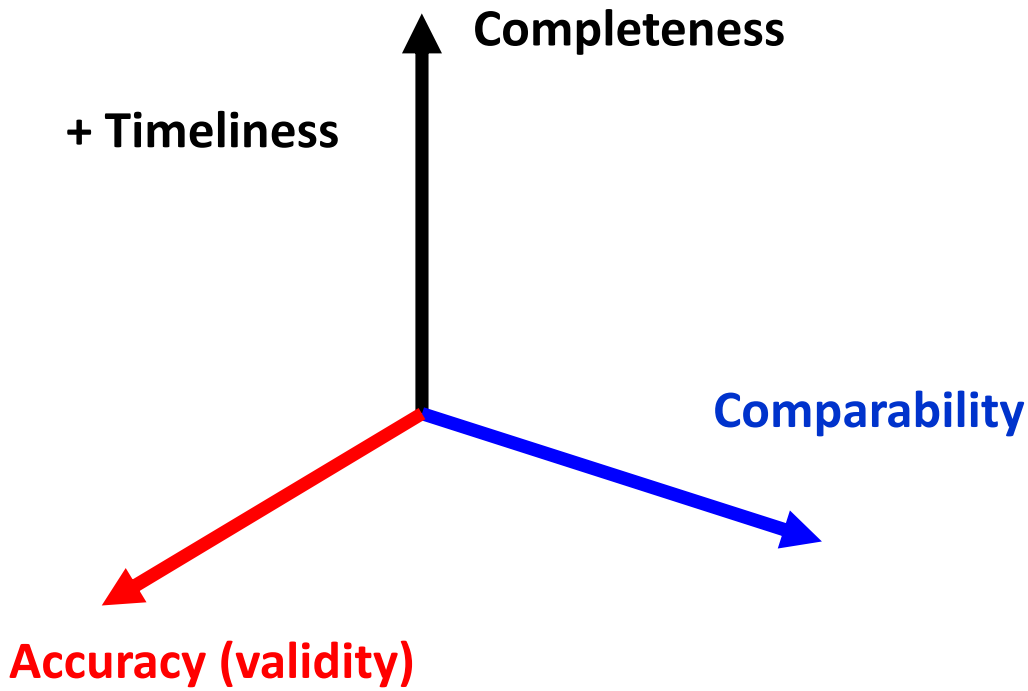
Cambridge dictionary

## … of completeness in cancer registration

"the extent to which all of the incident cancers occurring in the population are included in the registry database"

Parkin and Bray, 2009

# One of the three dimensions




Parkin and Bray, 2009

European Network of Cancer Registries

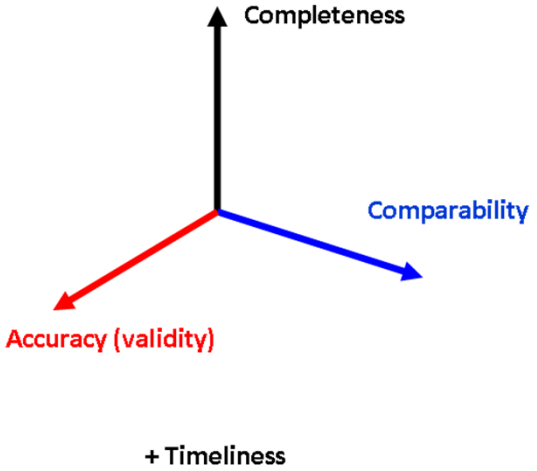
# Completeness and accuracy:
## independent dimensions?

Example:

Completeness of case ascertainment and survival time error in English cancer registries: impact on 1-year survival estimates. Moller et al, 2011

**Table 2** Difference in survival time from date of diagnosis in cancer registration and from earliest episode

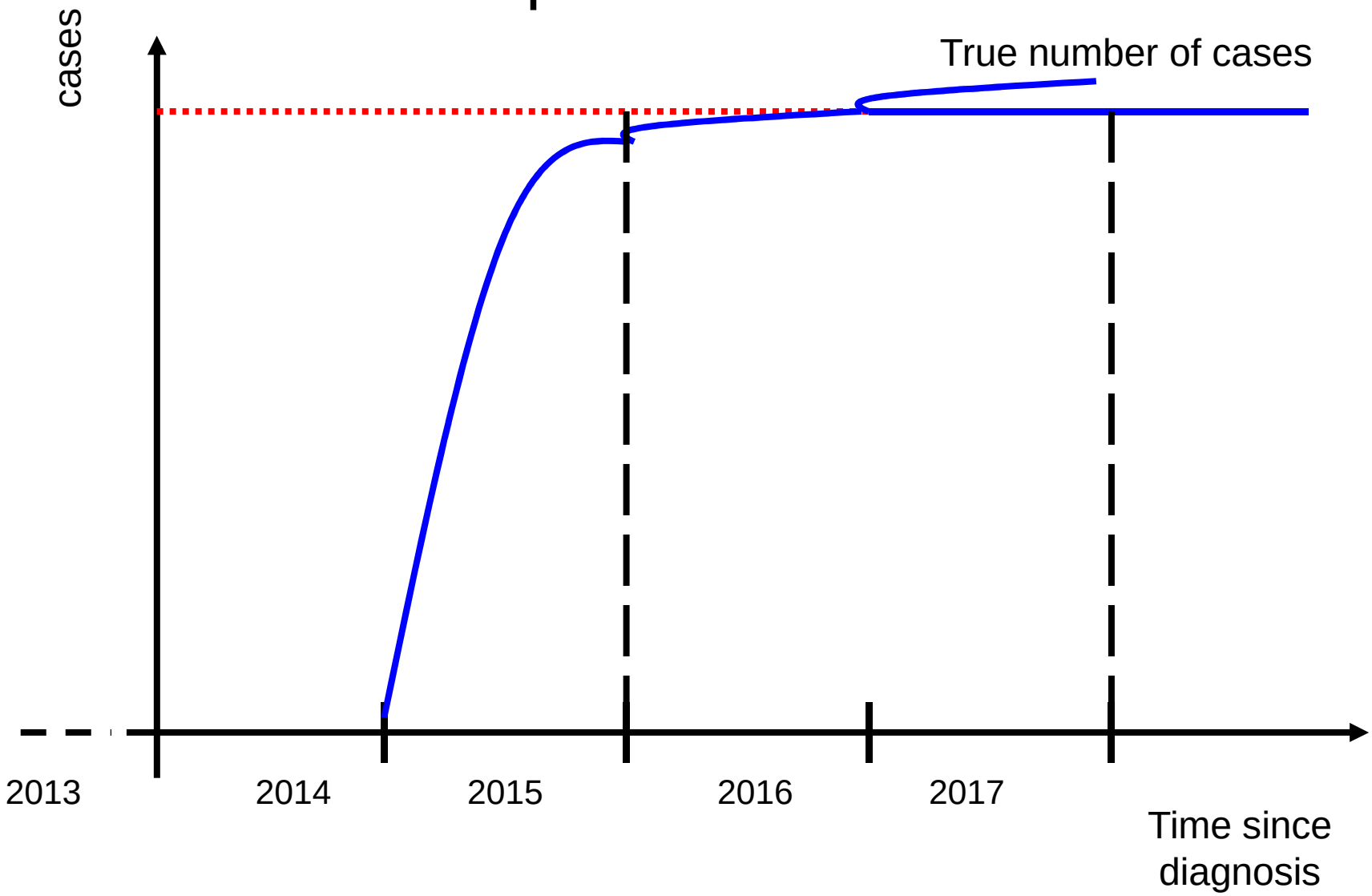
| | Proportion that changed (%) | | |
|---|---|---|---|
| | No change | More than 1 month | More than 1 year |
| *Colorectal cancer registry* | | | |
| Eastern Cancer Registration & Information Centre (ECRIC) | 56.4 | 4.4 | 0.9 |
| North West Cancer Intelligence Service | 55.1 | 11.1 | 0.9 |
| Northern & Yorkshire Cancer Registry & Information Service | 65.3 | 1.7 | 0.2 |
| Oxford Cancer Intelligence Unit | 62.9 | 3.1 | 0.3 |
| South West Cancer Intelligence Service | 70.3 | 4.3 | 1.0 |
| Thames Cancer Registry | 62.5 | 5.7 | 1.4 |
| Trent cancer registry | 65.9 | 4.0 | 0.9 |
| West Midlands Cancer Intelligence Unit | 57.3 | 4.9 | 0.6 |
| Total | 62.2 | 5.1 | 0.8 |
| *Lung cancer registry* | | | |
| Eastern Cancer Registration & Information Centre (ECRIC) | 66.2 | 3.2 | 0.3 |
| North West Cancer Intelligence Service | 65.5 | 11.9 | 0.7 |
| Northern & Yorkshire Cancer Registry & Information Service | 72.6 | 1.5 | 0.1 |
| Oxford Cancer Intelligence Unit | 72.3 | 2.3 | 0.2 |
| South West Cancer Intelligence Service | 75.2 | 3.6 | 0.5 |
| Thames Cancer Registry | 70.6 | 4.1 | 0.5 |
| Trent Cancer Registry | 74.3 | 3.0 | 0.5 |
| West Midlands Cancer Intelligence Unit | 64.0 | 4.8 | 0.3 |
| Total | 70.0 | 4.7 | 0.4 |
| *Breast cancer registry* | | | |
| Eastern Cancer Registration & Information Centre (ECRIC) | 78.7 | 4.5 | 2.1 |
| North West Cancer Intelligence Service | 67.3 | 9.4 | 3.3 |
| Northern & Yorkshire Cancer Registry & Information Service | 80.7 | 2.4 | 0.8 |
| Oxford Cancer Intelligence Unit | 86.5 | 1.9 | 1.0 |
| South West Cancer Intelligence Service | 83.6 | 6.0 | 2.7 |
| Thames Cancer Registry | 77.3 | 8.1 | 3.8 |
| Trent Cancer Registry | 78.6 | 8.0 | 4.4 |
| West Midlands Cancer Intelligence Unit | 78.3 | 4.1 | 1.4 |
| Total | 78.2 | 6.2 | 2.7 |



European Network of Cancer Registries

# Timeliness and completeness:

## independent dimensions?



# Completeness in cancer registration

- Overall
- By subgroup (i.e: cancer site, geographic area, age group)

- Homogeneous completeness

**Missing Completely At Random (MCAR)**

- Heterogeneous completeness

**Missing At Random (MAR)**
**Missing Not At Random (MNAR)**

Rubin, 1976

# Evaluation of completeness

- qualitative (or semi-quantitative) methods
  - role of the experts
  - automated evaluation (i.e: software)

- quantitative methods
  - analytical indicators derived from auxiliary variables

Parkin and Bray, 2009

# Qualitative methods

- Historic data methods:
  - Stability of incidence rates over time
  - Comparison of incidence rates in different populations
- Shape of age-specific curves
  - Incidence rates of childhood cancers
- Mortality/incidence ratios
- Number of sources/notifications per case
- Histological verification of diagnosis

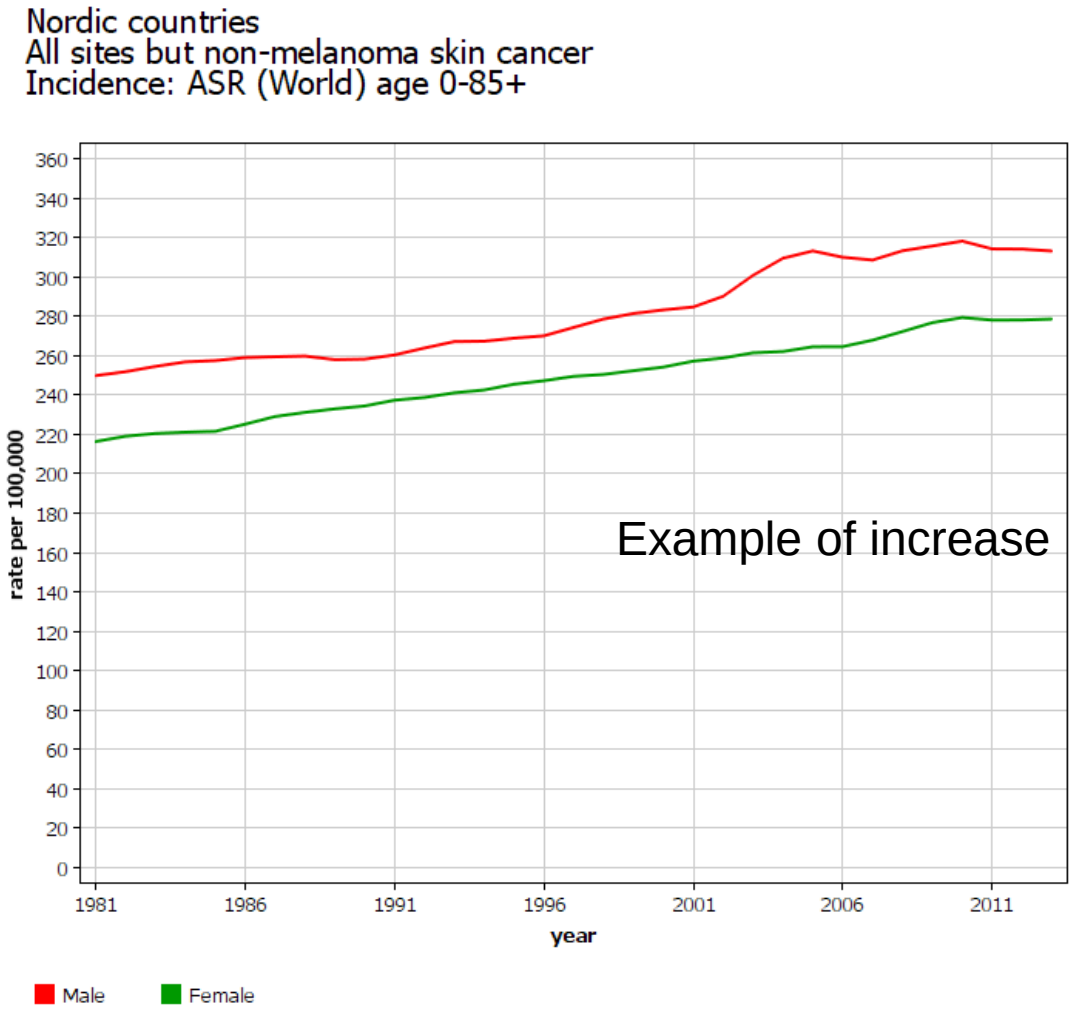
Parkin and Bray, 2009

# Stability of incidence rates over time

Variation in incidence trend may be due to:

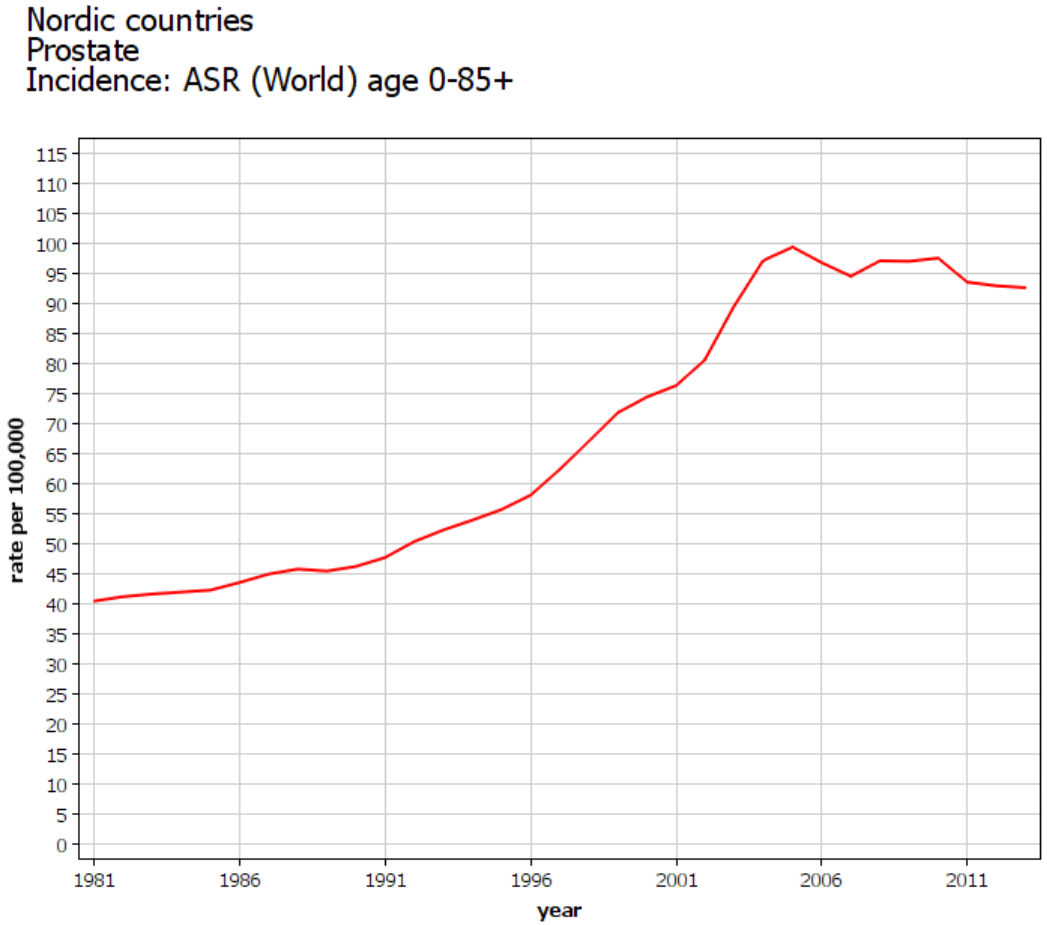
- an increased/decreased exposure to carcinogens (i.e: changes in prevalence of smoking during past years) slow variation
- organized screening programmes/early diagnosis activities rapid variation
- changes in classification systems

# Stability of incidence rates over time




NORDCAN © Association of the Nordic Cancer Registries (16.9.2016)

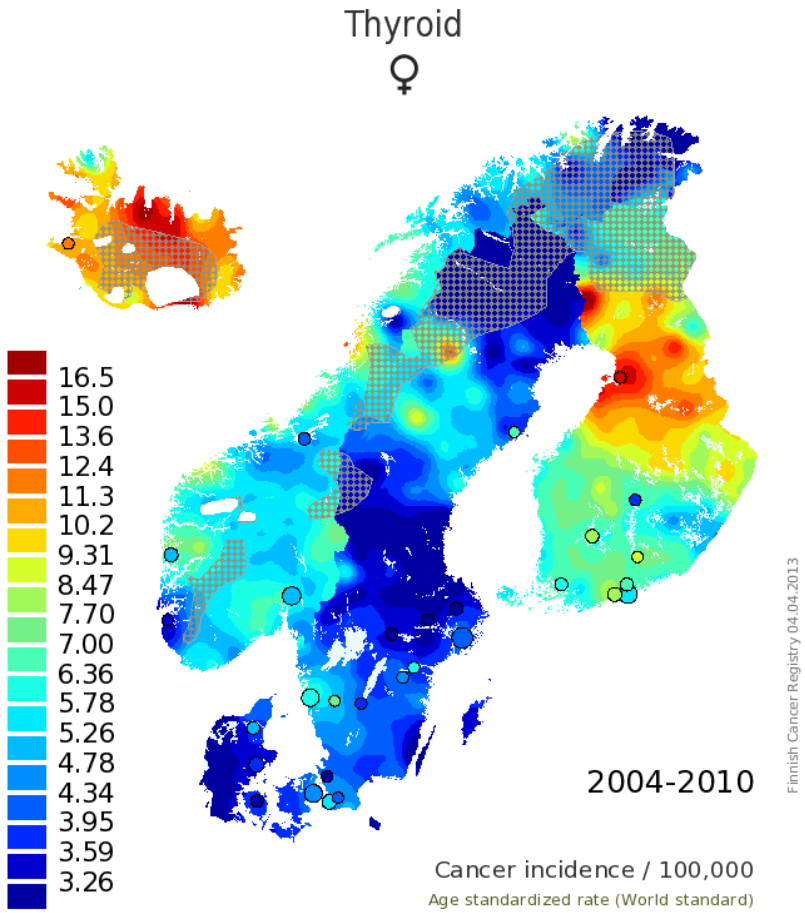
# Stability of incidence rates over time



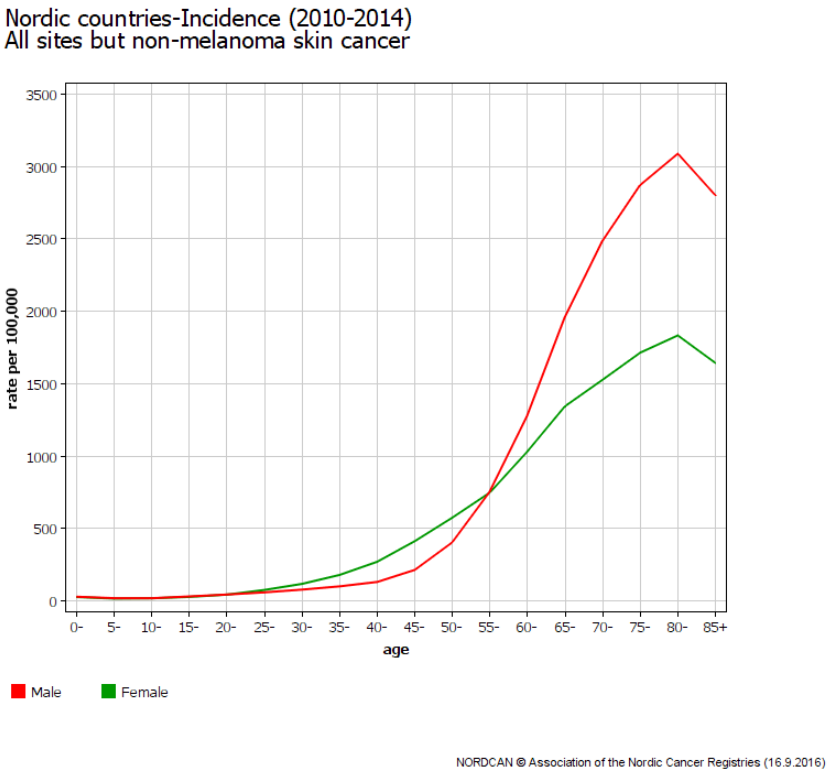

NORDCAN © Association of the Nordic Cancer Registries (16.9.2016)

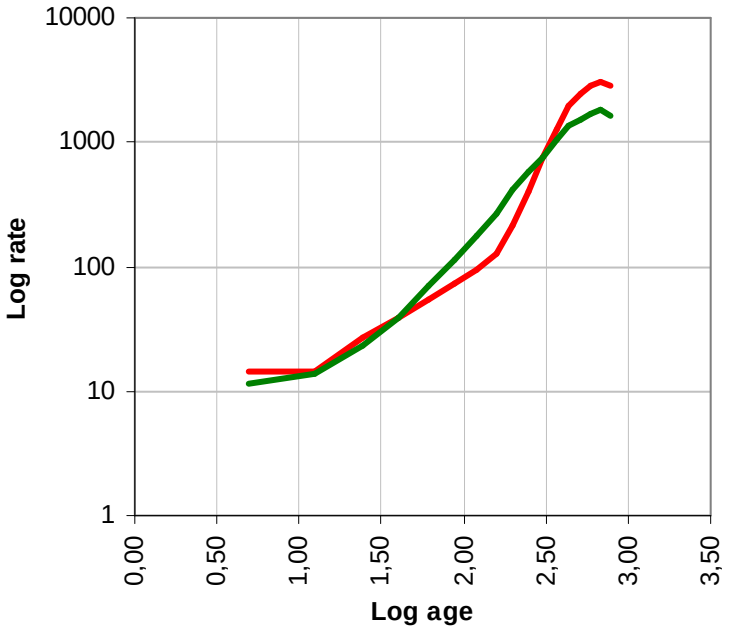
European Network of Cancer Registries

# Comparison of incidence rates in different populations



# Shape of age-specific curves



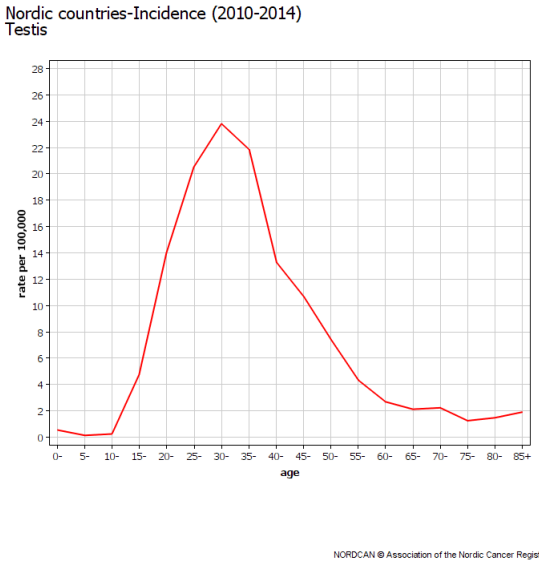



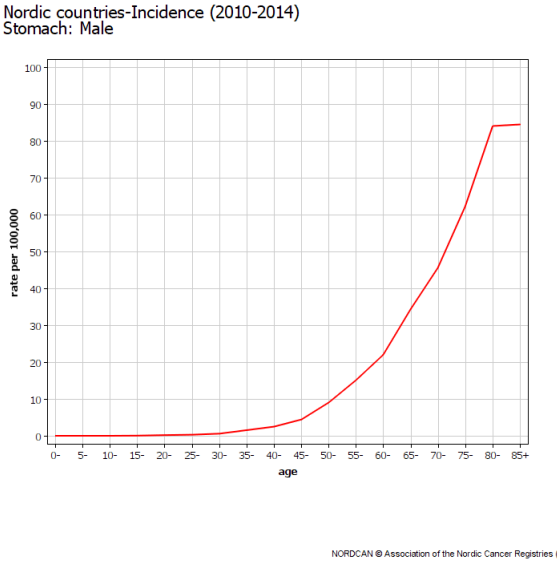

Linear on Log-scale

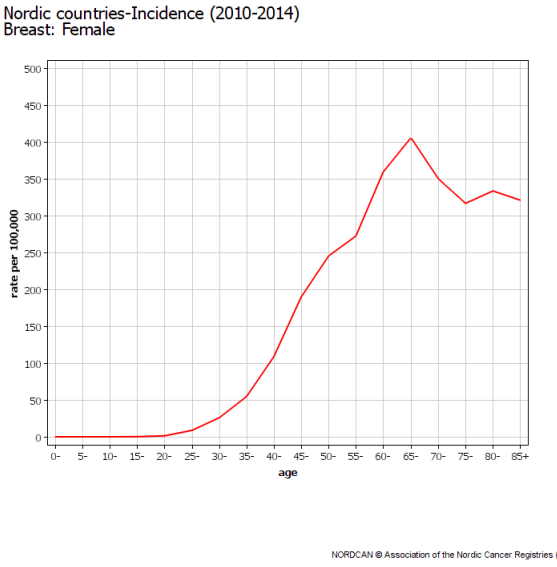
Armitage and Doll 1954

European Network of Cancer Registries

# Shape of age-specific curves




Different by cancer site

# Mortality/incidence ratios

- Constant in short period
- Measure of survival if short term time trend are stable

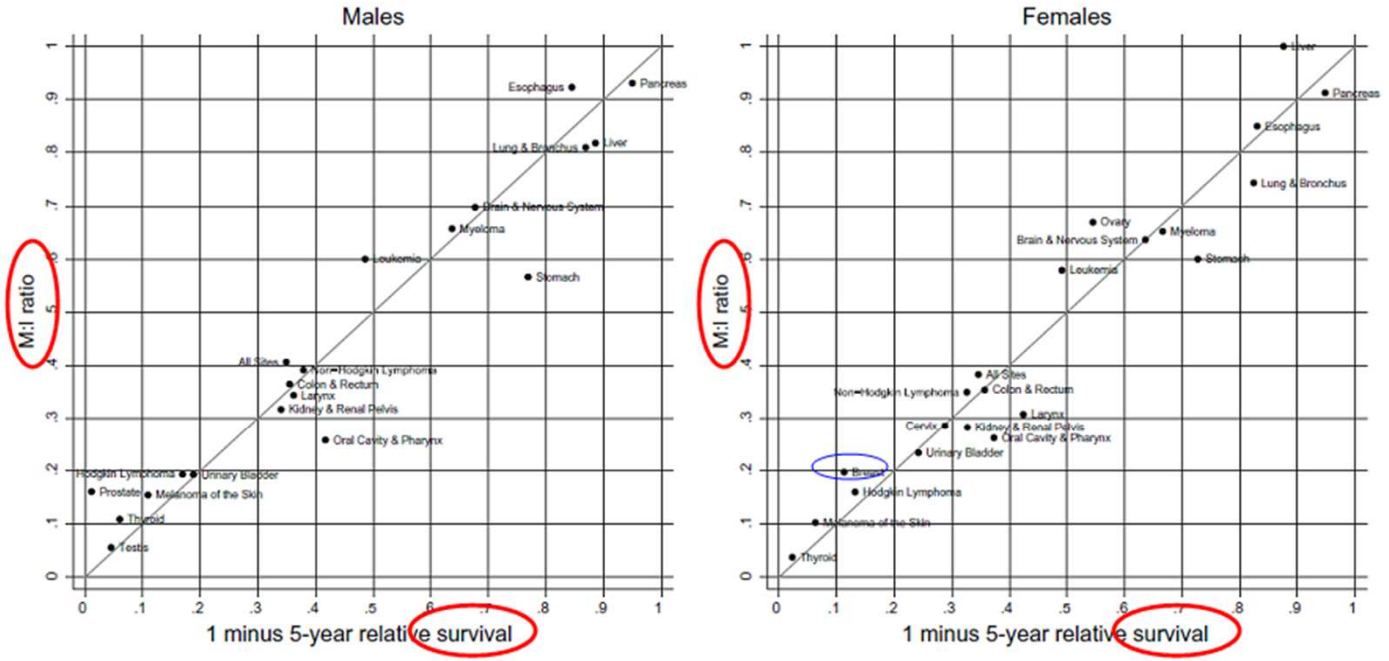



**Fig. 1 – Mortality:incidence ratios (2001–2005) versus 1 minus 5-years relative survival (1996–2004). Statistics based on data from the SEER 9 registries (*Source*: SEER Cancer Statistics Review, 1975–2005[35]).**

Ex: breast **M/I** 0.2 **S(t)** 0.90

# Number of sources/notifications per case

Many sources as possible

-> minimizing of the possibility of cancer diagnoses going unreported

-> increasing the completeness

- average number of sources per case,
- and the average number of notifications per case

Importance of record linkage

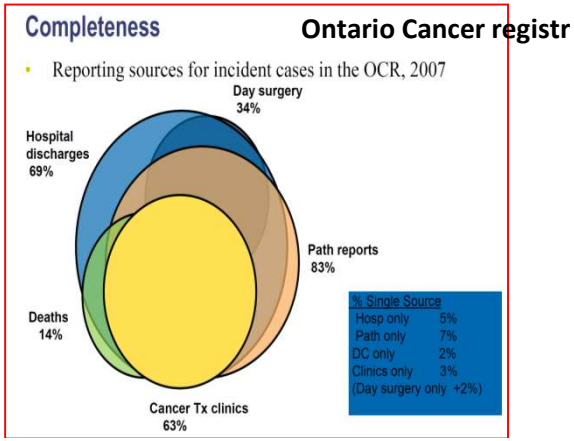
# Histological verification of diagnosis

- measure of validity/accuracy, and methods for comparing of observed and 'expected' values of MV%.

  i.e: Breast cancer vs ductal carcinoma of breast

- High proportion of cases diagnosed by histology or cytology/haematology –suggests over-reliance on the pathology laboratory as source of information, and failure to find cases diagnosed by other means.

  i.e: 100% MV of lung cancer: what does it mean?

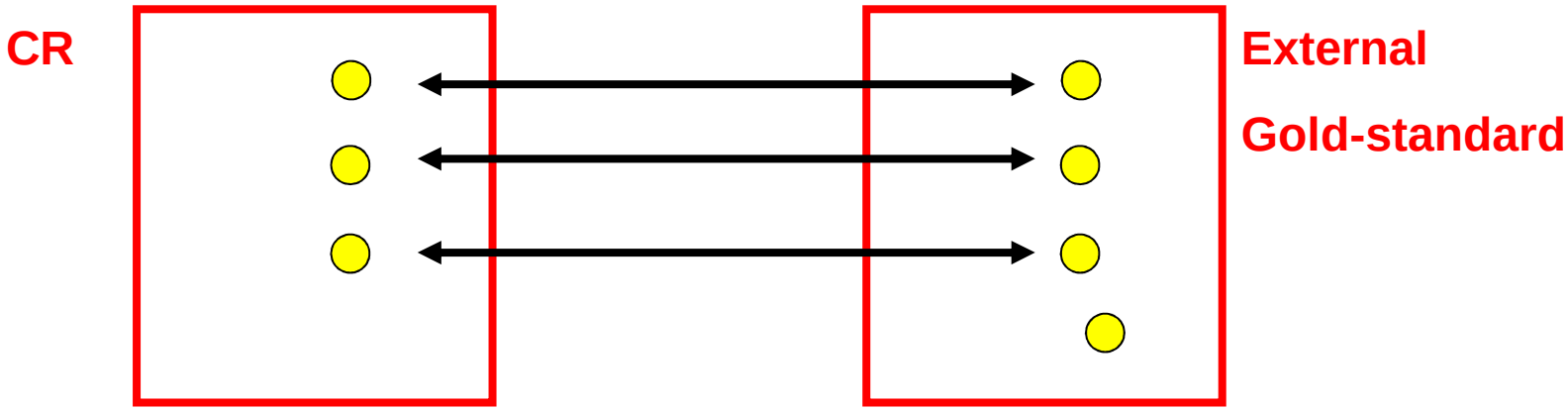


# Quantitative methods

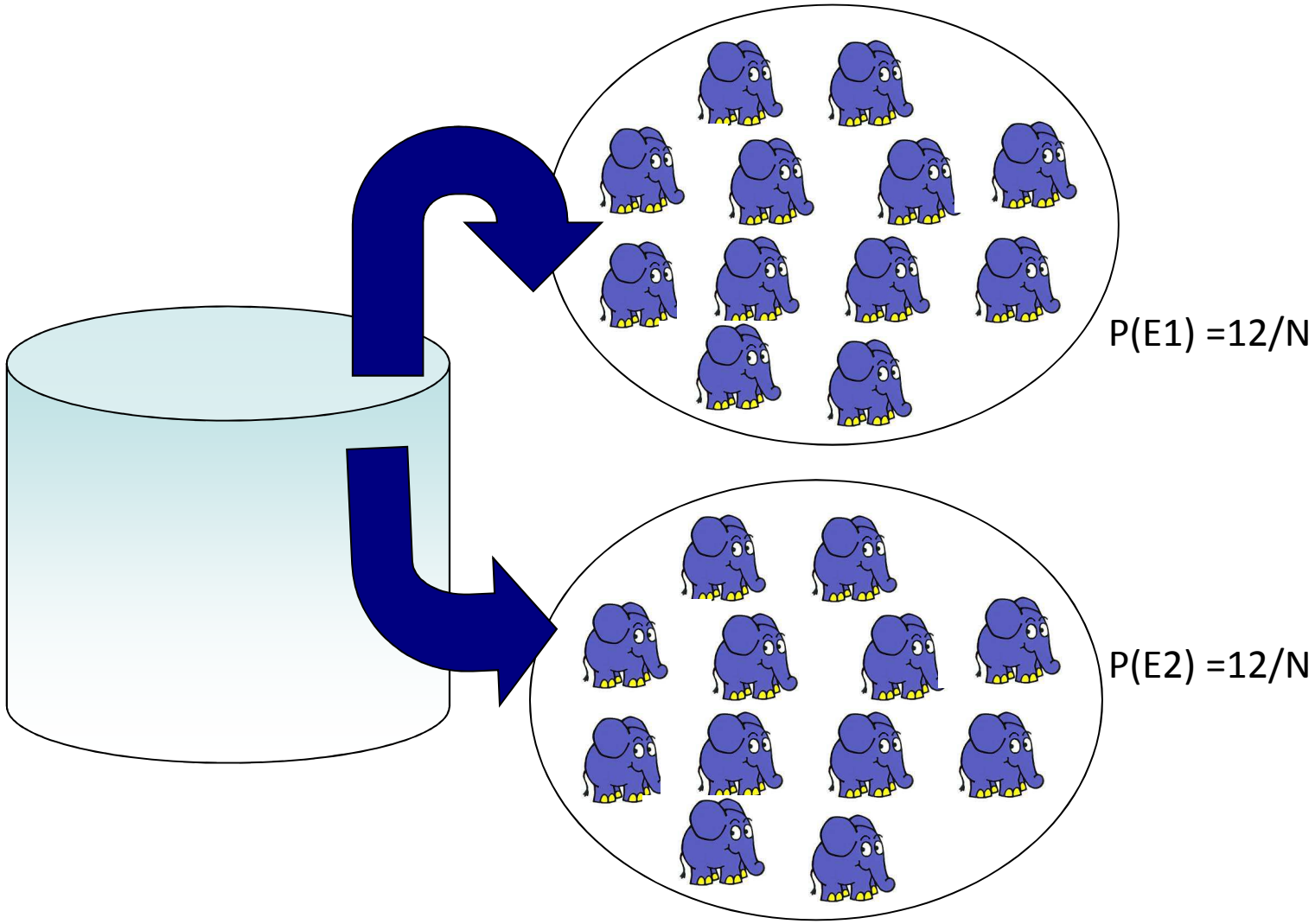
- Independent case ascertainment
- Capture–recapture method
- Death certificate methods
  - DCI method
  - ‘flow’ method

# Independent case ascertainment

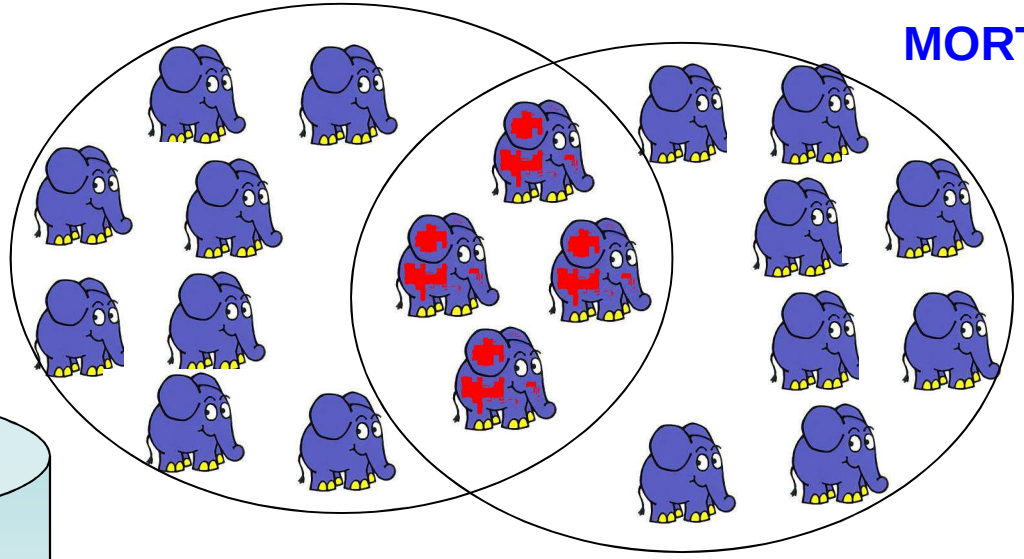
- Re-screening the sources that had been used by the registry, to detect any case missed during the registration process.
- The use of one or more independent sources of cancer cases, and comparison of the registry database with them

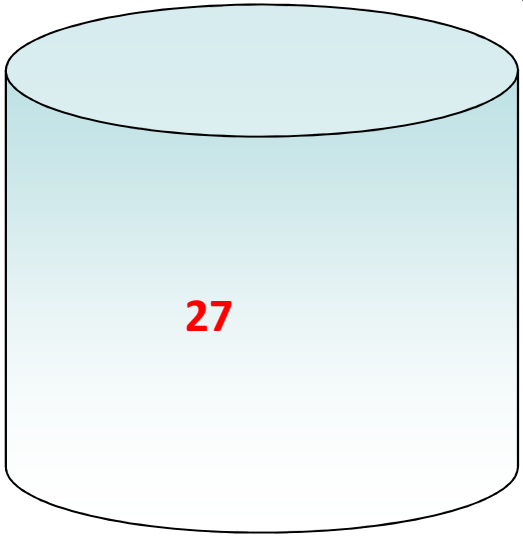


# Capture–recapture method



# Capture–recapture methods




$P(E1 \& E2) = P(E1) * P(E2) =$

$12/N * 12/N$

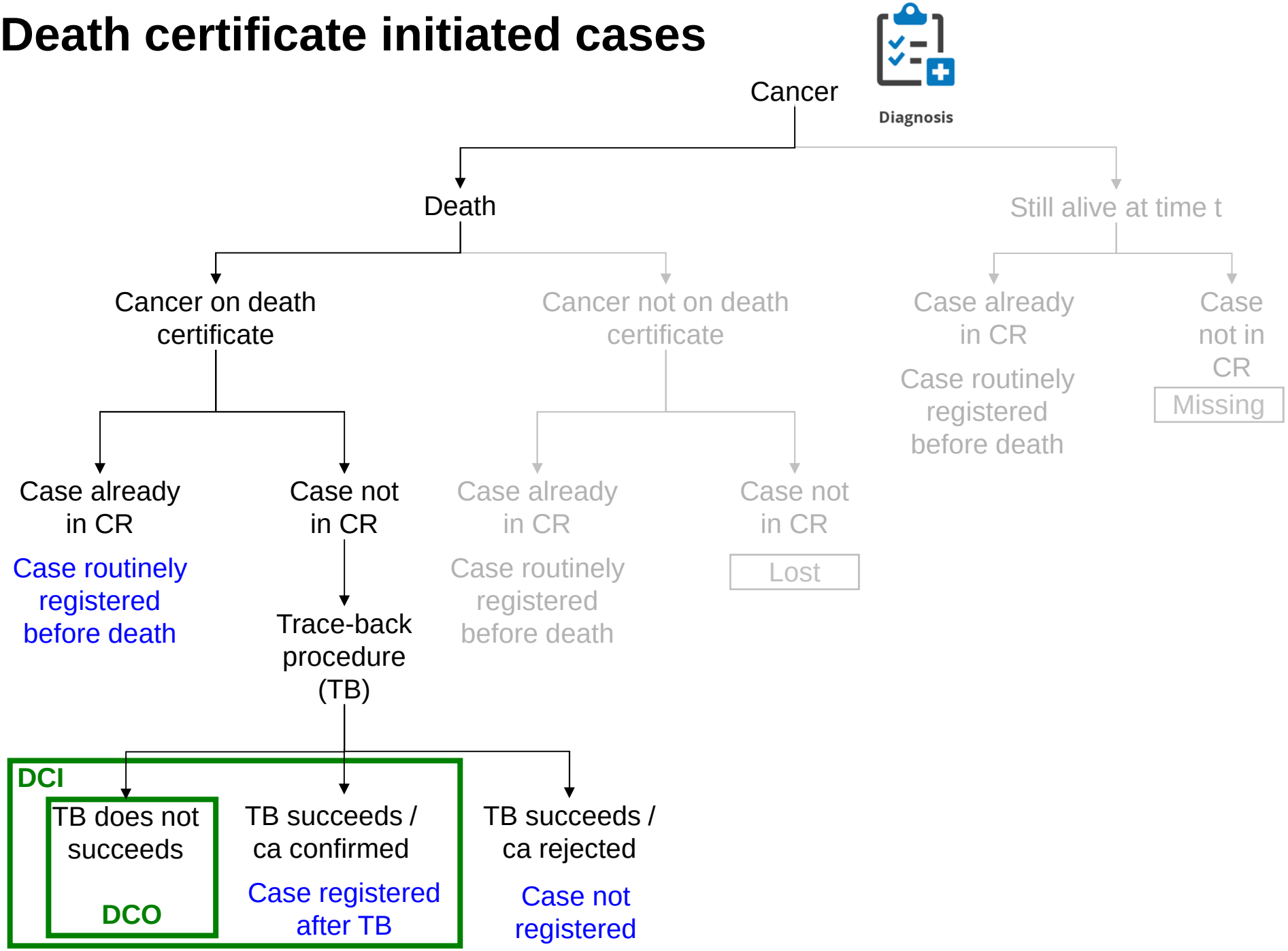
$144/ N^2$

$P(E1 \& E2) = 4/20$

$4/20 = 144/N^2$

$N = RADQ(720)$

# Death certificate initiated cases



# Death certificate initiated cases

- High DCO% -> informative system is not able to trace cancer cases history
- Low DCO%  -> efficient case-finding or efficient trace-back of cases
- High DCI% -> informative system is not able to capture all cancer cases
- A low DCI% is associated with an high completeness
- The DCI% will always be equal to, or greater than, the DCO%

Jpn J Clin Oncol 2007;37(2)150–155
doi:10.1093/jjco/hyl143

Epidemiology Report

# A Mathematical Estimation of True Cancer Incidence Using Data from Population-based Cancer Registries

Ken-ichi Kamo[1,2], Satoshi Kaneko[1,3], Kenichi Satoh[4], Hirokazu Yanagihara[5], Shoichi Mizuno[6] and Tomotaka Sobue[1]

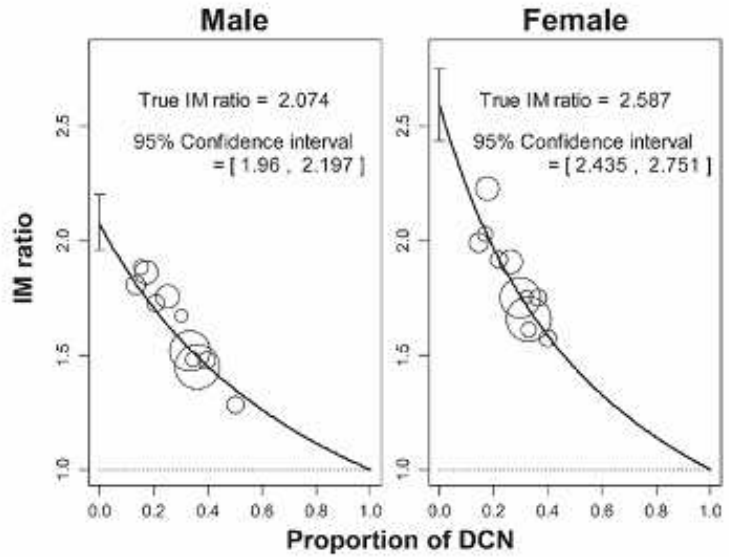



**Figure 2.** Regression curves for the estimate of the 'true IM ratios' for all cancer sites. The size of the plot is proportional to the population size covered by the registries. The line denotes the regression curve. A 95% confidence interval of the 'true IM ratio' is expressed at the left edge of regression curve. IM and DCN refer to incidence/mortality and death certificate notification, respectively.

# Death certificate initiation

| | MORT YES | MORT NO | |
|---|---|---|---|
| PatRep YES | **A** | **B** | $n_{1.}$ |
| PatRep NO | **C** | **D** | $n_{0.}$ |
| | $n_{.1}$ | $n_{.0}$ | $n_{..}$ |

Completeness = (A + B +C) / (A + B +C + D)

Assuming B/A = C/D therefore D = C*B/A

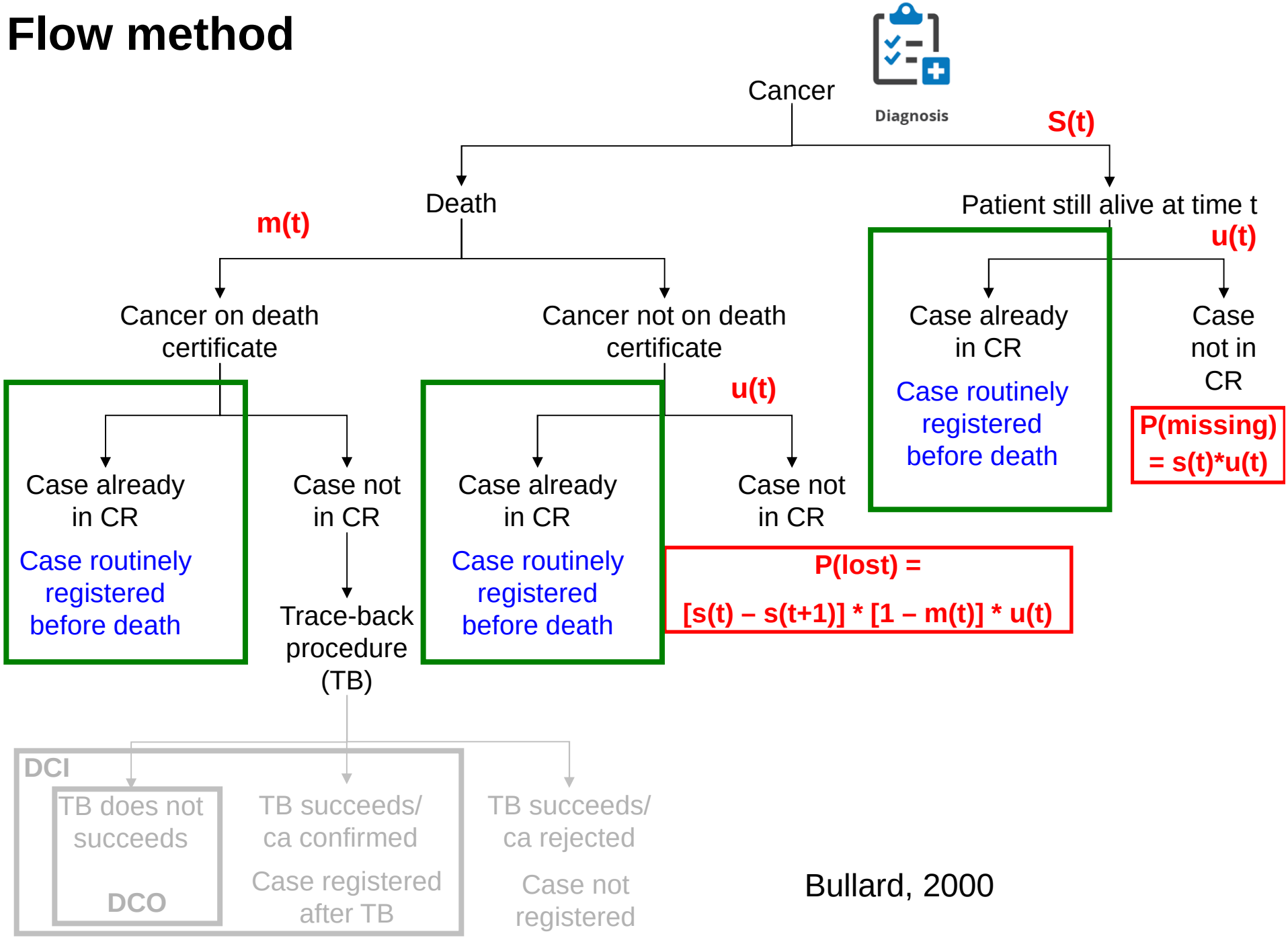
Completeness = (A + B +C) / (A + B +C **+** C*B/A)

1 / { 1 + B/A * C / (A + B +C )} %DCI

(A+B)/A

1 / { I:M * (%DCI)}

# Flow method

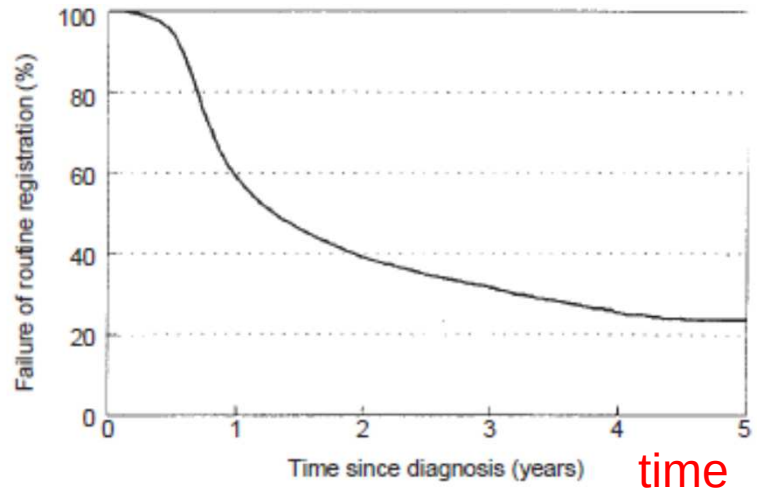



Bullard, 2000

# Flow method

- The survival distribution: standard indicators - s(t)
- The probability that cancer is mentioned on the death certificate: deaths for which the death certificate includes a mention of cancer over the total number of deaths – m(t)

s(t)*u(t)




time

**Figure 3** The probability that a surviving cancer patient remains unregistered by time since diagnosis, *u(t)*

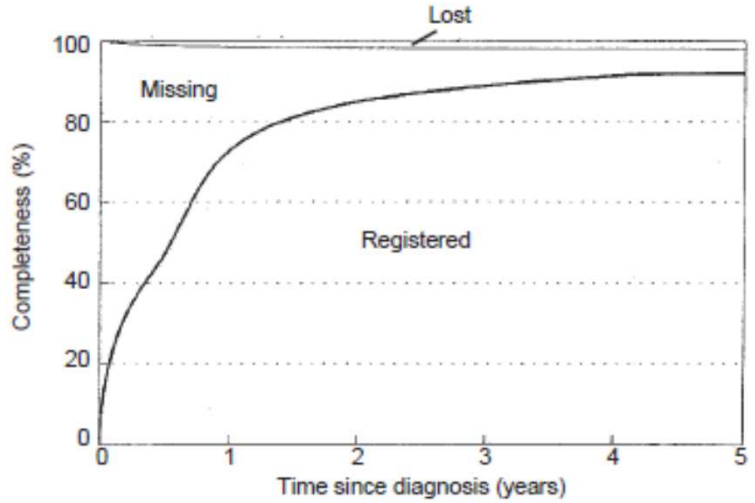


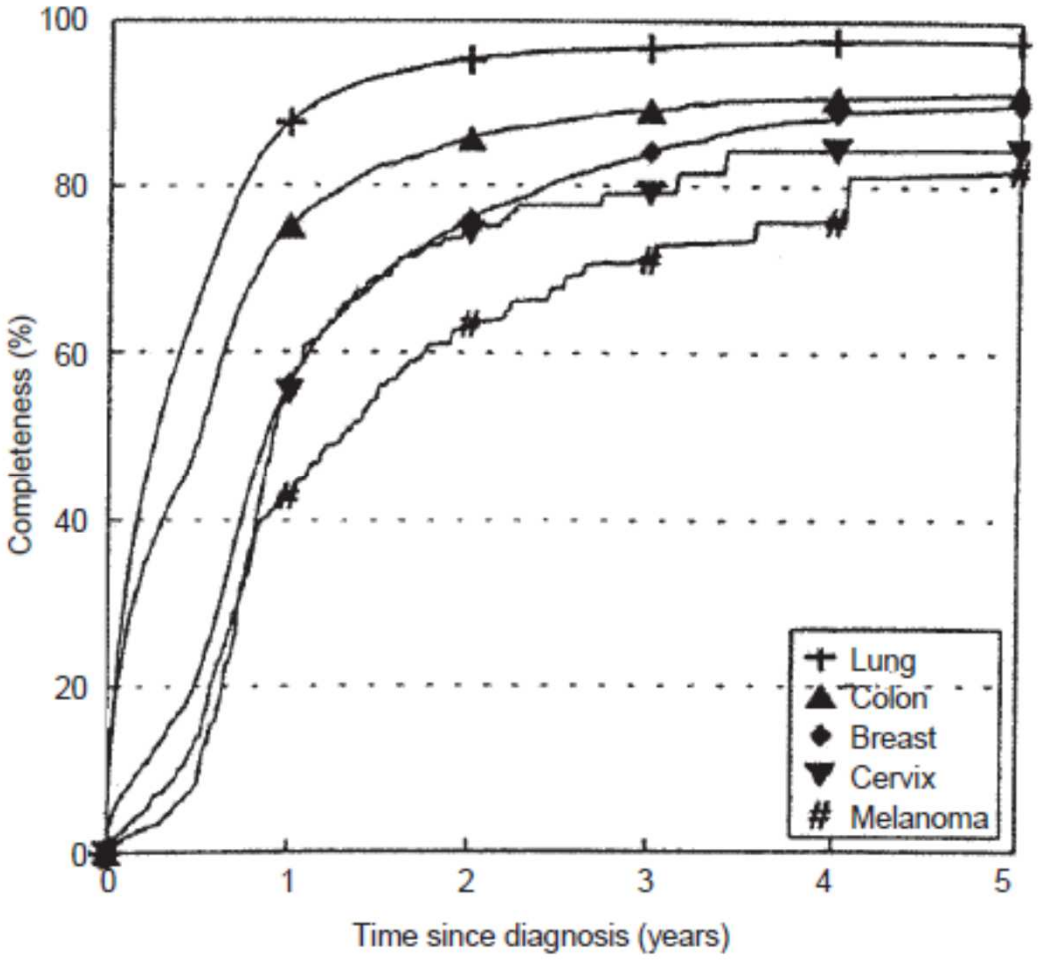

**Figure 4** The completeness of cancer registration *C(T)* at Thames Cancer Registry for all cancers diagnosed in 1987 (excluding non-melanoma skin cancer) by time since diagnosis

time

Bullard, 2000

# Flow method

Completeness(t)




time

Bullard, 2000

# Flow method (modified)

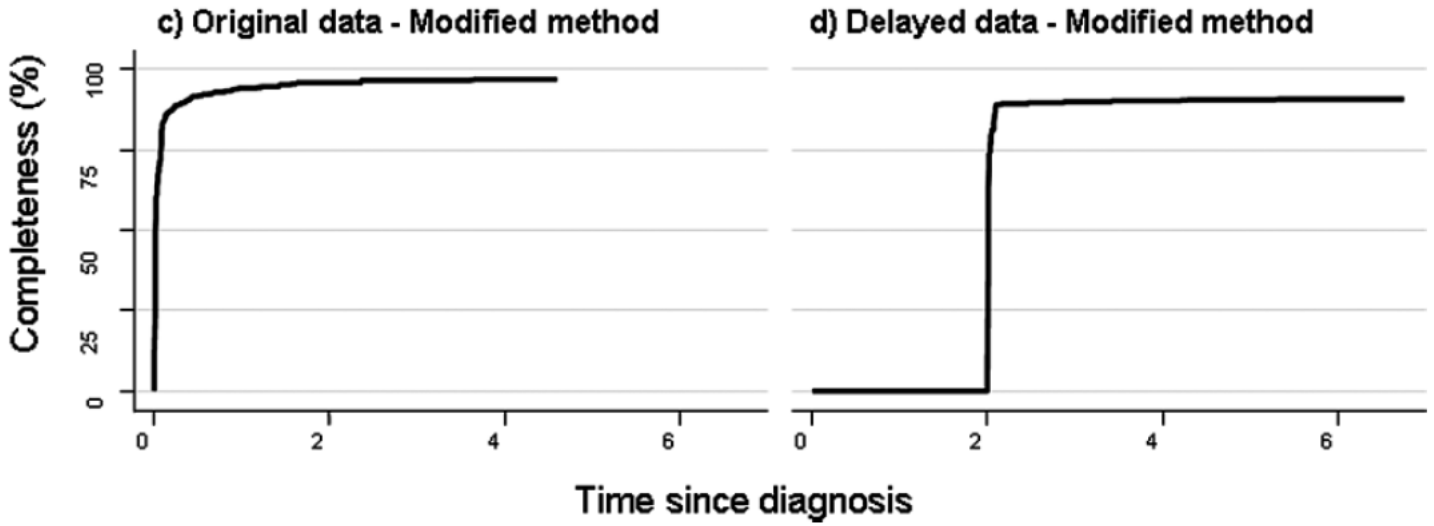



**Fig. 1** Completeness estimated by original (**a** and **b**) and modified (**c** and **d**) method using original data (**a**–**c**) and data with registration date artificially delayed by two years (**b** and **d**).

Montanaro, 2006

# Delay adjusting models in the US

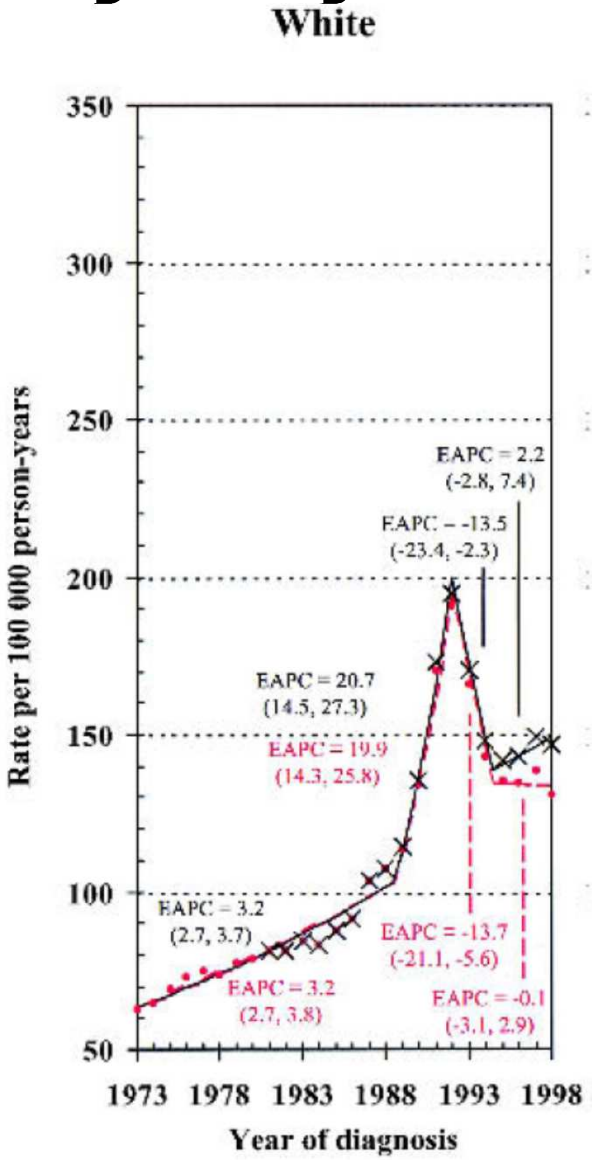



**Fig. 4.** Incidence and reporting-adjusted rates for prostate cancer by race. Rates are per 100 000 person-years and are age-adjusted to the 1970 U.S. standard population. (●) = incidence data and (×) = reporting-adjusted data. Regression lines are calculated using the joinpoint regression program. EAPC = estimated annual percentage change in the regression line. Numbers in parentheses are the 95% confidence intervals of the EAPC. Data are from the Surveillance, Epidemiology and End Results (SEER) program August 2000 submission.

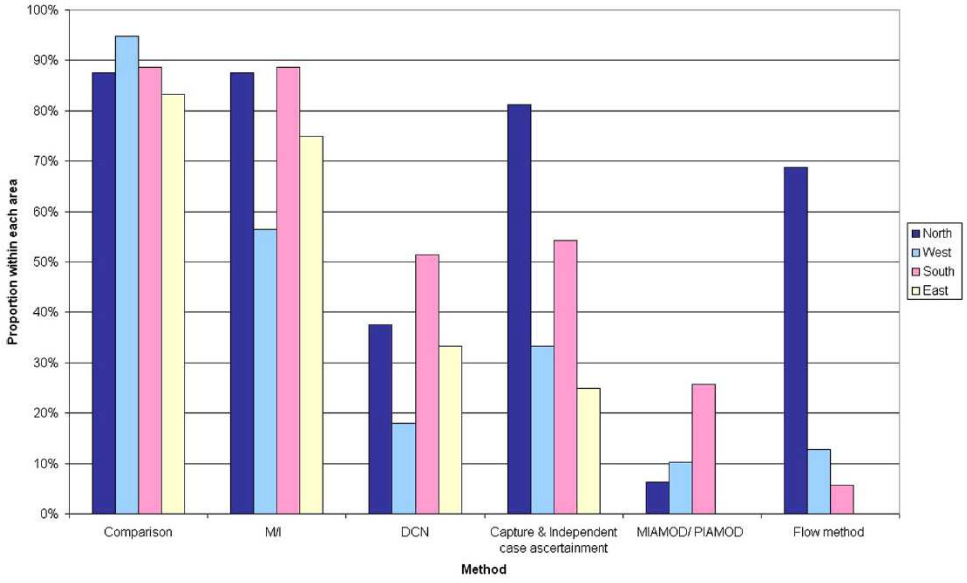
Table 4
Frequencies of use of different methods by general cancer registries (GCRs) and specialised cancer registries (SCRs); total number of registries (GCRs + SCRs) = 102. Multiple answers on methods were allowed.

| Method | GCRs | SCRs | % of total number of registries (102) |
|---|---|---|---|
| Historical comparison [7] | 68 (37) | 13 (3) | 79 |
| Compare incidence with incidence in reference registry [7] | 54 (35) | 9 (3) | 62 |
| Comparison with reference registry (indirect standardisation) [7] | 31 (16) | 3 (2) | 33 |
| Death certificate notification (DCN) method [7] | 31 (18) | 1 (1) | 31 |
| DCN method (Ajiki's formula) [13] | 9 (7) | 0 (0) | 9 |
| M/I ratio: compute and compare with other registries/national average [7] | 62 (34) | 2 (2) | 63 |
| M/I ratio: compute and compare with own registry in previous year(s) [7] | 68 (47) | 3 (3) | 70 |
| Log-linear models [8] | 11 (5) | 0 (0) | 11 |
| Independent case ascertainment [7] | 30 (15) | 4 (1) | 33 |
| Flow method (Bullard) [14–16] | 17 (12) | 1 (1) | 18 |
| MIAMOD/PIAMOD [9] | 14 (11) | 0 (0) | 14 |
| Capture recapture [10–12] | 27 (15) | 4 (1) | 30 |
| Other | 10 (6) | 3 (1) | 13 |

In parentheses: number of registries that reported the availability of dedicated software for that method. M/I ratio, mortality/incidence ratio.


Completeness and timeliness: Cancer registries could/should improve their performance

R. Zanetti [a,*], I. Schmidtmann [b], L. Sacchetto [a], F. Binder-Foucard [c], A. Bordoni [d], D. Coza [e], S. Ferretti [f], J. Galceran [g], A. Gavin [h], N. Larranaga [i], D. Robinson [j], L. Tryggvadottir [k], E. Van Eycken [l], V. Zadnik [m], J.W.W. Coebergh [n], S. Rosso [a]



# Pubmed: cancer registry completeness (2015-2016)

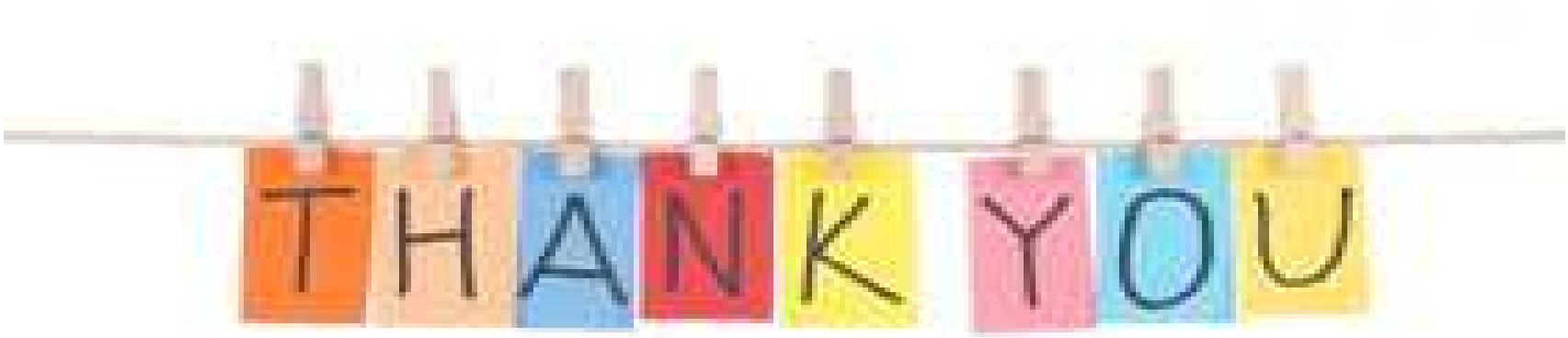
- Hansen S, Nielsen J, Laursen RJ, et al. The Danish Neuro-Oncology Registry: establishment, completeness and validity. BMC Res Notes. 2016 Aug 30;9(1):425.
- Kleijwegt M1, Ho V, Visser O, Godefroy W, van der Mey A. Real Incidence of Vestibular Schwannoma? Estimations From a National Registry (Netherlands). Otol Neurotol. 2016 Oct;37(9):1411-7.
- Jeffree SM, Mihat O, Lukman KA, et al. Surveillance Evaluation of the National Cancer Registry in Sabah, Malaysia. . Asian Pac J Cancer Prev. 2016;17(7):3123-9.
- Linder G, Lindblad M, Djerf P, Elbe P, Johansson J, Lundell L, Hedberg J. Validation of data quality in the Swedish National Register for Oesophageal and Gastric Cancer. Br J Surg. 2016 Sep;103(10):1326-35.
- Khodadost M, Yavari P, Khodadost B et al. Estimating the Esophagus Cancer Incidence Rate in Ardabil, Iran: A Capture-Recapture Method. Iran J Cancer Prev. 2016 Feb 17;9(1):e3972.
- Fung JW, Lim SB, Zheng H et al. Data quality at the Singapore Cancer Registry: An overview of comparability, completeness, validity and timeliness. Cancer Epidemiol. 2016 Aug;43:76-86
- Linder G, Lindblad M, Djerf J et al. Validation of data quality in the Swedish National Register for Oesophageal and Gastric Cancer. Br J Surg. 2016 Sep;103(10):1326-35.
- Morling JR, Grant R, Brewster DH. Completeness of primary intracranial tumour recording in the Scottish Cancer Registry 2011-2012. Public Health. 2016 Jun 28
- Arboe B, El-Galaly TC, Clausen MR et al. The Danish National Lymphoma Registry: Coverage and Data Quality. PLoS One. 2016 Jun 23;11(6)
- Khodadost M, Mosavi-Jarrahi A, Hashemian SS et al. Estimating the Completeness of Lung Cancer Registry in Ardabil, Iran with a Three-Source Capture-Recapture Method. Asian Pac J Cancer Prev. 2016;17 Spec No.:225-9.
- Mohammadi G, Akbari ME, Mehrabi Y, Ghanbari et al.. Estimating Completeness of Cancer Registration in Iran with Capture-Recapture Methods.Asian Pac J Cancer Prev. 2016;17 Spec No.:93-9.
- Puckett M, Neri A, Rohan E, Clerkin C et al. Evaluating Early Case Capture of Pediatric Cancers in Seven Central Cancer Registries in the United States, 2013. Public Health Rep. 2016 Jan-Feb;131(1):126-36.
- Schindler M, Mitter V, Bergstraesser E, Gumy et al. Death certificate notifications in the Swiss Childhood Cancer Registry: assessing completeness and registration procedures. Swiss Med Wkly. 2015 Dec 23;145:w14225
- Tran JM, Schwartz R, Fung K et al. Comprehensive capture of cutaneous melanoma by the Ontario Cancer Registry: validation study using community pathology reports. Cancer Causes Control. 2016 Jan;27(1):137-42. doi: 10.1007/s10552-015-0690-5. Epub 2015 Nov Sharifian R, SedaghatNia MH, Nematolahi M, Zare N, Barzegari S 4.
- Estimation of Completeness of Cancer Registration for Patients Referred to Shiraz Selected Centers through a Two Source Capture Re-capture Method, 2009 Data. Asian Pac J Cancer Prev. 2015;16(13):5549-56.
- Kropf S, Burger E, Radinski et al. Completeness and quality of baseline data and follow-up in cancer registry--an analysis on the example of colorectal cancer. Dtsch Med Wochenschr. 2015 May;140(11):e106-13
- Dimitrova N, Parkin DM. Data quality at the Bulgarian National Cancer Registry: An overview of comparability, completeness, validity and timeliness. Cancer Epidemiol. 2015 Jun;39(3):405-13.
- Bailly L, Daurès JP, Dunais B, Pradier C. Bayesian estimation of a cancer population by capture-recapture with individual capture heterogeneity and small sample. BMC Med Res Methodol. 2015 Apr 24;15:39.
- al-Haddad BJ, Jedy-Agba E, Oga et al. Comparability, diagnostic validity and completeness of Nigerian cancer registries. Cancer Epidemiol. 2015 Jun;39(3):456-64
- Khodadost M, Yavari P, Babaei M et al. Estimating the completeness of gastric cancer registration in Ardabil/Iran by a capture-recapture method using population-based cancer registry data. Asian Pac J Cancer Prev. 2015;16(5):1981-6.
- Kearney TM, Donnelly C, Kelly JM et al. Validation of the completeness and accuracy of the Northern Ireland Cancer Registry. Cancer Epidemiol. 2015 Jun;39(3):401-4.

# Conclusions

- Simple methods to evaluate completeness are available
- Information on DCI case, registration date, modification date are necessary
- A comprehensive evaluation of Cancer Registries data quality in Europe is needed

# References


- Bray F, Parkin DM. Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness. Eur J Cancer. 2009 Mar;45(5):747-55
- Parkin DM, Bray F. Evaluation of data quality in the cancer registry: principles and methods Part II. Completeness. Eur J Cancer. 2009 Mar;45(5):756-64.
- Zanetti R, Schmidtmann I, Sacchetto L et al. Completeness and timeliness: Cancer registries could/should improve their performance. Eur J Cancer. 2015 Jun;51(9):1091-8.
- Møller H, Richards S, Hanchett N eta al. Completeness of case ascertainment and survival time error in English cancer registries: impact on 1-year survival estimates. Br J Cancer. 2011 Jun 28;105(1):170-6
- Rubin DB. Inference and missing data. Biometrika, 1976; 63, 581-592.
- P. Armitage and R. Doll. The Age Distribution of Cancer and a Multi-stage Theory of Carcinogenesis. Br J Cancer. 1954 Mar; 8(1): 1–12.
- Bullard J, Coleman MP, Robinson D et al. Completeness of cancer registration: a new method for routine use. Br J Ca 2000; 82: 1111-1116.
- http://www.krebsdaten.de [Centre for Cancer Registry Data (ZfKD)]
- Contiero P, Tittarelli A, Maghini A et al. Comparison with manual registration reveals satisfactory completeness and efficiency of a computerized cancer registration system. J Biomed Inform. 2008 Feb;41(1):24-32. Epub 2007 Mar 21.
- Montanaro F, Robinson D, Bordoni A, Lutz JM.J Public Health (Oxf). 2006 Sep;28(3):274-7. Epub 2006 Jul 18. A modification to the flow method to estimate completeness in cancer registries with delayed registration.
- Crocetti, E; Miccinesi, G; Paci, E; Zappa, M  An application of the two-source capture–recapture method to estimate the completeness of the Tuscany Cancer Registry, Italy. European Journal of Cancer Prevention: October 2001 - Volume 10 - Issue 5 - pp 417-423
- Lewis DR, Chen HS, Cockburn M et al. Preliminary estimates of SEER cancer incidence for 2013. Cancer. 2016 May 15;122(10):1579-87.
- Lewis DR, Chen HS, Midthune DN et al. Cancer. Early estimates of SEER cancer incidence for 2012: Approaches, opportunities, and cautions for obtaining preliminary estimates of cancer incidence. 2015 Jun 15;121(12):2053-62.
- Huang L, Midthune D, Krapcho M et al. Adjusting for reporting delay in cancer incidence when combining different sets of cancer registries. Biom J. 2013 Sep;55(5):755-70.
- Clegg LX, Feuer EJ, Midthune DN et al. Impact of reporting delay and reporting error on cancer incidence rates and trends. J Natl Cancer Inst. 2002 Oct 16;94(20):1537-45.




Thanks to E. Crocetti (JRC) for revising my presentation