

The Use of a New Completeness Measure and its Application

Description and use of the “Flow” method of Bullard et al, British Journal of Cancer, 2000

Presented to NAACCR Annual Meeting, June 2010, Québec City

Ron Dewar
Vickey Bu

Surveillance and Epidemiology Unit



Estimation of Completeness of cancer registration

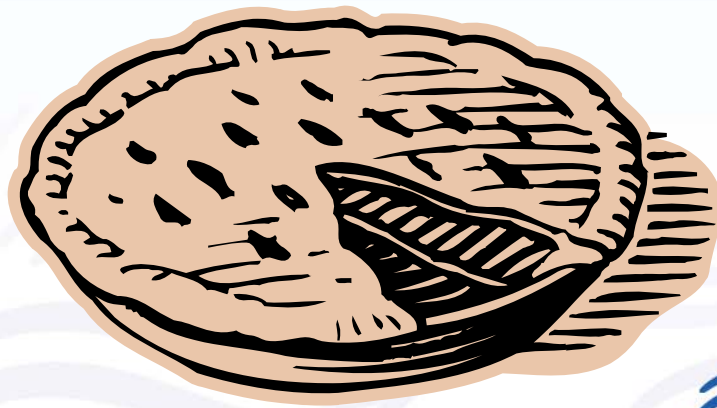
- No realistic 'Gold Standard' method
- All methods have strengths and limitations
- Look for consistency between methods, especially methods that use different assumptions or data sources



?



?



Methods – a brief overview

- Semi-Quantitative: completeness *relative* to an external / historic standard
 - historic data methods
 - mortality:incidence ratios
 - number of sources per case
 - % histologically verified



Methods – a brief overview (2)

Quantitative – use information about *observed* cases to estimate the number of cases that have been *unreported*

- independent case-finding
- capture-recapture
- spatio-temporal modeling
- delay adjustment / modeling
- death certificate (DC) methods

DC and M:I
'Flow'



Completeness

At any time after the diagnosis of cancer, cases not registered can be classified as:

- “**Missing**” – diagnosed but not yet registered. May be registered later...
- “**Lost**” – diagnosed, not registered, *deceased*, and cancer not mentioned on the death certificate (DC). Will *never* be registered.



Completeness: Logical Framework

Registered ?

*Cancer has been
diagnosed and
patient is ...*

Alive

Dead - ca. on DC

- ca. not on DC

Yes

No

Alive	Yes	No
Dead - ca. on DC	Yes	No
- ca. not on DC	Yes	No

Registered, or will
be

“Missing”

“Lost”

Registered, or will be	Yes	No
“Missing”	Yes	No
“Lost”	Yes	No



*Cancer Care
Nova Scotia*

Completeness

➤ To estimate proportion 'Missing' and 'Lost' at time t since diagnosis, we need time-dependent proportions:

- survival $s(t)$
- mention of cancer on DC $m(t)$
- time to registration $u(t)$

All estimated from *registered patients*

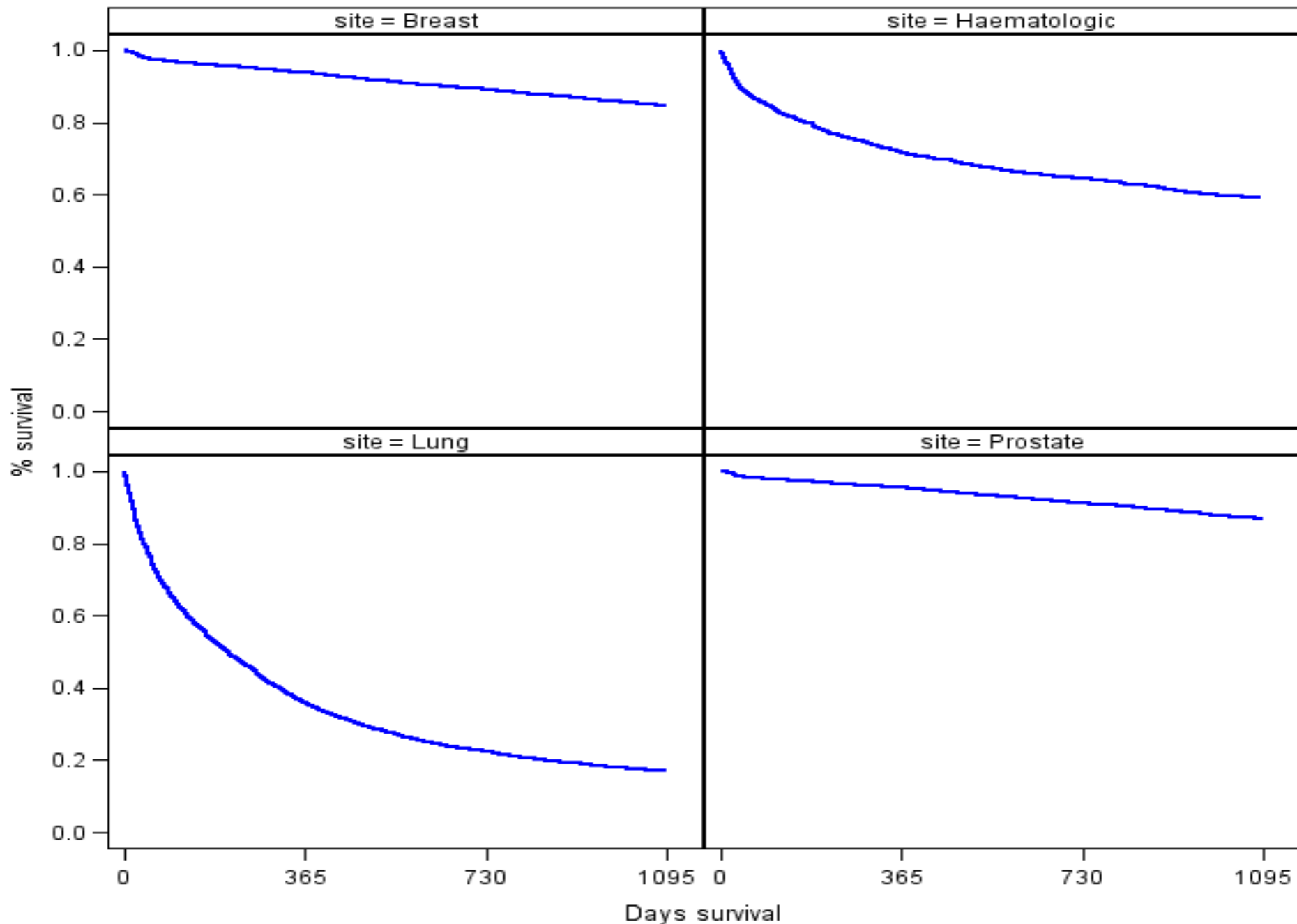


Survival:

$s(t)$: % alive at start of interval

- standard registry product
- time to event process with censoring
- requirements:
 - date of diagnosis
 - date of death (any cause)
 - end of follow-up / date last contact
 - age, sex, cancer type, etc...

Survival Proportion, selected cancers by time since diagnosis



Mention:

m(t): % with cancer on DC those dying in interval

Among *registered* patients who die, what proportion have a cancer diagnosis on their DC?

Cancer may *not* be the cause of death

Requirements:

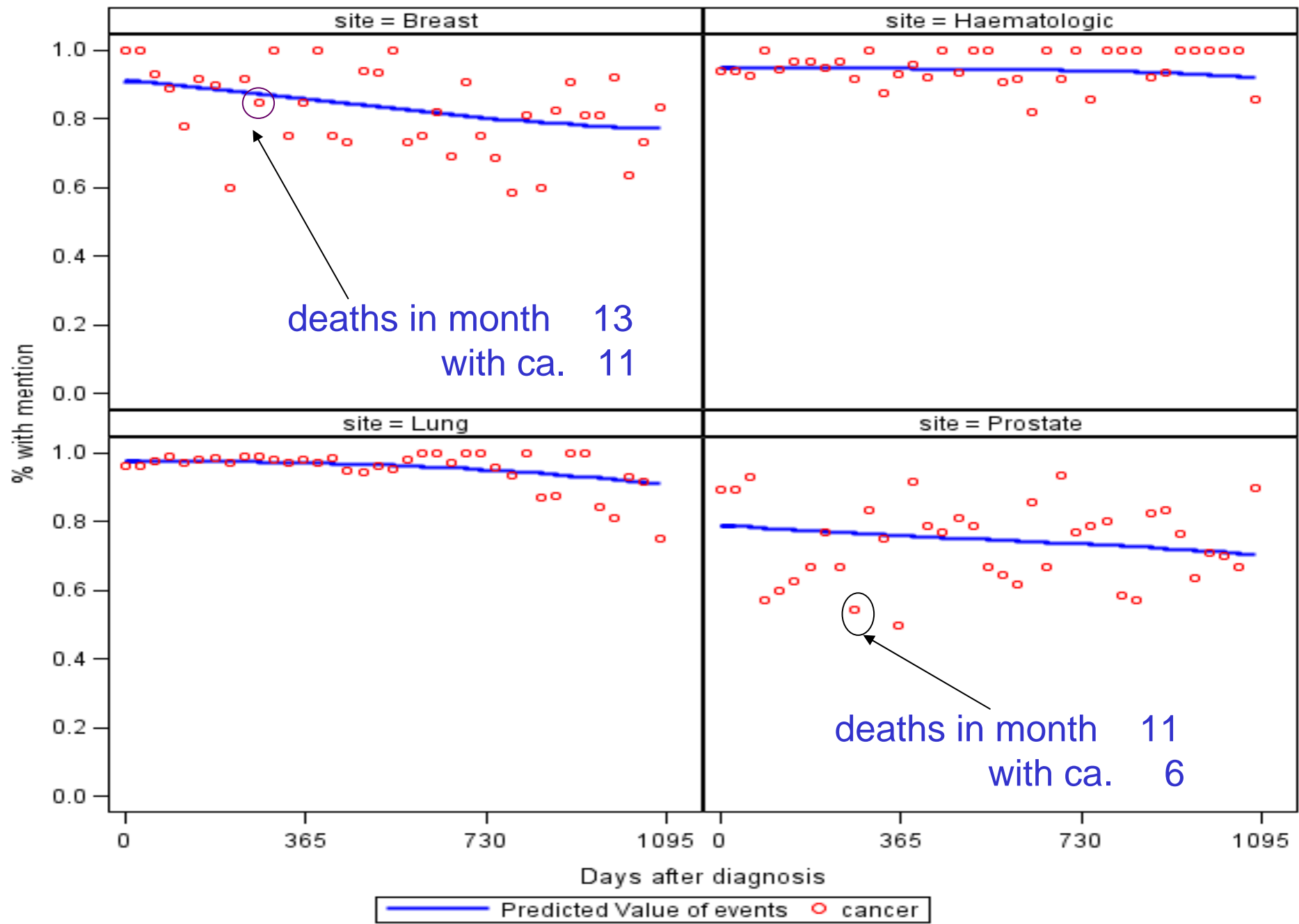
- registry receives all DCs

- note any cancer mentioned on DC

- date of diagnosis



Mention of cancer on Death Certificate, selected cancers by time since diagnosis



Registration

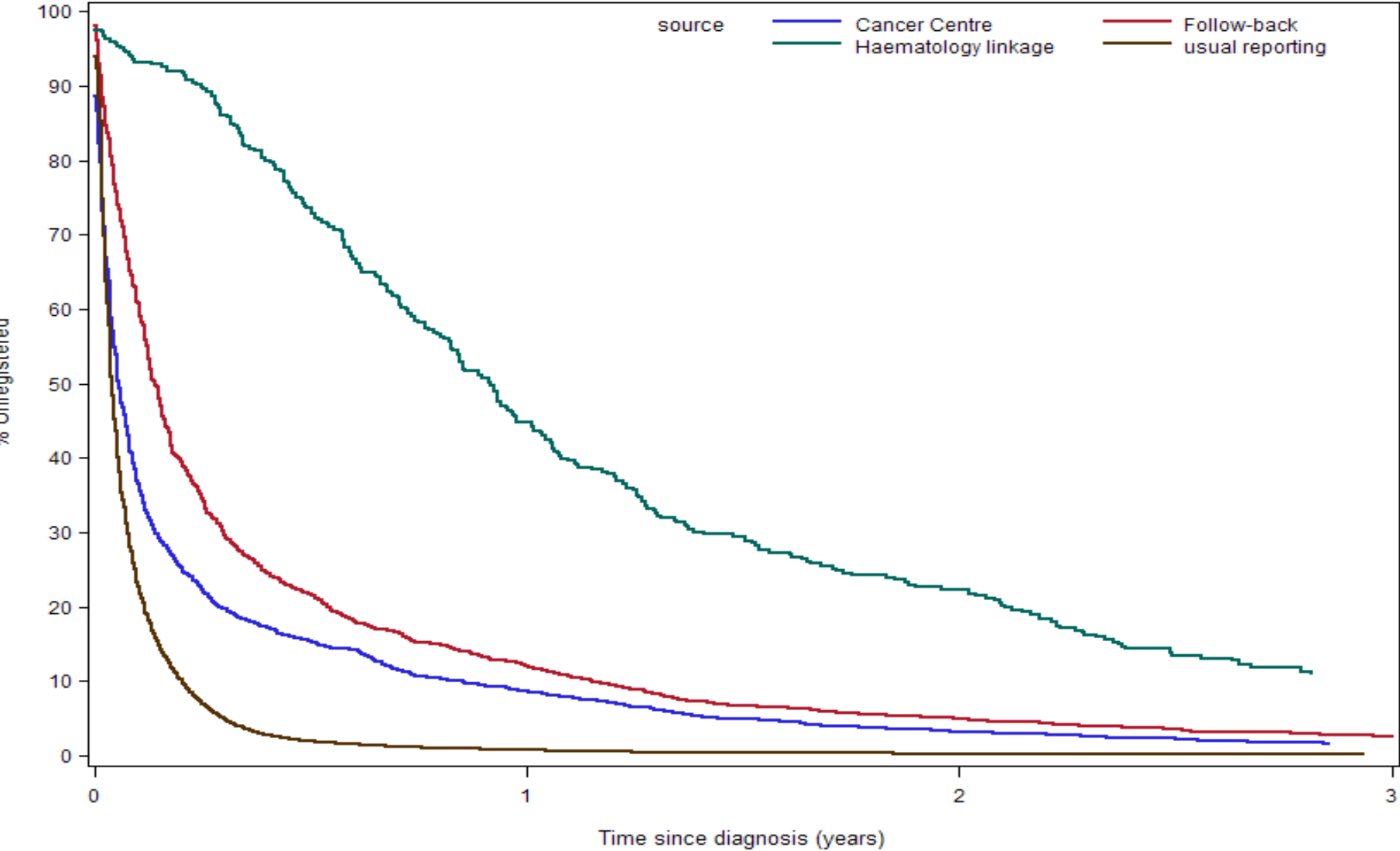
$u(t)$: % unregistered at time t

‘registration’: *earliest* date the case became known, after which it would not be possible for the registry to miss the case

- not overwritten by subsequent data entry
- will depend on registry processes
- estimated from cases selected from *diagnosis* or *registration* point of view



Unregistered by Time Since Diagnosis (Registration: 2002 - 2008) by Source of Registration



Estimates at time t_i

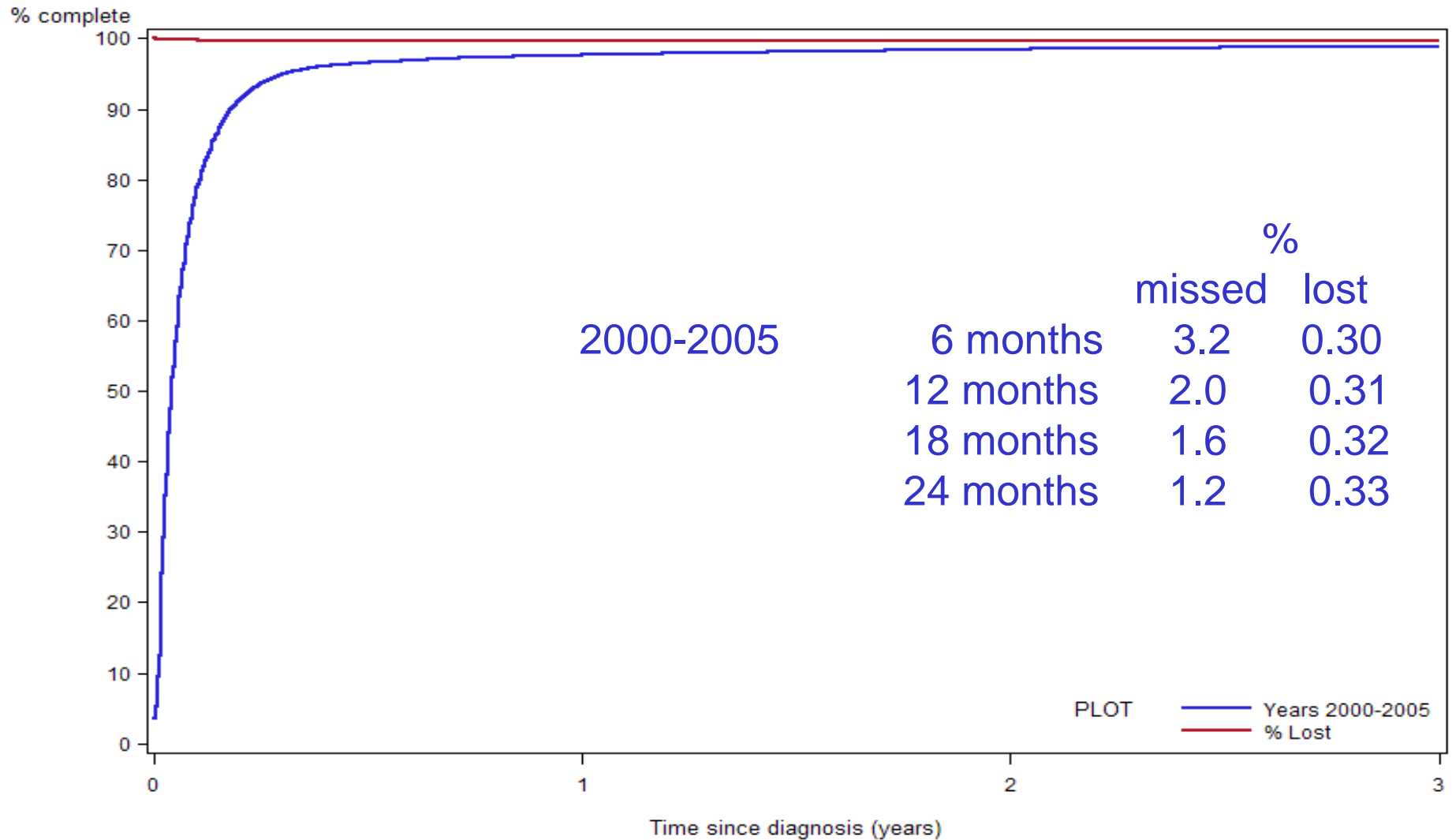
- **Missing:** p (alive *and* unregistered)

$$= s(t_i) * u(t_i)$$

- **Lost:** p (death in interval t_i
and not registered yet
and cancer **not** on DC)

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

completeness by time since diagnosis



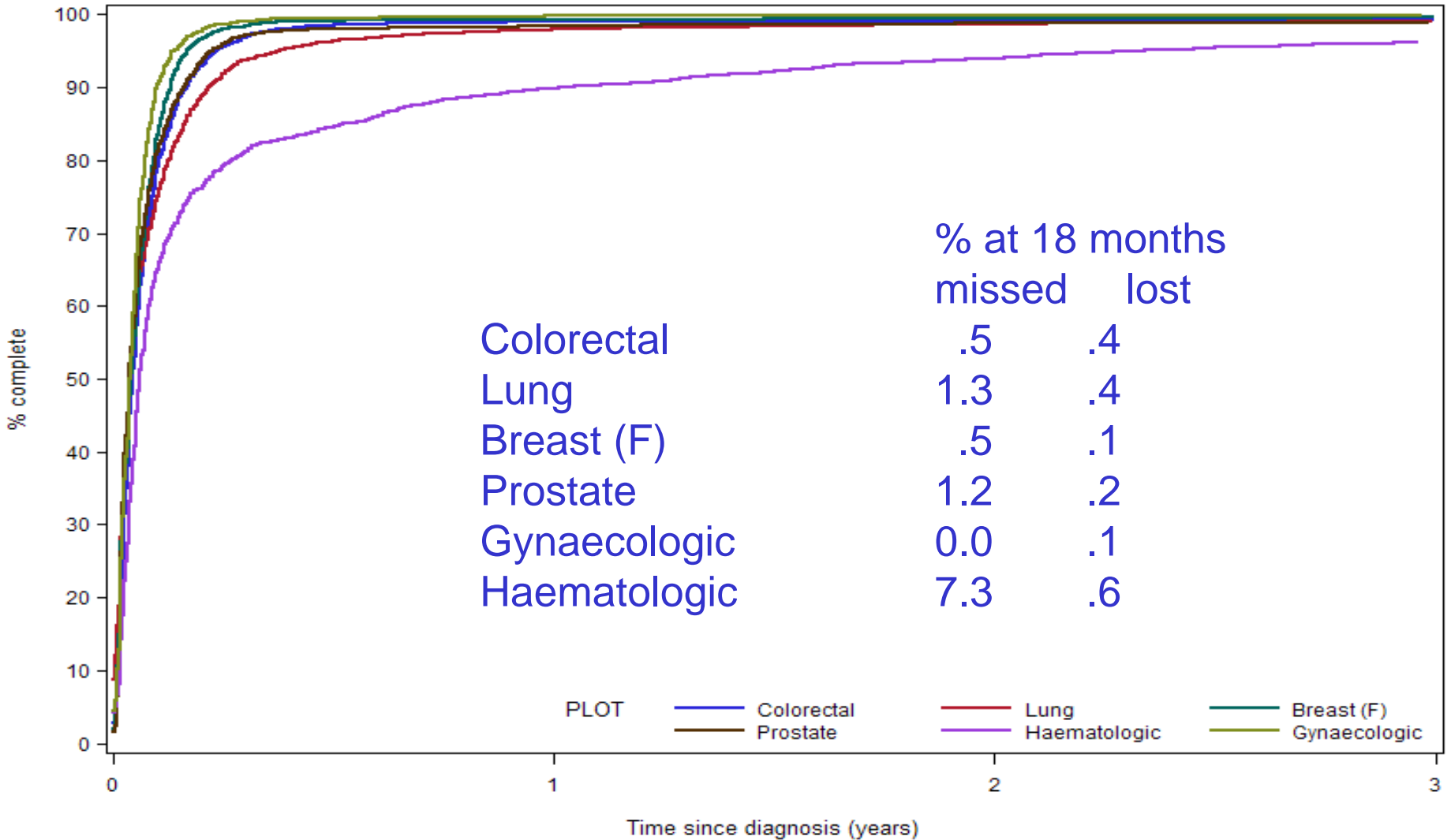
2000-2005

	% missed lost	
6 months	3.2	0.30
12 months	2.0	0.31
18 months	1.6	0.32
24 months	1.2	0.33

PLOT — Years 2000-2005
— % Lost

Years 2000-2005

completeness by time since diagnosis



Colorectal
Lung
Breast (F)
Prostate
Gynaecologic
Haematologic

% at 18 months
missed lost

PLOT

— Colorectal
— Prostate

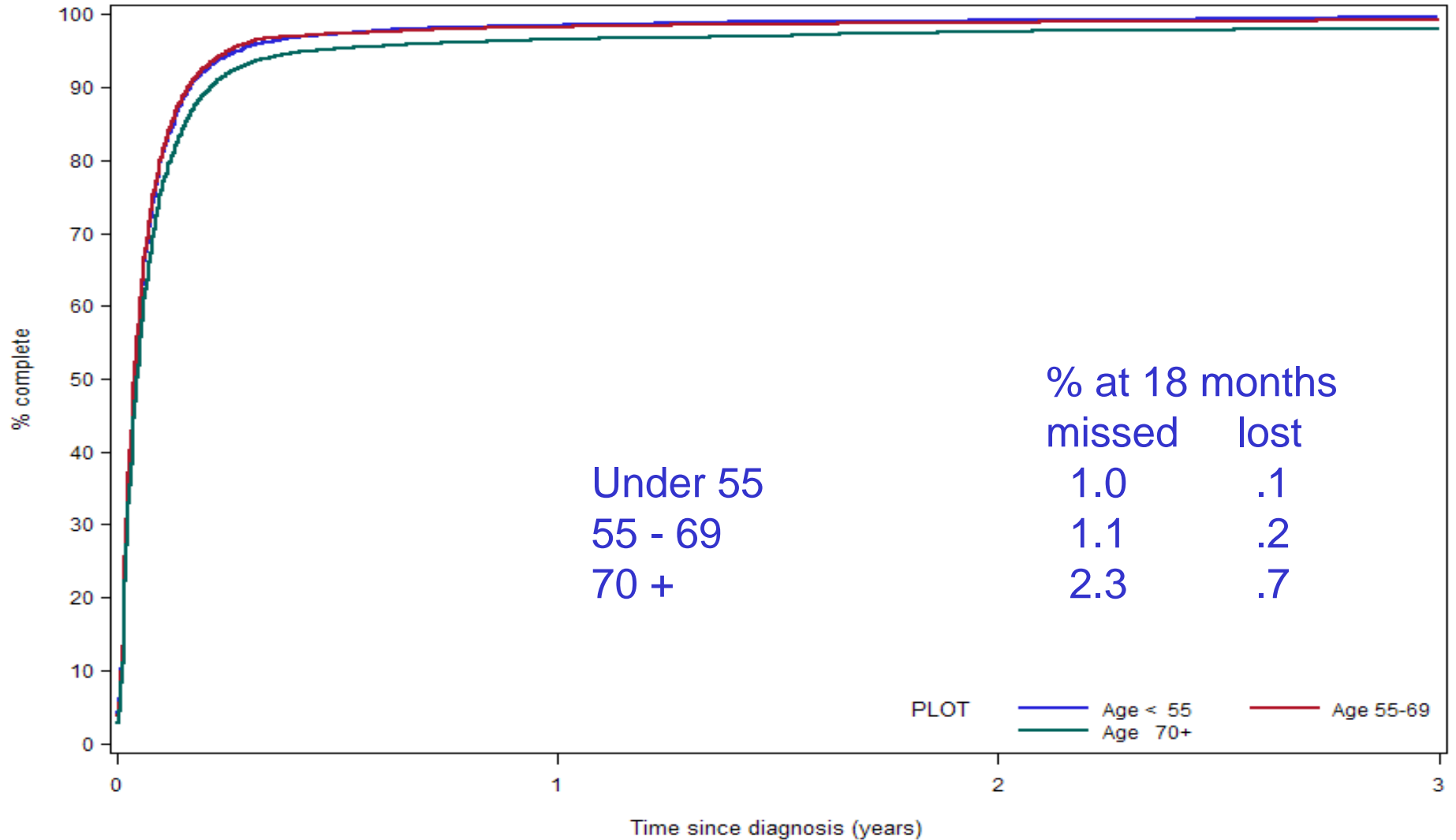
— Lung
— Haematologic

— Breast (F)
— Gynaecologic

Time since diagnosis (years)

by cancer type

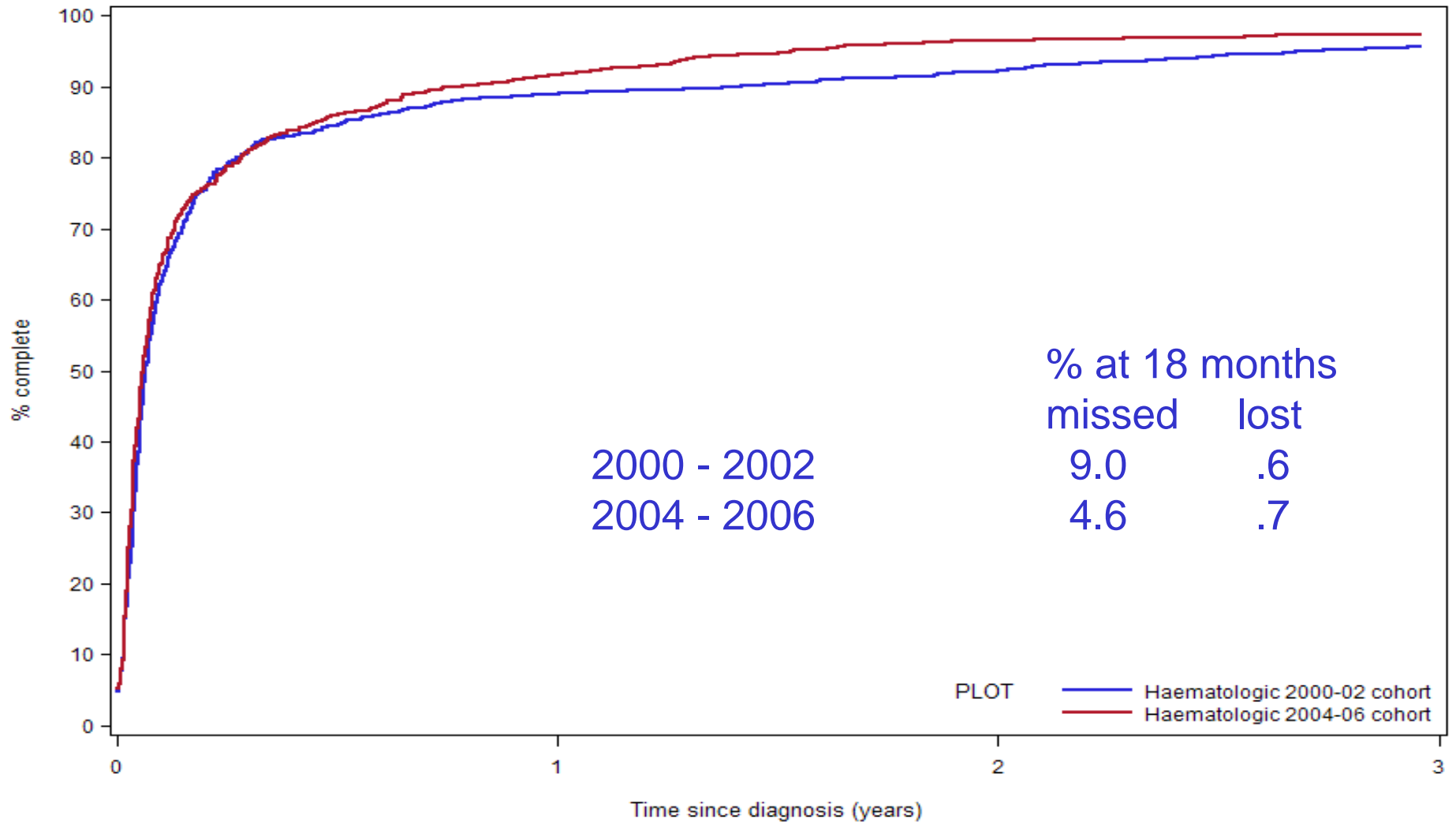
completeness by time since diagnosis



by age at diagnosis

completeness by time since diagnosis

diagnosed 2000 - 2005



% at 18 months
missed lost

2000 - 2002

9.0 .6

2004 - 2006

4.6 .7

PLOT

— Haematologic 2000-02 cohort
— Haematologic 2004-06 cohort

by year of diagnosis

Completeness: 'Flow' method

- Extensive data, unusual programming
- Identifies situations where cases may be relatively incomplete:
 - hematology, lung, older ages
- Assumptions:
 - independence of events
 - relevance of data on *registered* patients to understanding of *unregistered* patients

Completeness: work in progress

- Time to registration is key
what does 'registration' mean?
all cases, or just those deceased?
- How to handle DCO and DCI cases?
concern where DCO or DCI % is large
- Is there a fit to NAACCR certification needs?
(single year's incidence data)
- Confidence intervals?
- Effect of fatal disease on time to registration (lack of statistical independence)



conclusions

- + estimation at any time post diagnosis
- + independent of external standard
- - data requirements
- - programming effort involved
- ? Sensitivity to failure of assumptions
- May be most useful for internal comparisons (eg, targeted ascertainment strategies)



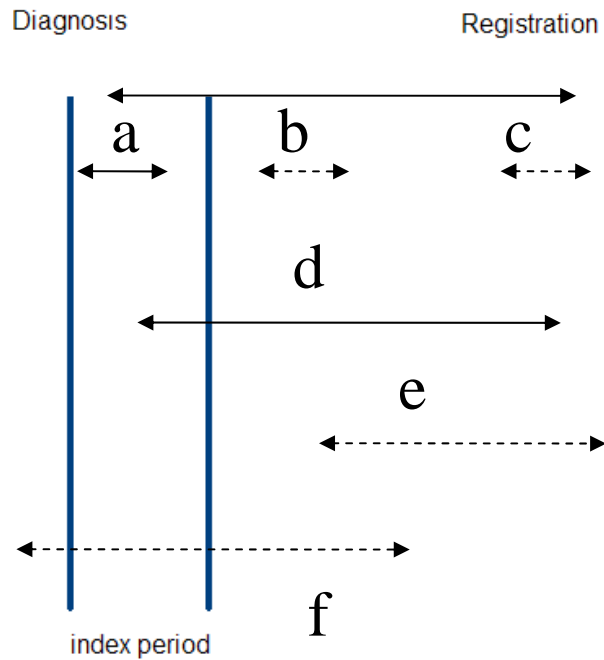
Completeness: take-home message

- Many methods are available
- All require a leap of faith (just different!)
- A comprehensive review of completeness should include several methods
 - internal comparisons
 - external standards



$u(t)$: % unregistered

DIAGNOSIS viewpoint



REGISTRATION viewpoint

