Estimating net survival using a life table approach

Enzo Coviello[1]

Joint work with
Paul Dickman[2], Karri Seppä[3], and Arun Pokhrel[3]

[1] Epidemiology Unit ASL BT, Barletta, Italy
[2] Karolinska Institutet, Stockholm, Sweden
[3] Finnish Cancer Registry, Helsinki, Finland

enzo.coviello@alice.it

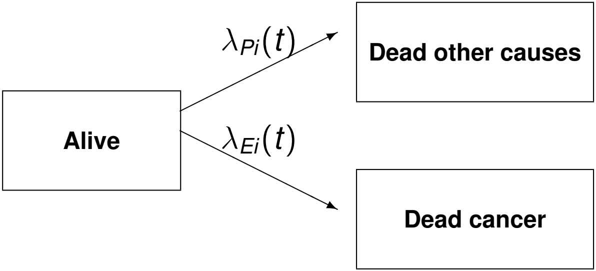
Italian Stata Users Group, Florence, November 2013

A key indicator

For cancer cases net survival is the probability of survival in the hypothetical scenario where the cancer under study is the only possible cause of death.

Although it is a hypothetical concept, in practice it is the key indicator for comparing cancer survival between countries and over time as it is independent of the mortality due to other diseases that also varies between countries and over time.

Competing risks



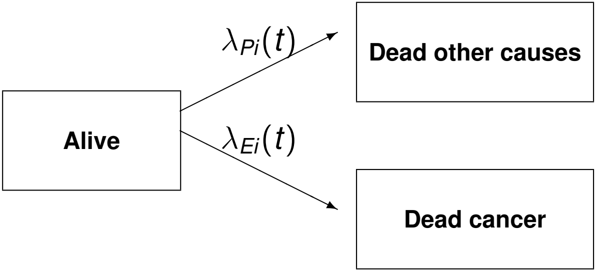
Additive Model

$$\lambda_{Oi}(t) = \lambda_{Pi}(t) + \lambda_{Ei}(t)$$

Net Survival

$$NS(t) = S_E(t) = exp(-\int_0^t \lambda_E)$$

Competing risks



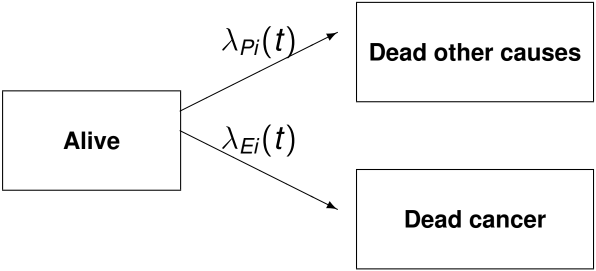
Additive Model

$$\lambda_{Oi}(t) = \lambda_{Pi}(t) + \lambda_{Ei}(t)$$

Net Survival

$$NS(t) = S_E(t) = exp(-\int_0^t \lambda_E)$$

Competing risks



Additive Model

$$\lambda_{Oi}(t) = \lambda_{Pi}(t) + \lambda_{Ei}(t)$$

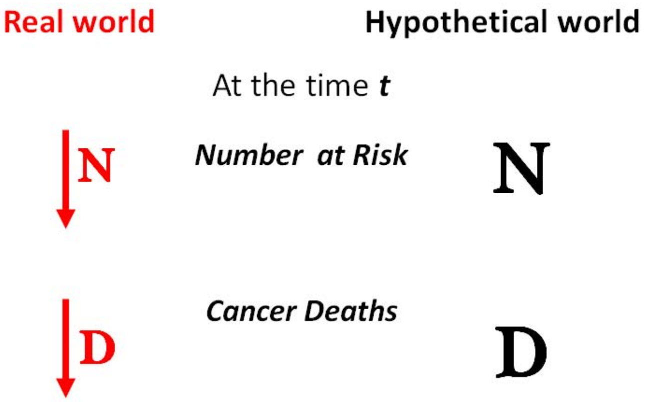
Net Survival

$$NS(t) = S_E(t) = exp(-\int_0^t \lambda_E)$$

Usual Estimators

Cause-specific and relative survival are two estimators of the net survival. They both require that the hazards for cancer and for other causes are independent (conditional on covariates), but this condition is usually not met.

Effect of the Competing Risks



This effect is not uniform being stronger in groups with higher risk to die from competing risks (informative censoring).

Effect of the Competing Risks

Real world		Hypothetical world
	At the time ***t***	
N	***Number at Risk***	N
D	***Cancer Deaths***	D

This effect is not uniform being stronger in groups with higher risk to die from competing risks (informative censoring).

Practical consequences

Simulation

10000 cancer cases were simulated divided in five age-groups. Time of death due to cancer has been generated from an exponential distribution. The effect of age has been simulated by defining an increasing excess hazard ratios for the defined age-groups.

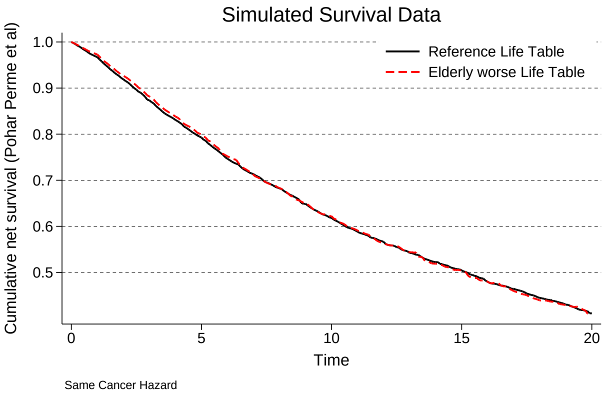
Two times to death from causes other than cancer have been generated from two population life tables, LT-A and LT-B. In both population life tables the probability of dying from causes other than cancer increases with age, but **in LT-B the probabilities of death among elderly are higher (the survival probabilities are worse) than in LT-A.**

Finally we calculated a first overall time to death by taking the minimum of the cancer survival time and the other causes survival time generated from LT-A (**first simulated data**) and a **second overall time to death** by taking the minimum of the cancer survival time and the other causes survival time generated from LT-B (**second simulated data**).

Note that cancer hazard is the same in both data sets. Therefore net survival should not change because it depends only on the cancer hazard.

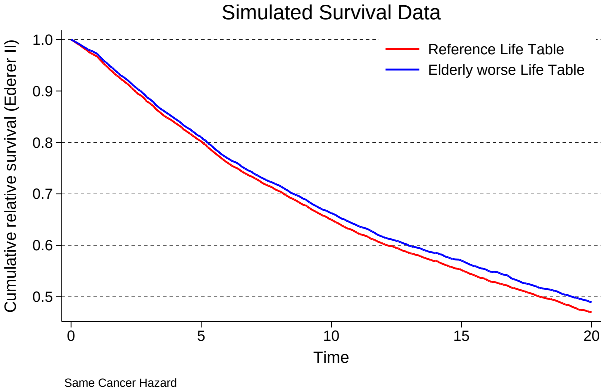
Unbiased new estimator

Looking at the net survival (new estimator) we correctly realize that cancer survival is the same in both data sets.



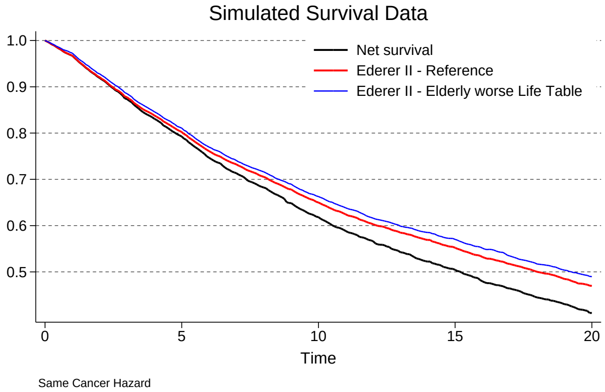
Biased old estimator

Cancer relative survival is apparently improved in the second data set as effect of the worsening of the other causes survival probabilities in elderly people.



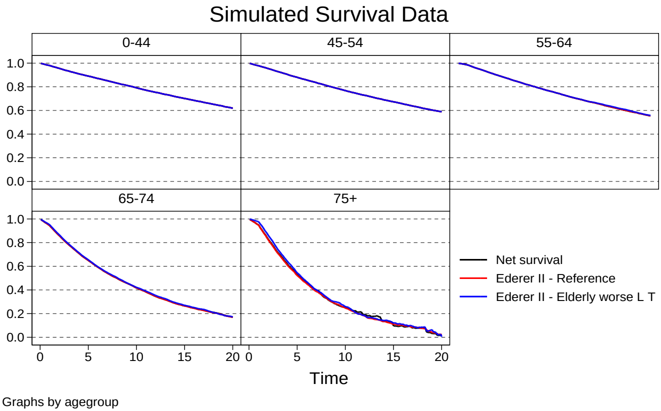
Effect of the informative censoring

When we consider patients of all ages, i.e. patients heterogeneous by age, cancer relative survival is biased towards the survival of the groups with better other causes survival, i.e. towards the survival of the younger patients.



Age-specific estimates

When we consider patients with homogeneous age, i.e. patients within age groups, the differences between the new and the old estimator almost disappear.



Weights

Inverse Probability Weights

$$w_i(t) = \frac{1}{S_{iE}(t)}$$

Real world	Hypothetical world
$N \times w = \mathbf{N}$	$\mathbf{N}$
$D \times w = \mathbf{D}$	$\mathbf{D}$

More on weights

Weights are always greater than 1. Therefore, each individual represents more than one person.

Elderly patients with low expected survival can have large weights. Each of them represents many other individuals died from competing causes.

Large weights cause also large variability of the net survival estimates. Intuitively, we expect large variance if our estimates rely on just a few individuals with large weights. The variance formula of the new estimator includes w^2.

More on weights

Weights are always greater than 1. Therefore, each individual represents more than one person.

Elderly patients with low expected survival can have large weights. Each of them represents many other individuals died from competing causes.

Large weights cause also large variability of the net survival estimates. Intuitively, we expect large variance if our estimates rely on just a few individuals with large weights. The variance formula of the new estimator includes w^2.

Life table approach

In the life table approach we divide the survival time in intervals and compute an interval-specific net survival probability.

Then the cumulative net survival at the end of interval t is the product of the interval-specific net survival up to this time.

Two Stata Commands

- -strs- specifying -pohar- option
- -stnet-

Life table approach

In the life table approach we divide the survival time in intervals and compute an interval-specific net survival probability.

Then the cumulative net survival at the end of interval t is the product of the interval-specific net survival up to this time.

Two Stata Commands

- -strs- specifying -pohar- option
- -stnet-

Life table approach

In the life table approach we divide the survival time in intervals and compute an interval-specific net survival probability.

Then the cumulative net survival at the end of interval t is the product of the interval-specific net survival up to this time.

Two Stata Commands

- **-strs- specifying -pohar- option**
- **-stnet-**

Life table approach

In the life table approach we divide the survival time in intervals and compute an interval-specific net survival probability.

Then the cumulative net survival at the end of interval t is the product of the interval-specific net survival up to this time.

Two Stata Commands

- **-strs- specifying -pohar- option**
- **-stnet-**

Formulae

Two different formulae are applied by `strs`, `pohar` and `stnet`, but they produce net survival estimates very similar.

-strs- weighted actuarial approach

$$NS_i = \frac{1 - \frac{d_i^w}{n_i^w - c_i^w/2}}{exp\left(-\frac{\sum_j^{n_i} \lambda_j^{Pw} - \sum_j^{w_i} \lambda_j^{Pw}/2 - \sum_j^{d_i} \lambda_j^{Pw}/2}{n_i^w - d_i^w/2 - c_i^w/2}\right)}$$

-stnet- weighted hazard transformed

$$NS_i = exp(-(\Lambda_i^{Ow} - \Lambda_i^{Pw})) = exp\left(-k_i \frac{d_i^w - d_i^{Pw}}{y_i^w}\right)$$

Formulae

Two different formulae are applied by `strs, pohar` and `stnet`, but they produce net survival estimates very similar.

-strs- weighted actuarial approach

$$NS_i = \frac{1 - \frac{d_i^w}{n_i^w - c_i^w/2}}{exp\left(-\frac{\sum_j^{n_i} \lambda_j^{Pw} - \sum_j^{w_i} \lambda_j^{Pw}/2 - \sum_j^{d_i} \lambda_j^{Pw}/2}{n_i^w - d_i^w/2 - c_i^w/2}\right)}$$

-stnet- weighted hazard transformed

$$NS_i = exp(-(\Lambda_i^{Ow} - \Lambda_i^{Pw})) = exp\left(-k_i \frac{d_i^w - d_i^{Pw}}{y_i^w}\right)$$

Details

- When net survival estimates are made by using the so-called period or hybrid analysis (see next slides) `strs` and `stnet` apply the same formula (hazard transformed) and net survival estimates they produce match exactly.
- Internally `strs` expands the data set. For each individual as many records are created as the number of the intervals. When the number of cases is large the execution may become slow and memory problems may be encountered.
- `stnet` does not expand the data set. Therefore, it runs faster and without memory problems

Details

- When net survival estimates are made by using the so-called period or hybrid analysis (see next slides) `strs` and `stnet` apply the same formula (hazard transformed) and net survival estimates they produce match exactly.
- Internally `strs` expands the data set. For each individual as many records are created as the number of the intervals. When the number of cases is large the execution may become slow and memory problems may be encountered.
- `stnet` does not expand the data set. Therefore, it runs faster and without memory problems

Details

- When net survival estimates are made by using the so-called period or hybrid analysis (see next slides) `strs` and `stnet` apply the same formula (hazard transformed) and net survival estimates they produce match exactly.
- Internally `strs` expands the data set. For each individual as many records are created as the number of the intervals. When the number of cases is large the execution may become slow and memory problems may be encountered.
- `stnet` does not expand the data set. Therefore, it runs faster and without memory problems

Basic syntax

-stset- data

```
.  use colon_net,clear
(Finnish colon cancer 1975-94, follow-up 1995)

.  stset exit, origin(dx) f(status) scale(365.24)
```

The exit variable contains the exit date from the study and the variable dx contains the date of diagnosis. The timescale must be time since diagnosis in years so we have applied a scale factor of 365.24.

-stnet- syntax

```
.  stnet using popmort, mergeby(_year sex _age) ///
   breaks(0(.08333333)10) diagdate(dx) birthdate(birthdate)///
   listyearly
```

Basic syntax

-stset- data

```
.  use colon_net,clear
(Finnish colon cancer 1975-94, follow-up 1995)

.  stset exit, origin(dx) f(status) scale(365.24)
```

The exit variable contains the exit date from the study and the variable dx contains the date of diagnosis. The timescale must be time since diagnosis in years so we have applied a scale factor of 365.24.

-stnet- syntax

```
.  stnet using popmort, mergeby(_year sex _age) ///
   breaks(0(.08333333)10) diagdate(dx) birthdate(birthdate)///
   listyearly
```

Not Options

```
.  stnet using popmort, ...
```

popmort is the file containing general population survival probabilities.

```
.  stnet .., mergeby(_year sex _age)
```

-mergeby(_year sex _age)- specifies the variables which uniquely determine the records in the popmort file.

```
.  stnet .., breaks(0(.08333333)10)
```

`-breaks(range)-` specifies the cut-points for the life table intervals as a range in the `-forvalues-` command. The units must be years.

Not Options

```
.   stnet using popmort, ...
```

popmort is the file containing general population survival probabilities.

```
.   stnet .., mergeby(_year sex _age)
```

-mergeby(_year sex _age)- specifies the variables which uniquely determine the records in the popmort file.

```
.   stnet .., breaks(0(.08333333)10)
```

`-breaks(range)-` specifies the cut-points for the life table intervals as a range in the `-forvalues-` command. The units must be years.

Not Options

```
.  stnet using popmort, ...
```

popmort is the file containing general population survival probabilities.

```
.  stnet .., mergeby(_year sex _age)
```

-mergeby(_year sex _age)- specifies the variables which uniquely determine the records in the popmort file.

```
.  stnet .., breaks(0(.08333333)10)
```

–`breaks(range)`– specifies the cut-points for the life table intervals as a range in the –`forvalues`– command. The units must be years.

Not Options

```
.   stnet .., diagdate(dx) birthdate(birthdate)
```

The date of diagnosis, variable -dx-, and the date of birth, variable -birthdate-, must also be supplied.

```
.   stnet .., listyearly
```

We have chosen to use one-month intervals to estimate net survival, but the option `listyearly` displays the results only at the end of each year of the follow-up.

Not Options

```
. stnet .., diagdate(dx) birthdate(birthdate)
```

The date of diagnosis, variable -dx-, and the date of birth, variable -birthdate-, must also be supplied.

```
. stnet .., listyearly
```

We have chosen to use one-month intervals to estimate net survival, but the option `listyearly` displays the results only at the end of each year of the follow-up.

Net survival estimator

Cumulative net survival according to Pohar Perme, Stare and Estève method.

start	end	n	d	cns	locns	upcns	secns
.9167	1	2393	56	0.6650	0.6484	0.6811	0.0084
1.917	2	1918	17	0.5682	0.5500	0.5859	0.0091
2.917	3	1677	18	0.5234	0.5043	0.5421	0.0097
3.917	4	1490	12	0.4952	0.4751	0.5150	0.0102
4.917	5	1344	14	0.4709	0.4493	0.4923	0.0110
5.917	6	1232	8	0.4577	0.4343	0.4807	0.0118
6.917	7	1150	7	0.4576	0.4325	0.4824	0.0127
7.917	8	1078	5	0.4623	0.4349	0.4893	0.0139
8.917	9	1010	9	0.4666	0.4358	0.4969	0.0156
9.917	10	936	5	0.4762	0.4415	0.5100	0.0175

Confidence bounds and standard error

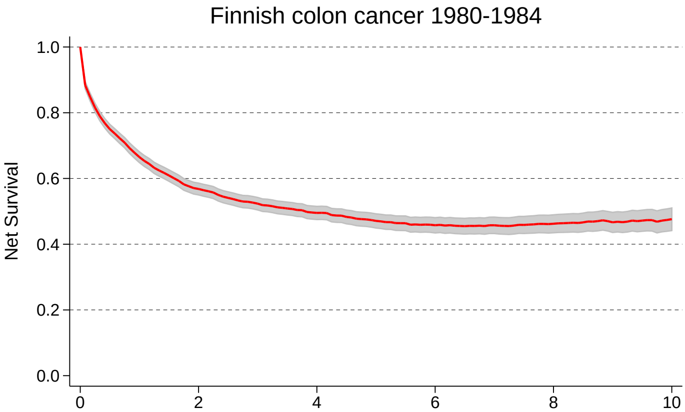
Cumulative net survival according to Pohar Perme, Stare and Estève method.

start	end	n	d	cns	locns	upcns	secns
.9167	1	2393	56	0.6650	0.6484	0.6811	0.0084
1.917	2	1918	17	0.5682	0.5500	0.5859	0.0091
2.917	3	1677	18	0.5234	0.5043	0.5421	0.0097
3.917	4	1490	12	0.4952	0.4751	0.5150	0.0102
4.917	5	1344	14	0.4709	0.4493	0.4923	0.0110
5.917	6	1232	8	0.4577	0.4343	0.4807	0.0118
6.917	7	1150	7	0.4576	0.4325	0.4824	0.0127
7.917	8	1078	5	0.4623	0.4349	0.4893	0.0139
8.917	9	1010	9	0.4666	0.4358	0.4969	0.0156
9.917	10	936	5	0.4762	0.4415	0.5100	0.0175

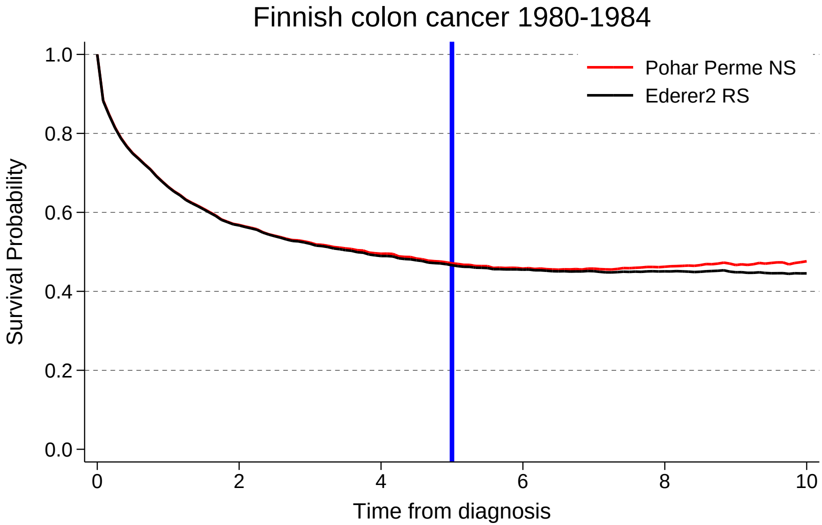
Graph

```
.   stnet ..., saving(colon_results,replace)
```

```
.  use colon_results,clear
.  twoway (rarea locns upcns end, color(gs12))
       (line cns end, ...), ...
```



Pohar Perme and Ederer II



Length of the intervals

The life table approach assumes that the excess hazard is constant within the interval. Therefore net survival estimates may be sensitive to the choice of the length of the interval.

Time Intervals

Interval	5Y-NS	10Y-NS	
One Week	47.12	47.71	
One Month	47.09	47.62	
Three Months	47.04	47.46	
One Year	47.00	46.58	
Time Continuous	47.13	47.52	rs.surv function on R

Length of the intervals

The life table approach assumes that the excess hazard is constant within the interval. Therefore net survival estimates may be sensitive to the choice of the length of the interval.

Time Intervals

Interval	5Y-NS	10Y-NS	
One Week	47.12	47.71	
One Month	47.09	47.62	
Three Months	47.04	47.46	
One Year	47.00	46.58	
Time Continuous	47.13	47.52	rs.surv function on R

Grouped survival times

Sometimes survival times are provided only in months or in years from diagnosis.

Time continuous approach to the estimation of the net survival, developed on the rs.surv function on R and on `stns` on Stata, may be more sensitive to the precision of the survival times than the life table approach .

Precision of Time

	5Y-NS		10Y-NS	
Time in	**stnet**	**rs.surv**	**stnet**	**rs.surv**
Days	47.09	47.13	47.62	47.52
Months	47.09	47.20	47.62	47.87
Years	47.00	47.82	46.58	49.17

Grouped survival times

Sometimes survival times are provided only in months or in years from diagnosis.

Time continuous approach to the estimation of the net survival, developed on the rs.surv function on R and on `stns` on Stata, may be more sensitive to the precision of the survival times than the life table approach .

Precision of Time

	5Y–NS		10Y–NS	
Time in	**stnet**	**rs.surv**	**stnet**	**rs.surv**
Days	**47.09**	**47.13**	**47.62**	**47.52**
Months	**47.09**	**47.20**	**47.62**	**47.87**
Years	**47.00**	**47.82**	**46.58**	**49.17**

Period and Hybrid analysis

To produce more up-to-date survival estimates we can apply a period or an hybrid analysis. Both approaches consider the survival experience of the cancer cases within a time window.

This entails that some patients are observed after their diagnosis (late entry)

The life table approach allows to estimate the Pohar Perme net survival by applying a period or a hybrid analysis. The time-continuous approach currently implemented in available softwares does not allow late entry. Therefore, period and hybrid analysis cannot be performed.

Period and Hybrid analysis

To produce more up-to-date survival estimates we can apply a period or an hybrid analysis. Both approaches consider the survival experience of the cancer cases within a time window.

This entails that some patients are observed after their diagnosis (late entry)

The life table approach allows to estimate the Pohar Perme net survival by applying a period or a hybrid analysis. The time-continuous approach currently implemented in available softwares does not allow late entry. Therefore, period and hybrid analysis cannot be performed.

Period Window

Anni Diagnosi	Anni Follow-up																				
	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
1975	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1976		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1977			0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1978				0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1979					0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1980						0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1981							0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1982								0	1	2	3	4	5	6	7	8	9	10	11	12	13
1983									0	1	2	3	4	5	6	7	8	9	10	11	12
1984										0	1	2	3	4	5	6	7	8	9	10	11
1985											0	1	2	3	4	5	6	7	8	9	10
1986												0	1	2	3	4	5	6	7	8	9
1987													0	1	2	3	4	5	6	7	8
1988														0	1	2	3	4	5	6	7
1989															0	1	2	3	4	5	6
1990																0	1	2	3	4	5
1991																	0	1	2	3	4
1992																		0	1	2	3
1993																			0	1	2
1994																				0	1

Period Analysis

```
. stset exit, origin(dx) failure(status) scale(365.24) ///
      enter(time mdy(1,1,1990)) exit(time mdy(12,31,1994))
```

Hybrid Window

Anni Diagnosi	Anni Follow-up																				
	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
1975	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1976		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1977			0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1978				0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1979					0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1980						0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1981							0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1982								0	1	2	3	4	5	6	7	8	9	10	11	12	13
1983									0	1	2	3	4	5	6	7	8	9	10	11	12
1984										0	1	2	3	4	5	6	7	8	9	10	11
1985											0	1	2	3	4	5	6	7	8	9	10
1986												0	1	2	3	4	5	6	7	8	9
1987													0	1	2	3	4	5	6	7	8
1988														0	1	2	3	4	5	6	7
1989															0	1	2	3	4	5	6
1990																0	1	2	3	4	5
1991																	0	1	2	3	4
1992																		0	1	2	3
1993																			0	1	2
1994																				0	1

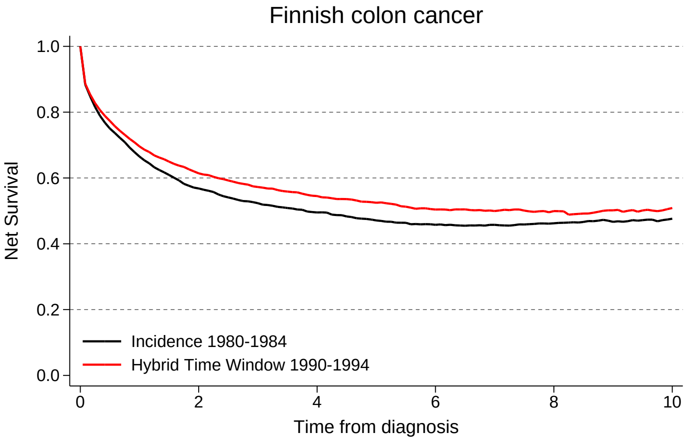
Hybrid Analysis

```
. g long hybridtime = cond(yydx>1989, dx, mdy(1,1,1991))

. stset exit, origin(dx) failure(status) scale(365.24) ///
      enter(time hybridtime)
```

Most up-to-date net survival estimates

We can then apply `stnet` in the usual manner to obtain net survival estimates



Age standardization

Net survival depends on age of patients. Therefore, age-standardization is required to compare net survival across populations or over time.

Deriving age-standardised net survival estimates by means of `strs` or `stnet` is straightforward. We first generate age groups and weights:

```
. egen agegr =cut(age), at(0 45(10)75 100) icodes
. recode agegr 0=0.07 1=0.12 2=0.23 3=0.29 4=0.29, gen(standw)
```

Then, age-standardised NS estimates are directly produced

```
. stnet using popmort [iw=standw], mergeby(_year sex _age) ///
        br(0(.083333333)10) diagdate(dx) birthdate(birthdate) ///
        standstrata(agegr) by(sex)
```

Age standardization

Net survival depends on age of patients. Therefore, age-standardization is required to compare net survival across populations or over time.

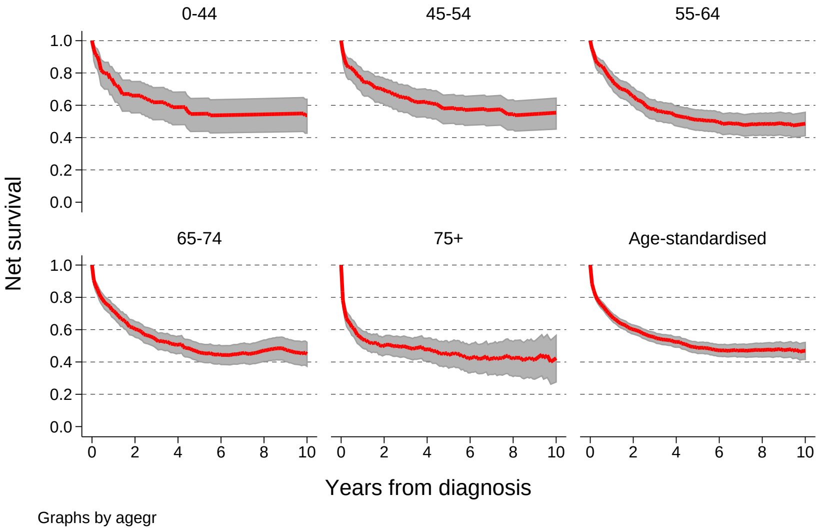
Deriving age-standardised net survival estimates by means of `strs` or `stnet` is straightforward. We first generate age groups and weights:

```
. egen agegr =cut(age), at(0 45(10)75 100) icodes
. recode agegr 0=0.07 1=0.12 2=0.23 3=0.29 4=0.29, gen(standw)
```

Then, age-standardised NS estimates are directly produced

```
. stnet using popmort [iw=standw], mergeby(_year sex _age) ///
       br(0(.083333333)10) diagdate(dx) birthdate(birthdate) ///
       standstrata(agegr) by(sex)
```

Net survival by age and age-standardized



What's New

Pohar Perme et al., by developing an unbiased estimator of net survival, significantly advanced the field of estimating net survival of cancer patients in a relative survival framework. Their approach was developed for continuous survival times and implemented in R (*rs.surv*) and more recently in Stata (`stns`).

We have adapted the approach to discrete survival times and hope the new Stata commands, `stnet` and `strs,pohar` will enable users to easily compute this unbiased net survival estimator. In particular, we hope that it will be useful for survival analysis by population cancer registries.

What's New

Pohar Perme et al., by developing an unbiased estimator of net survival, significantly advanced the field of estimating net survival of cancer patients in a relative survival framework. Their approach was developed for continuous survival times and implemented in R (*rs.surv*) and more recently in Stata (`stns`).

We have adapted the approach to discrete survival times and hope the new Stata commands, `stnet` and `strs,pohar` will enable users to easily compute this unbiased net survival estimator. In particular, we hope that it will be useful for survival analysis by population cancer registries.

Under submission

The Stata Journal (*yyyy*) *vv*, Number *ii*, pp. 1–13

Estimating net survival using a life table approach

Enzo Coviello
ASL BT
Barletta, Italy
enzo.coviello@tin.it

Paul W. Dickman
Karolinska Institutet
Stockholm, Sweden
paul.dickman@ki.se

Karri Seppä
Finnish Cancer Registry
Helsinki, Finland
karri.seppa@cancer.fi

Arun Pokhrel
Finnish Cancer Registry
Helsinki, Finland
arun.pokhrel@cancer.fi

Abstract. Cancer registries are often interested in estimating net survival, the probability of survival if the cancer under study is the only possible cause of death. In 2011 Pohar Perme et al. proposed a new estimator of net survival based on an inverse probability weighting. They demonstrated that existing estimators of net survival based on relative survival were biased, whereas the new estimator was unbiased. However, the Pohar Perme estimator was developed for continuous survival times yet cancer registries often only have discrete survival times (e.g., survival time in completed months or completed years). We propose an approach to estimation when survival times are discrete and life table estimation is applied. This article describes the command `stnet` for life table estimation of net survival, adapting the Pohar Perme approach to life table estimation. Age-standardised survival estimates are available. In addition to traditional cohort/complete approach, estimates can also be made using also a period or hybrid approach.

Web resources

`strs` is available on Paul Dickman web site on

`http://www.pauldickman.com`

and can be installed by typing on the command window:

```
net install http://www.pauldickman.com/rsmodel/stata_colon/strs
```

`stnet` can be downloaded from the SSC Archive by typing :

```
ssc install stnet
```


Thanks
38