The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go "beyond the Stata manual" in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

The *Stata Journal* is indexed and abstracted by *CompuMath Citation Index*, *Current Contents/Social and Behavioral Sciences*, *RePEc: Research Papers in Economics*, *Science Citation Index Expanded* (also known as *SciSearch*), *Scopus*, and *Social Sciences Citation Index*.

For more information on the *Stata Journal*, including information for authors, see the webpage

http://www.stata-journal.com

Subscription rates listed below include both a printed and an electronic copy unless otherwise mentioned.

U.S. and Canada		Elsewhere	
Printed & electronic		**Printed & electronic**	
1-year subscription	$115	1-year subscription	$145
2-year subscription	$210	2-year subscription	$270
3-year subscription	$285	3-year subscription	$375
1-year student subscription	$ 85	1-year student subscription	$115
1-year institutional subscription	$345	1-year institutional subscription	$375
2-year institutional subscription	$625	2-year institutional subscription	$685
3-year institutional subscription	$875	3-year institutional subscription	$965
Electronic only		**Electronic only**	
1-year subscription	$ 85	1-year subscription	$ 85
2-year subscription	$155	2-year subscription	$155
3-year subscription	$215	3-year subscription	$215
1-year student subscription	$ 55	1-year student subscription	$ 55

Back issues of the *Stata Journal* may be ordered online at

http://www.stata.com/bookstore/sjj.html

Individual articles three or more years old may be accessed online without charge. More recent articles may be ordered online.

http://www.stata-journal.com/archives.html

The *Stata Journal* is published quarterly by the Stata Press, College Station, Texas, USA.

Address changes should be sent to the *Stata Journal*, StataCorp, 4905 Lakeway Drive, College Station, TX 77845, USA, or emailed to sj@stata.com.

Copyright © 2015 by StataCorp LP

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
University of Alberta
Edmonton, Canada
apokhrel@ualberta.ca

Abstract. Cancer registries are often interested in estimating net survival (NS), the probability of survival if the cancer under study is the only possible cause of death. Pohar Perme, Stare, and Estéve (2012, *Biometrics* 68: 113–120) proposed a new estimator of NS based on inverse probability weighting. They demonstrated that existing estimators of NS based on relative survival were biased, whereas the new estimator was unbiased. The new estimator was developed for continuous survival times, yet cancer registries often have only discrete survival times (for example, survival time in completed months or years). Therefore, we propose an approach to estimation for when survival times are discrete. In this article, we describe the stnet command for life-table estimation of NS, adapting the Pohar Perme estimation approach to life-table estimation. Estimates can be made using a period or hybrid approach in addition to the traditional cohort (or complete) approach, and age-standardized survival estimates are available.

Keywords: st0375, stnet, net survival, relative survival, competing risks, life table, cancer survival, age standardization, period analysis

1 Introduction

Cancer registries are often interested in net survival (NS), the probability of survival when the cancer under study is the only possible cause of death. Although NS is defined under a hypothetical condition (Tsiatis 2005), it is useful in practice for comparing cancer survival between countries (Coleman et al. 2008) and over time (Dickman and Adami 2006) because it is independent of mortality due to other diseases that varies between countries and over time.

NS is typically estimated using one of two methods: cause-specific survival, which relies on accurate information on cause of death, and relative survival (Estève et al. 1990). Relative-survival estimates NS by contrasting the all-cause mortality of the patients to the expected mortality of a comparable group from the general population (that is, a group with the same demographics of the patients). This method is preferred by the cancer registries because it does not require the cause of death, which may be unavail-

able or unreliable, and it captures mortality that is indirectly caused by the cancer, such as mortality due to treatment side effects. On the other hand, relative survival assumes that the expected mortality of the general population correctly reflects the mortality from causes other than cancer in the patients (Sarfati, Blakely, and Pearce 2010).

Pohar Perme, Stare, and Estève (2012) showed that none of the available methods actually estimate NS in a relative-survival setting because the methods do not account for the effect of competing causes of mortality. They proposed a new estimator[1] (often called the Pohar Perme estimator), which was shown in a simulation study to produce unbiased estimates of NS (Danieli et al. 2012), provided that patients have the same survival experience as the general population if they did not have cancer (Rachet et al. 2010; Hinchliffe et al. 2012).

The Pohar Perme estimator was developed for continuous survival times; however, cancer registries often have only discrete survival times (for example, survival time in completed months or years). The continuous survival time approach forces NS to be estimated using the cumulative expected survival computed at each event time, which may result in some approximation. Furthermore, continuous survival time and actuarial methods differ in their handling of individuals with identical survival times: the former assumes that all censored individuals were at risk at their time of death, whereas the latter assumes that only half the individuals were at risk at their time of death. As a result, when there are ties between censored and observed survival times, the continuous survival time approach tends to overestimate survival. Therefore, actuarial estimates are usually preferred (Therneau and Grambsh 2000).

We propose an actuarial approach to the estimation of NS that adapts the Pohar Perme estimation method to life-table estimation. In this article, we describe the stnet command for life-table estimation of NS.

2 Methods

For computing NS, the Pohar Perme estimation method weights[2] the contribution of each individual according to the inverse of his or her cumulative expected survival probability. As a consequence of the inverse probability weighting, the number at risk at time t becomes the sum of the weights of the individuals at risk, and the number of deaths becomes the sum of the weights of individuals dying at that time. Intuitively, the effect of the weights is to increase the number of people at risk and the number of deaths to account for person-time and deaths not observed as a result of mortality due to competing causes.

---

1. The estimator has been implemented in the rs.surv function, available in the relsurv package of R software (Pohar and Stare 2006; Pohar Perme 2013), and more recently in the user-written stns command for Stata (Clerc-Urmés, Grzebyk, and Hedelin 2014).
2. All information for the weights can be obtained from the population mortality tables.

In our approach, the interval-specific NS is estimated by transforming the estimated cumulative weighted excess hazard, Λ_{ei}^w; the difference of the cumulative weighted observed hazard, Λ_i^w; and the cumulative weighted expected hazard, Λ_i^{*w}:

$$\text{NS}_i = \exp(-\Lambda_{ei}^w) = \exp\{-(\Lambda_i^w - \Lambda_i^{*w})\} = \exp\left(-k_i\frac{d_i^w - d_i^{*w}}{y_i^w}\right)$$

where d_i^w is the weighted number of deaths in the interval, d_i^{*w} is the weighted number of expected deaths in a comparable general-population group, y_i^w is the weighted time at risk, and k_i is the length of the interval. The weighted quantities d_i^w, d_i^{*w}, and y_i^w are the sums of interval-specific observations of n patients weighted by the inverse of their cumulative expected survival probabilities; that is,

$$d_i^w = \sum_{j=1}^n \frac{d_{ij}}{S_{ij}^*}, \quad d_i^{*w} = \sum_{j=1}^n \frac{-\log(p_{ij}^*)y_{ij}}{S_{ij}^*}, \quad \text{and} \quad y_i^w = \sum_{j=1}^n \frac{y_{ij}}{S_{ij}^*}$$

where d_{ij} is the indicator of death, p_{ij}^* is the expected probability of surviving one year, y_{ij} is the time at risk for patient j in interval i, and the cumulative expected survival probability S_{ij}^* is computed at the midpoint of interval i.

The cumulative NS at the end of interval t is the product of the interval-specific NS probabilities:

$$\text{NS}(t) = \prod_{i=1}^t \text{NS}_i$$

The variance of the cumulative NS is approximated by the delta method $\text{Var}\{\text{NS}(t)\} = \text{NS}(t)^2\text{Var}\{\Lambda_e^w(t)\}$, where the variance of the cumulative weighted excess hazard, $\Lambda_e^w(t) = \sum_{i=1}^t \Lambda_{ei}^w$, at the end of interval t is

$$\text{Var}\{\Lambda_e^w(t)\} = \sum_{i=1}^t k_i^2 \frac{\sum_{j=1}^n \frac{d_{ij}}{S_{ij}^{*2}}}{(y_i^w)^2}$$

Confidence bounds are then computed after a log-log transformation. Because this approach assumes that the hazard and weights are constant within the interval, stnet is sensitive to the choice of interval length. In practice, an interval length of one month is a safe choice; in our experience using this interval, we have never found differences of any practical significance when survival times were recorded in days and results were also obtained after specifying shorter time intervals or using the continuous survival time estimator. The same interval length may also be applied when survival times are grouped in months or larger time units. However, note that continuous survival time estimates do not constitute a good reference in this case (see section 4).

3   The stnet command

In general, two data files are required to estimate NS in a relative-survival setting: a file containing individual-level data on the patients and a file containing expected probabilities of death for a comparable group from the general population (the popmort.dta file;

see section 3.3). The stnet command is for use with survival-time (st) data; therefore, the patient data file must be stset (see [ST] **stset**) with time since entry in years as the timescale before using stnet.

3.1 Syntax

stnet <u>u</u>sing *filename* $\begin{bmatrix} if \end{bmatrix}$ $\begin{bmatrix} in \end{bmatrix}$ $\begin{bmatrix} weight \end{bmatrix}$, <u>br</u>eaks(*range*) <u>mer</u>geby(*varlist*)
   <u>diag</u>date(*varname*) <u>birth</u>date(*varname*) $\begin{bmatrix} by(varlist) \end{bmatrix}$ attage(*newvar*)
   attyear(*newvar*) <u>surv</u>prob(*varname*) maxage(#) <u>ederer2</u>
   <u>stand</u>strata(*varname*) <u>list</u>(*varlist*) listyearly <u>f</u>ormat(%*fmt*) <u>notables</u>
   <u>level</u>(#) saving(*filename*$\begin{bmatrix} , replace \end{bmatrix}$) <u>savst</u>and(*filename*$\begin{bmatrix} , replace \end{bmatrix}$) $\big]$

using *filename* specifies a file containing general-population survival probabilities (see section 3.3). Importance weights (iweights) are allowed and should be specified as proportions; they produce age-standardized estimates (see the example in section 4.1).

3.2 Options

breaks(*range*) specifies the cutpoints for the life-table intervals as a *range* in the forvalues command (see [P] **forvalues**). The units must be years; for example, specify breaks(0(0.08333)5) for monthly intervals up to five years. breaks() is required.

mergeby(*varlist*) specifies the variables that uniquely determine the records in the file of general-population survival probabilities (the using file, also known as the popmort.dta file). The using file must also be sorted by these variables. mergeby() is required.

diagdate(*varname*) specifies the variable containing the date of diagnosis. diagdate() is required.

birthdate(*varname*) specifies the variable containing the date of birth. birthdate() is required.

by(*varlist*) specifies the life-table stratification variables. One life table is estimated for each combination of these variables.

attage(*newvar*) specifies the variable containing attained age (that is, age at the time of follow-up). This variable cannot exist in the patient data file (it is created as the integer part of age at diagnosis plus follow-up time) but must exist in the using file. The default is attage(_age).

attyear(*newvar*) specifies the variable containing attained calendar year (that is, calendar year at the time of follow-up). This variable cannot exist in the patient data file (it is created as the integer part of year of diagnosis plus follow-up time) but must exist in the using file. The default is attyear(_year).

survprob(*varname*) specifies the variable in the using file that contains the general-population survival probabilities. The default is survprob(prob).

maxage(#) specifies the maximum age for which general-population survival probabilities are provided in the using file. Probabilities for individuals older than this value are assumed to be the same as for the maximum age. The default is maxage(99).

ederer2 specifies that Ederer II relative-survival estimates be calculated. Note that stnet calculates the observed survival by transforming the interval-specific cumulative hazard; therefore, the results are not exactly equal to those obtained by using ltable (Dickman 2010; see [ST] **ltable**).

standstrata(*varname*) specifies a variable defining strata across which to average the cumulative survival estimate. With this option, a *weight* must also be specified as follows: [iweight=*varname*].

list(*varlist*) specifies the variables to be listed in the life table. The variables start and end are included by default; however, if only one of these is specified in the list() option, then the other is suppressed.

listyearly displays life-table survival estimates only at the end of each year of follow-up.

format(%*fmt*) specifies the format for variables containing survival estimates. The default is format(%6.4f).

notables suppresses display of the life tables.

level(#) sets the confidence level based on the value of global macro S_level. The default is level(95).

saving(*filename*[, replace]) specifies to save in *filename* a dataset containing one observation for each life-table interval.

savstand(*filename*[, replace]) specifies to save standardized estimates in *filename*.

3.3 The population mortality file

The population mortality file (typically named popmort.dta) must contain general-population survival probabilities (conditional probabilities of surviving one year) stratified by those variables which uniquely determine the records and upon which it is assumed expected survival depends. Typically, those variables are age, sex, and period, but further variables may be included, such as race, region of residence, or social class (Coleman et al. 1999). Such probabilities (or corresponding rates that can be transformed to probabilities) are available from the Human Mortality Database[3] for many populations or can be obtained from local government authorities (typically the central statistics office). The filename is specified via the using option, and the variables by which the file is sorted are specified using the mergeby(*varlist*) option. The following is a listing of the first five rows of the Finnish popmort.dta file:

---

3. See http://www.mortality.org/.

```
. use popmort
. list sex _year _age prob in 1/5, noobs
```

sex	_year	_age	prob
1	1951	0	.96429
1	1951	1	.99639
1	1951	2	.99783
1	1951	3	.99842
1	1951	4	.99882

Probabilities must be provided for every year that the patients will attain during follow-up; if data are not available for recent years, it is standard practice to assume the probabilities are the same as those most recently available. When estimating expected survival, we require the expected probabilities of death according to age and year at time of follow-up (rather than time of diagnosis). The command must, therefore, keep track of both.

The patient data file must contain variables for date of diagnosis (`diagdate()`) and date of birth (`birthdate()`). Internally, `stnet` computes age and calendar year at diagnosis. These variables are then updated with follow-up. By default, these "time-updated" variables are named `_age` and `_year`, but alternative names can be specified using the appropriate options. `popmort.dta` must contain variables `_age` and `_year` (or as otherwise specified) because the expected survival probabilities are merged using these variables.

4   Example

Here we illustrate the command using data provided by the Finnish Cancer Registry on patients diagnosed with colon carcinoma in Finland, 1975–1994. In the original data, the diagnosis and censoring dates were recorded only to the nearest month, so we added a random day to these dates. Date of birth has also been randomly generated from the recorded age. We first estimate NS in patients diagnosed from 1980–1984 up to 10 years from diagnosis.

```
. use colon_net
(Colon carcinoma, all stages, Finland 1975-94, follow-up to 1995. Date of diagno)

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

. stnet using popmort if yydx>=1980 & yydx<1985, mergeby(_year  sex  _age)
> breaks(0(.083333333)10) diagdate(dx) birthdate(birthdate)
> list(n d cns locns upcns secns) listyearly

          failure _d:  status
     analysis time _t:  (exit-origin)/365.24
               origin:  time dx

Cumulative net survival according to Pohar Perme, Stare and Estéve 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

The variable `exit` 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. We have chosen to use one-month intervals to estimate NS, but the option `listyearly` displays the results only at the end of each year of follow-up. Variables containing dates of diagnosis (`dx`) and dates of birth (`birthdate`) must also be supplied to `stnet`. The Pohar Perme NS estimate is listed in the column labeled `cns`, log-log confidence bounds are in the columns `locns` and `upcns`, and standard errors are in the column `secns`.

Sometimes survival times are provided only in months or years from diagnosis. In the `colon_net` data, we can mimic these situations by creating new variables with survival times in months and years.

```
. generate surv_mm = floor((exit-dx)/365.24*12)+.5
. generate surv_yy = floor((exit-dx)/365.24)+.5
```

We can use the survival time in months after

```
. quietly stset surv_mm, failure(status) scale(12)
```

and use the survival time in years after

```
. quietly stset surv_yy, failure(status)
```

We can then compute NS estimates in the original data, where time is recorded in days, and in the derived data, where time is recorded in months and years, by applying time-continuous (`stns`) and life-table approaches (`stnet`).

In the estimation of NS, note that the time-continuous approach is forced to use cumulative expected survival computed at each event time (Clerc-Urmès, Grzebyk, and Hedelin 2014), yet the life-table approach computes cumulative expected survival at the specified time intervals. Therefore, we can continue to use time intervals in months when survival times are provided in years, thus resulting in more accurate estimates of the cumulative expected survival.

Table 1. NS estimates using time-continuous and life-table approaches with survival times recorded in days, months, and years

| Years | Approach | | Time | |
		Days	Months	Years
5	Life table (stnet)	0.4709	0.4709	0.4752
	Time cont. (stns)	0.4713	0.4721	0.4784
10	Life table (stnet)	0.4762	0.4762	0.4807
	Time cont. (stns)	0.4754	0.4790	0.4921

Table 1 shows that in the derived data, the time-continuous approach overestimates the survival in the original data mostly when time is recorded in years. The life-table approach yields the same estimates when times are grouped in months and shows smaller differences with the original data when times are recorded in years. When survival times are grouped because of rounded survival times, the life-table approach is less sensitive than the time-continuous approach to the precision of survival times (Seppä, Hakulinen, and Pokhrel Forthcoming).

4.1 Example 2: Age-standardized NS

NS generally depends on age at diagnosis, and the age distribution of cancer patients varies over time and across populations. As such, age standardization is recommended for comparing estimates between different populations or over time within the same population (Pokhrel and Hakulinen 2009). Here we derive age-standardized NS estimates using the Corazziari, Quinn, and Capocaccia (2004) standard cancer patient population number 1 as the standard population. We use the by() option together with standstrata() to produce age-standardized estimates for each sex.

```
. 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, generate(standw)
(15564 differences between agegr and standw)

. stnet using popmort if yydx>=1980 & yydx<1985 [iw=standw],
> mergeby(_year  sex  _age) breaks(0(.083333333)10)
> diagdate(dx) birthdate(birthdate) notables noshow
> standstrata(agegr) listyearly by(sex) savstand(agestand_sex_NS,replace)
file agestand_sex_NS.dta saved
```

Specifying the `standstrata()` option results in `stnet` first producing stratified life tables for each level of the variables specified in `standstrata()` and then producing standardized estimates using the weights contained in the variable specified as `iweight`. (The weights should be specified as proportions.) The `notables` option requests suppression of output, and the `savstand()` option saves life-table survival estimates in the file agestand_sex_NS.

The following command produces figure 1, which illustrates the age-standardized NS and confidence bounds for each sex.

```
. use agestand_sex_NS, clear

. twoway (rarea locns upcns end, col(gs10))
> (line cns end, lc(black) lw(medthick) lp(l)),
> by(sex, legend(off)) xlabel(0(2)10) xtitle("Years from diagnosis")
> ytitle("Net survival") ylabel(0(.2)1, format(%2.1f))
```
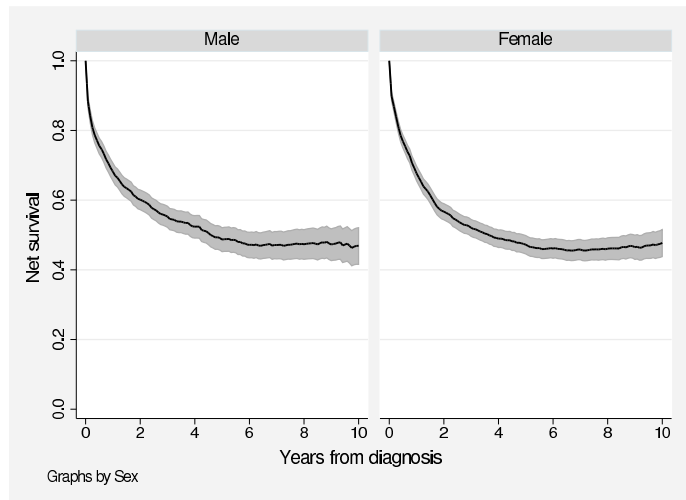

Figure 1. NS by sex for colon cancer cases diagnosed in Finland from 1980–1984

4.2   Example 3: Period and hybrid estimation

To produce more up-to-date survival estimates, Brenner and associates suggested that life-table survival estimates be made using a period analysis (Brenner et al. 2004; Brenner and Gefeller 1996). The fundamental characteristic of the period analysis is that

time at risk is left-truncated at the start of the period window and right-censored at the end.

Our approach to period estimation using Stata is to first identify the time at risk during the period window for each individual by applying stset with calendar time as the timescale. For example, we might be interested in the period between 1 January 1990 and 31 December 1994 (the last five years for which incidence data were collected in this dataset).

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

We can then apply stnet to obtain NS estimates according to the period analysis.

```
. stnet using popmort, mergeby(_year  sex  _age) breaks(0(.083333333)10)
> diagdate(dx) birthdate(birthdate) listyearly nosh
```

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

start	end	n	d	cns	locns	upcns	secns
.9167	1	3188	73	0.6910	0.6766	0.7050	0.0072
1.917	2	2535	33	0.6013	0.5853	0.6170	0.0081
2.917	3	2126	17	0.5608	0.5435	0.5776	0.0087
3.917	4	1878	13	0.5321	0.5136	0.5502	0.0094
4.917	5	1624	11	0.5105	0.4907	0.5298	0.0100
5.917	6	1404	11	0.4862	0.4639	0.5082	0.0113
6.917	7	1244	5	0.4848	0.4604	0.5087	0.0123
7.917	8	1162	2	0.4862	0.4588	0.5130	0.0138
8.917	9	1054	6	0.4848	0.4522	0.5165	0.0164
9.917	10	947	4	0.4905	0.4539	0.5260	0.0184

Application of this approach may be problematic if the follow-up period extends beyond the period for which incident cases are accrued. For example, our sample dataset contains patients diagnosed until December 1994 with follow-up until December 1995. For this reason, we censored the follow-up of all individuals on 31 December 1994 in the previous example. What would we do if we wanted to perform period analysis with a window from 1 January 1991 to 31 December 1995? In such a situation, Brenner and Rachet (2004) suggested that the period window be widened for the first year (it should be 1 January 1990 to 31 December 1995 so that patients diagnosed in 1989–1994 will contribute person-time). They called this approach hybrid analysis.

The distinctive feature of this approach is that the date at which individuals become at risk (the start of the period window) differs according to year of diagnosis. This is relatively easy to apply in Stata:

```
. generate long hybridtime = cond(yydx>1989, dx, mdy(1,1,1991))
. stset exit, origin(dx) failure(status) enter(time hybridtime)
> scale(365.24)
. stnet using popmort, mergeby(_year  sex  _age)
> breaks(0(.08333333)10) diagdate(dx) birthdate(birthdate)
```

We create the new variable `hybridtime` to hold the date at which each individual becomes at risk. This corresponds to the date of diagnosis for patients diagnosed in 1990–1994 and corresponds to 1 January 1991 for patients diagnosed before 1 January 1990.[4] We then `stset` the data with `hybridtime` as the start of the time at risk (using the `enter()` option) and call `stnet` in the usual manner.

5 Conclusion

Pohar Perme, Stare, and Estève (2012), by developing an unbiased estimator of NS, significantly advanced the field of estimating NS of cancer patients in a relative-survival framework. Their approach was developed for continuous survival times. We have adapted an actuarial approach to the estimation of NS that is helpful in cases of grouped survival times. The new command, `stnet`, will enable users to easily compute this unbiased estimator, directly obtain age-standardized NS, and produce more up-to-date survival estimates by using period or hybrid analysis. In particular, we hope that `stnet` will be useful for survival analysis by population cancer registries.

6 Acknowledgments

7 References

Brenner, H., V. Arndt, O. Gefeller, and T. Hakulinen. 2004. An alternative approach to age adjustment of cancer survival rates. *European Journal of Cancer* 40: 2317–2322.

Brenner, H., and O. Gefeller. 1996. An alternative approach to monitoring cancer patient survival. *Cancer* 78: 2004–2010.

Brenner, H., and B. Rachet. 2004. Hybrid analysis for up-to-date long-term survival rates in cancer registries with delayed recording of incident cases. *European Journal of Cancer* 40: 2494–2501.

Clerc-Urmès, K., M. Grzebyk, and G. Hedelin. 2014. Net survival estimation with stns. *Stata Journal* 14: 87–102.

Coleman, M. P., P. Babb, P. Damiecki, P. Grosclaude, S. Honjo, J. Jones, G. Knerer, A. Pitard, M. J. Quinn, A. Sloggett, and B. De Stavola. 1999. Cancer survival trends in England and Wales, 1971–1995: Deprivation and NHS Region. Office for National Statistics, London, UK.

---

4. A diagram such as the one used in Brenner and Rachet (2004) can assist in defining the entry dates.

Coleman, M. P., M. Quaresma, F. Berrino, J. M. Lutz, R. D. Angelis, R. Capocaccia, P. Baili, B. Rachet, G. Gatta, T. Hakulinen, A. Micheli, M. Sant, H. K. Weir, J. M. Elwood, H. Tsukama, S. Koifman, G. A. E. Silva, S. Francisci, M. Santguilani, A. Verdecchia, H. H. Storm, and J. C. Young. 2008. Cancer survival in five continents: A worldwide population-based study (CONCORD). *Lancet Oncology* 9: 730–756.

Corazziari, I., M. Quinn, and R. Capocaccia. 2004. Standard cancer patient population for age standardising survival ratios. *European Journal of Cancer* 40: 2307–2316.

Danieli, C., L. Remontet, N. Bossard, L. Roche, and A. Belot. 2012. Estimating net survival: The importance of allowing for informative censoring. *Statistics in Medicine* 31: 775–786.

Dickman, P. W. 2010. Standard errors of observed and relative survival in strs. http://www.pauldickman.com/rsmodel/stata_colon/standard_errors.pdf.

Dickman, P. W., and H. O. Adami. 2006. Interpreting trends in cancer patient survival. *Journal of Internal Medicine* 260: 103–117.

Estève, J., E. Benhamou, M. Croasdale, and L. Raymond. 1990. Relative survival and the estimation of net survival: Elements for further discussion. *Statistics in Medicine* 9: 529–538.

Hinchliffe, S. W., M. J. Rutherford, M. J. Crowther, C. P. Nelson, and P. C. Lambert. 2012. Should relative survival be used with lung cancer data? *British Journal of Cancer* 106: 1854–1859.

Pohar, M., and J. Stare. 2006. Relative survival analysis in R. *Computer Methods and Programs in Biomedicine* 81: 272–278.

Pohar Perme, M. 2013. relsurv: Relative survival. R. package version 2.0-4. http://CRAN.R-project.org/package=relsurv.

Pohar Perme, M., J. Stare, and J. Estève. 2012. On estimation in relative survival. *Biometrics* 68: 113–120.

Pokhrel, A., and T. Hakulinen. 2009. Age-standardisation of relative survival ratios of cancer patients in a comparison between countries, genders and time periods. *European Journal of Cancer* 45: 642–647.

Rachet, B., L. Ellis, C. Maringe, T. Chu, M. Quaresma, A. Shah, S. Walters, L. Woods, D. Forman, and M. P. Coleman. 2010. Socioeconomic inequalities in cancer survival in England after the NHS cancer plan. *British Journal of Cancer* 103: 446–453.

Sarfati, D., T. Blakely, and N. Pearce. 2010. Measuring cancer survival in populations: Relative survival vs. cancer-specific survival. *International Journal of Epidemiology* 39: 598–610.

Seppä, K., T. Hakulinen, and A. Pokhrel. Forthcoming. Choosing the net survival method for cancer survival estimation. *European Journal of Cancer*.

Therneau, T. M., and P. Grambsh. 2000. *Modeling Survival Data: Extending the Cox Model.* New York: Springer.

Tsiatis, A. A. 2005. Competing Risks. In *Encyclopedia of Biostatistics*, ed. P. Armitage and T. Colton, 2nd ed., 1025–1035. Chichester, UK: Wiley.

About the authors

Enzo Coviello is an epidemiologist in the Unit of Statistics and Epidemiology at ASL BT in Barletta, Italy. He is a longtime Stata user and enthusiast as well as the author of some popular user-written Stata commands, including `stcascoh`, `stcompet`, `strs`, and `distrate`. His main interest is in the analysis of population-based cancer registries data.

Paul Dickman is a professor of biostatistics at Karolinska Institutet in Stockholm, Sweden. He conducts research in population-based epidemiology, with a particular focus on cancer epidemiology. His main interest is in the development and application of statistical methods for studying the survival of cancer patients. Stata is his primary teaching tool, and he is coauthor of the `strs` command.

Karri Seppä is a statistician at the Finnish Cancer Registry in Helsinki, Finland. His main interest is in the development of statistical methods for population-based cancer survival analysis.

Arun Pokhrel is a statistician at the Women & Children's Health Research Institute at the University of Alberta in Edmonton, Canada. His main interest is in the development and application of statistical methods for studying the survival of cancer patients.