An alternative approach to age adjustment of survival rates

by

Hermann Brenner, Volker Arndt, Olaf Gefeller, Timo Hakulinen

a Dept. of Epidemiology, German Centre for Research on Ageing, Bergheimer Str. 20, D-69115 Heidelberg, Germany

b Department of Medical Informatics, Biometry and Epidemiology, University of Erlangen-Nuremberg, Waldstrasse 6, D-91504 Erlangen, Germany

c Finnish Cancer Registry, Liisankatu 21 B, FIN-00170 Helsinki, Finland

d Dept. of Public Health, University of Helsinki, FIN-00170 Helsinki, Finland

Summary

The most commonly reported measures of prognosis for patients with cancer or some other chronic disease are 5- or 10-year absolute or relative survival rates. Because survival rates often vary by age of patients, and because the age structure of patients often varies between populations, age adjustment is crucial for comparative analyses of survival rates. However, traditional age adjustment often breaks down for long-term survival rates, particularly for sparse data, and it may provide inconsistent results for relative survival rates, even if age adjustment is made to the study population's own age distribution. An alternative approach to age adjustment of both absolute and relative survival rates was recently proposed by Brenner et al. [1] to overcome both the practical and the conceptual problems inherent in traditional age adjustment.
Introduction

The most commonly reported measures of prognosis for patients with cancer or some other chronic disease are long-term survival rates, such as 5- or 10-year survival rates [2]. Typically, both absolute (observed) and relative survival rates are reported. Relative survival rates are derived as the ratios of the observed survival rates and the expected survival rates in the absence of the disease of interest (e.g. cancer) in a population of comparable age and sex distribution (the latter are usually estimated from population life tables). The relative survival rates thereby can be thought of as “net” measures of patient survival indicating patient survival in the hypothetical situation in which the disease of interest is the only cause of death [3].

Absolute, and, to a lesser extent, relative survival rates often vary by age of patients. Therefore, comparisons of patient survival between populations or at various calendar periods within one population may be confounded by age, if there are differences in the age distribution of patients. To overcome this problem, age standardization is commonly employed in comparative analyses of survival rates. Traditionally, age standardized survival rates have been calculated as a weighted average of age specific survival rates within subgroups of patients defined by age at diagnosis, with weights equal to the proportion of patients in those subgroups in some standard population, such as the European standard cancer population or the world standard cancer population [4-8]. A major problem in this context is often the sparseness of data within certain (mostly older) age groups, which may hinder calculation of age specific survival rates. Furthermore, it has been shown that, for relative survival rates, this procedure may provide adjusted survival rates which are conceptually inconsistent with the crude rates and may substantially differ from the latter even if adjustment is made to the study population’s own age distribution [9].

Brenner et al. therefore proposed an alternative approach to age adjustment of both absolute and relative survival rates to overcome both the practical and the conceptual problems inherent in traditional age adjustment. In particular, it does not require to carry out individual age specific survival analyses.

Principles of the Alternative Approach

When employing traditional age adjustment one first calculates the age specific estimates of survival in different age groups, and then combines these estimates in a weighted average (with weights reflecting the age distribution of the standard population). In the alternative approach, specific weights are first individually assigned to all patients in different age groups, and one then carries out conventional survival analysis using the “weighted individual data”, in which the weights are applied to the contributions of patients to the numbers of persons at risk and death. Whereas in the unadjusted (crude) analysis each patient in the study population and her/his contributions to the numbers of persons at risk and deaths are (implicitly) entered with a weight of 1, the proposed form of age adjustment gives weights higher (lower) than 1 to patients in age groups which are underrepresented (overrepresented) in the study population compared to the standard population.
More formally, let \( r_i \) be the relative proportion of patients in age group \( i \) in the study population of total sample size \( n \), and \( s_i \) the corresponding proportion in the standard population. Then, each patient (whether alive during the full follow-up period of interest, censored or died) and all of her/his contributions to the numbers of persons at risk and deaths are assigned a weight of \( \frac{s_i}{r_i} \). This procedure implies, firstly, that the weighted study population has exactly the same age distribution as the standard population (as the sum of weights in each age group \( i \) equals \( n \cdot r_i \cdot \frac{s_i}{r_i} = n \cdot s_i \)), and, secondly, that the total sum of weights equals the total number \( n \) of study participants (as in the crude analysis), as

\[
\sum_i n \cdot \frac{s_i}{r_i} = n.
\]

Compared to traditional age adjustment of survival rates, this procedure has two obvious advantages: firstly, and probably of most practical importance, it does not break down if none of the patients within one or more age groups is followed over the entire follow-up period of interest, a problem commonly encountered in traditional age adjustment. Secondly, this procedure also assures, that age adjustment of relative survival rates to the study population’s own age distribution, which assigns a weight of 1 to each patient, yields exactly the same result as obtained in the crude analysis. This ‘natural’ property is usually not fulfilled for traditionally adjusted relative survival rates as previously shown [9].

**Computational Realization**

For easy implementation of the alternative method two recently developed SAS macros for both relative and absolute survival rates [10,11] were extended. These macros as well as their extension described below may not only be used for traditional “cohortwise” analysis of survival rates, but also for “period analysis”, a more recently introduced method to derive more up-to-date long-term survival estimates [12].

A detailed description of the previously available macros, which can be downloaded free of charge from the internet, has been given elsewhere [10-12] and can be found at http://www.imbe.med.uni-erlangen.de/issan/SAS/period/period.htm. Briefly, two macros have been developed, one called “period” in which relative survival rates are calculated according to the Ederer II method [13], and one called “periodh” in which they are calculated according to Hakulinen’s method [14]. Here, we provide extensions of these macros, denoted “adperiod” and “adperiodh”, in which all contributions of study participants to the observed and expected person-time at risk and to the numbers of deaths are weighted. Thus, the only additional step to be done by the investigator is a priori assignment of appropriate weights to the study participants as described above. The macros will then provide estimates of the adjusted survival rates in a one-step analysis. In particular, there is no need to carry out individual age specific survival analyses.

To facilitate application, the necessary preparatory steps and an exemplary SAS program for specification of the age structure of the standard population are given in Appendix I. Appendix II provides a SAS macro denoted "adweight" that allows for convenient, flexible assignment of appropriate weights to study participants. The extended, commented macros for carrying out weighted survival analyses are listed in Appendix III and Appendix IV. All programs can be downloaded free of charge from http://www.imbe.med.uni-erlangen.de/issan/issan.htm. The source code of the SAS macros is open code under the
conditions of the GNU-GPL licence [15] and thus can be modified by future users to adapt the macros to specific needs.

An empirical illustration of the proposed method using data from the nationwide Finnish Cancer Registry and further in depth discussion of the method is given in the pertinent publication by Brenner et al. [1].

References

Appendix I

Preparatory steps and specifications

/* List of variables needed for each patient and required specifications */

/* Variables needed for all patients: */
/* ---------------------------------- */
/* diagage     age at diagnosis (in years) */
/* sex         gender (1=males, 2=females) */
/* dy          year of diagnosis (e.g. 1990) */
/* dm          month of diagnosis (e.g. 6 for June) */
/* fy          year of end of follow-up (e.g. 1995) */
/* fm          month of end of follow-up (e.g. 12 for December) */
/* vitstat     vital status at the end of follow-up (1=alive, 2=dead) */
/* */

/* Specifications: */
/* * */
/* * */
/* k           length of follow-up (e.g. k=10 for 10-year survival rates) */
/* perbeg      first calendar year included in survival analysis */
/*             (e.g. 1985) */
/* perend      last calendar year included in survival analysis */
/*             (e.g. 1999) */
/* P [2,100]   2 x 100 dimensional array */
/* Elements P [sex,toage] indicate conditional survival rates */
/* from age toage-1 to age toage of male (sex=1) or */
/* female (sex=2) subjects in the general population */
/* [note: may be updated for subsequent calendar years */
/* included in survival analysis as previously described */
/* (Eur J Cancer 2002; 38: 690-695) */

/* Specify age categories (here: 0-44, 45-54, 55-64, 65-74, 75+) */

proc format;
value agecat
  0-44=1
  45-54=2
  55-64=3
  65-74=4
  75-high=5;

/* Specify age distribution of standard population */
/* Example: World standard cancer patient populations (all tumour sites) */
/* Reference: Cancer Survival in Developing Countries, */
/* IARC Scientific Publications No. 145, Lyon 1998 */

data weight_s;
input i s; /* i=age category; s=percentage of subjects in age category */
datalines;
  1 16.4
  2 14.3
  3 20.2
  4 20.8
  5 28.3;
Appendix II

SAS macro "adweight" for assignment of appropriate weights to patients prior to invocation of the macros "adperiod" (Appendix III) or "adperiodh" (Appendix IV)

%macro adweight;
	/*----------------------------------------------------------*/
	/* Determine age distribution of study population */
	/*----------------------------------------------------------*/
	options compress=yes;

data studpop;
set {source data study population};
i=input(put(diagage,agecat.),2.0);               /* recode diagage into i */
proc sort; by i;                   /* data sets will later be merged by i */

proc freq data=studpop;
tables i/out=weight_r noprint; /* weight_r= no. of patients by age group */
	/*----------------------------------------------------------*/
	/* Merge age distributions referring to study and standard population and*/
	/* assign the appropriate weight to each patient */
	/*----------------------------------------------------------*/
data include ;
merge studpop
  weight_r(rename=percent=r drop=count)
  weight_s;
by i;
w=s/r;  /* w = individual weight assigned to each patient */
%mend adweight;
Appendix III

SAS macro “adperiod” for calculating adjusted absolute and relative survival rates. Relative survival rates are calculated according to the Ederer II method.

%macro adperiod(k,perbeg,perend);
** Step 1: Determining the contribution of each patient to persons at risk (perl[i]), observed deaths (perd[i]), and expected deaths in the absence of cancer (perde[i]) for each follow-up year i (1<=i<=k).
Each of these contributions is weighted by the patient’s weight w;
array perl [&k]; array perd [&k]; array perde [&k];
do i=1 to &k; perl[i]=0; perd[i]=0; perde[i]=0; end;
do cy=&perbeg to &perend;
i=cy-dy;
if 0<=i<&k and cy<=fy then do;
age=diagage+(cy-dy);
if age>99 then gp=p[sex,100];
else if age=0 then gp=p[sex,1];
else gp=(p[sex,age]+p[sex,age+1])/2;
** note: gp=expected 1-year survival probability of a person of the same sex and age in the general population;
if i>0 then do;
perl[i]=perl[i]+0.5*w; perde[i]=perde[i]+0.5*(1-gp)*w; end;
if (vitstat=1 and cy<(fy-1)) or (vitstat=2 and cy<fy) or fm>=dm) and i<&k then do;
perl[i+1]= perl[i+1]+0.5*w; perde[i+1]= perde[i+1]+0.5*(1-gp)*w; end;
if vitstat=2 and fm<dm and i>0 and fy=cy then perd[i]=1*w;
if vitstat=2 and fm=dm and i>0 and fy=cy and i<&k then do;
perl[i+1]=1*w; perde[i+1]=(1-gp)*w; perd[i+1]=1*w; end;
end;
end;
** Step 2: Summing up numbers of persons at risk, observed and expected deaths over all patients (dataset sumvar, variables sperl[i], sperd[i] and perde[i], respectively);
proc means noprint sum; var perl1-perl&k perd1-perd&k perde1-perde&k;
output out=sumvar sum=sperl1-sperl&k sperd1-sperd&k sperde1-sperde&k; run;
** Step 3: Determining absolute cumulative survival rates (abssur[i]), expected cumulative survival rates in the absence of cancer (expsur[i]), and relative cumulative survival rates (relsur[i]) for each follow-up year i (1<=i<=k);
data sumvar; set sumvar;
array sperl[&k]; array sperd[&k]; array perde[&k]; array abssur[&k];
array expsur[&k]; array relsur[&k];
do i=1 to &k;
if sperd[i]>0 and sperd[i]>=sperl[i] then abssur[i]=0;
else if i=1 then abssur[i]=1-sperd[i]/sperl[i];
else if abssur[i-1]=0 then abssur[i]=0;
else abssur[i]=abssur[i-1]*(1-sperd[i]/sperl[i]);
if i=1 then expsur[i]=1-sperde[i]/sperl[i];
else expsur[i]=expsur[i-1]*(1-sperde[i]/sperl[i]);
if abssur[i]=0 then relsur[i]=0;

else relsur[i]=abssur[i]/expsur[i];
end;

** Step 4: specification of output;

proc means mean;  var  abssur1-abssur&k
                     relsur1-relsur&k; run;
%mend adperiod;
Appendix IV

SAS macro “adperiodh” for calculating adjusted absolute and relative survival rates. Relative survival rates are calculated according to the Hakulinen method (13).

%macro adperiodh(k,perbeg,perend);

** Step 1: Determining the contribution of each patient to persons at risk (perl[i]), observed deaths (perd[i]), and expected deaths in the absence of cancer (perde[i]) for each follow-up year i (1<=i<=k).
Each of these contributions is weighted by the patient’s weight w;

array perl [&k]; array perd [&k]; array perle [&k]; array perde [&k];
do i=1 to &k; perl[i]=0; perd[i]=0; perle[i]=0; perde[i]=0; end;
do cy=&perbeg to &perend;
i=cy-dy;
if 0<=i<&k and cy<fy then do;
if i>0 then perl[i]=perl[i]+0.5*w;
if ((vitstat=1 and cy<(fy-1)) or (vitstat=2 and cy<fy) or fm>=dm) and i<&k then perl[i+1]= perl[i+1]+0.5*w;
if vitstat=2 and fm<dm and i>0 and fy=cy then perd[i]=1*w;
if vitstat=2 and fm>=dm and fy=cy and i<&k then do;
perl[i+1]=1*w; perd[i+1]=1*w;
end;
end;
if 0<=i=<&k then do;
age=diagage+(cy-dy);
if age>99 then gp=p[sex,100];
elif if age=0 then gp=p[sex,1];
else gp=(p[sex,age]+p[sex,age+1])/2;
wt=1;
do j=dy+1 to cy;
agej=diagage+(j-dy);
if agej>99 then cp=p[sex,100];
elif cp=p[sex,agej];
wt=wt*cp;
end;
** note: gp=expected 1-year survival probability of a person of the same sex and age in the general population
wt=expected cumulative probability of survival from age at diagnosis to current age of a person of the same sex and age in the general population;

if i=0 then do;
perle[i+1]=perle[i+1]+0.5*wt*w;
perde[i+1]=perde[i+1]+0.5*wt*(1-gp)*w;
end;
if 0<i<&k then do;
perle[i]=perle[i]+0.5*wt*w;
perle[i+1]=perle[i+1]+0.5*wt*w;
perde[i]=perde[i]+0.5*wt*(1-gp)*w;
perde[i+1]=perde[i+1]+0.5*wt*(1-gp)*w;
end;
if 0<i=&k then do;
perle[i]=perle[i]+0.5*wt*w;
perde[i]=perde[i]+0.5*wt*(1-gp)*w;
end;

%end;
** Step 2: Summing up observed and expected numbers of persons at risk and deaths over all patients (dataset sumvar, variables sperl[i], sperd[i], perle[i] and perde[i], respectively);

proc means noprint sum; var perl1-perl&k perdl-perd&k perle1-perle&k sperd1-sperd&k sperde1-sperde&k; output out=sumvar sum=sperl1-sperl&k sperd1-sperd&k sperle1-sperle&k sperd1-sperd&k; run;

** Step 3: Determining absolute cumulative survival rates (abssur[i]), expected cumulative survival rates in the absence of cancer (expsur[i]), and relative cumulative survival rates (relsur[i]) for each follow-up year i (1<=i<=k);

data sumvar; set sumvar;
array sperl[&k]; array sperd[&k]; array sperle[&k]; array sperde[&k];
array abssur[&k]; array expsur[&k]; array relsur[&k];
do i=1 to &k;
if sperd[i]>0 and sperd[i]>=sperl[i] then abssur[i]=0;
else if i=1 then abssur[i]=1-sperd[i]/sperl[i];
else if abssur[i-1]=0 then abssur[i]=0;
else abssur[i]=abssur[i-1]*(1-sperd[i]/sperl[i]);
if i=1 then expsur[i]=1-sperde[i]/sperle[i];
else expsur[i]=expsur[i-1]*(1-sperde[i]/sperle[i]);
if abssur[i]=0 then relsur[i]=0;
else relsur[i]=abssur[i]/expsur[i];
end;

** Step 4: specification of output;

proc means mean; var  abssur1-abssur&k
relsur1-relsur&k; run;

%mend adperiodh;