Point and Interval Estimation of Risk and Rate Advancement Periods in Epidemiologic Studies Using the SAS® Software

Annette Pfahlberg¹, Olaf Gefeller², Hermann Brenner³

¹Inst. Medical Informatics & Biomathematics, University of Münster
²Dept. Medical Statistics, University of Göttingen
³Inst. Medical Informatics, Biometry & Epidemiology, University of Munich

ABSTRACT

In epidemiologic studies, the impact of a risk factor on the occurrence of a particular disease has traditionally been quantified by variants of the attributable risk. For many chronic diseases, however, risk factors merely advance the disease onset, and traditional attributable risk measures do not convey information on the time dimension of disease onset. Therefore, a new concept termed “risk and rate advancement periods” (RAP) has recently been introduced for chronic diseases which describes the time period by which the risk or rate of disease is advanced among exposed subjects. In the paper, the statistical derivation of point and interval estimates for RAP is briefly outlined. For the computational realization, a flexible SAS® macro called RAPESTIM has been developed. The macro works on any SAS data set created by some statistical procedure of the SAS/STAT® software with the OUTEST-option. The SAS program code of RAPESTIM as well as its implementation and usage are explained.

1. Introduction

In epidemiologic studies, the relationship between some exposure characteristic and the occurrence of a particular disease (or other health-related events, such as accidents, suicides etc.) can be quantified employing different approaches. Whenever the focus of the investigation lies on estimating the impact of the exposure on the disease burden of the population, variants of the attributable risk are usually the epidemiologic risk measure of choice. The concept of attributable risk, originally introduced by Levin [1] in 1953, has received considerable attention in recent years (see Gefeller [2] for a bibliography of the pertinent literature). A common drawback of all variants of the attributable risk is that they do not adequately reflect the time dimension of disease risk. In many situations, however, when the disease occurs is of as much concern than whether it occurs. This concern is especially important for inevitable occurrences such as death from any cause. As argued by Greenland and Robins [3], in such situations, the preferable measures of exposure impact might often be measures of exposure effect on the time of death or disease onset, such as “potential years of (disease-free) life lost” (PYLL) due to the exposure of interest [4]. The application of the PYLL concept suffers, however, from a number of serious drawbacks inherent in the necessary assumptions for and the practical realization of the estimation of PYLL. A detailed discussion of these issues is beyond the scope of this paper (see Robins and Greenland [5]).
Brenner, Gefeller and Greenland [6] have recently proposed two alternative simple measures of the exposure impact on timing of disease, the “risk advancement period” and the “rate advancement period”, which are applicable for diseases whose rate monotonically increases with age. The measures are suited for risk factors that accelerate progression to chronic diseases and thereby modify the increase of disease risk with age. In this paper, the statistical derivation of point and interval estimates for the measures is briefly outlined, and the computational realization of the method within the SAS software is delineated.

2. Definition of Risk and Rate Advancement Periods

Let $E$ denote the exposure of primary interest (which may be measured on either a categorical or a continuous scale) and $D$ the binary disease characteristic whose risk $R$ of occurrence during some fixed time interval of length $T$ is investigated in an epidemiologic study. As always, if $D$ is an age–related chronic condition, the effect of age $(A)$ as well as a number of additional covariables $C_i, i = 1, \ldots, n$, (potentially including interaction terms) have to be taken into account.

Denote $R_T(A, E, C)$ the (average) risk of disease within $T$ years after some baseline age $A$, conditional on disease–free survival to age $A$, among population members with exposure $E$ and given vector of covariables $C := (C_1, \ldots, C_n)'$. Suppose we have two exposure levels $E_1$ and $E_0$ such that this risk is always higher among persons with $E_1$ than $E_0$. For any age $A_1$ at baseline and any time period of length $T$ after baseline, one may ask, “For what baseline age $A_0$ would we expect the $E_0$-exposed to have the $T$-year risk that we would expect the $E_1$-exposed to have at age $A_1$?” Formally,

$$R_T(A_0, E_0, C) = R_T(A_1, E_1, C)$$

has to be solved for the earliest possible age $A_0$. If this $A_0$ exists, the “risk advancement period” (RAP) among the $E_1$-exposed relative to the $E_0$-exposed is defined by

$$\text{RAP} := A_0 - A_1.$$

If this $A_0$ does not exist, i.e., if the $E_0$-exposed will never attain the disease risk of the $E_1$-exposed at age $A_1$, then define $\text{RAP} = \infty$. The uniqueness of the RAP measure is guaranteed by the restriction of the application to diseases whose rates increase monotonically with age.

Often, epidemiologic data are obtained in the form of estimated rates rather than risks. The preceding construction and definition can be repeated with $R(A, E, C)$ defined as the (person–time) incidence rate of disease at age $A$, conditional on disease–free survival to age $A$, rather than the $T$-year disease risk. In this case, for a given age $A_1$ the earliest age $A_0$ at which the $E_0$-exposed are expected to have the same incidence rate that the $E_1$-exposed are expected to have at age $A_1$ is sought. If such an $A_0$ exists, the difference $A_0 - A_1$ is defined as the “rate advancement period” (RAP). Otherwise, define $\text{RAP} = \infty$ as above.
3. Estimation of Risk and Rate Advancement Periods

In most cases, the statistical analysis of the data is conducted by employing some form of generalized linear model

\[ f(R) = \beta_0 + \beta_1 \cdot E + \beta_2 \cdot A + \sum_{i=1}^{n} \beta_{i+2} \cdot C_i, \tag{1} \]

where \( f(R) \) is some smooth (strictly monotonic) function of the disease risk or rate and \( \beta_i, i = 0, \ldots, n + 2 \) denote the parameters of the generalized linear model. In most studies, multiplicative models are used [7], such as multiple logistic regression (in which case \( f(R) \) reflects the log odds of the disease risk) or Cox’s proportional hazards model [8] (in which case \( f(R) = \log(R) \) and the model intercept is a function of time), but our method is equally applicable to other types of models, such as additive regression models [7]. This general applicability results from the fact that the derivation of the RAP measures is conditional on the type of model chosen for the analysis. In other words, the validity of the RAP parameter in quantifying the time period by which disease risk or rate is advanced for exposed subjects depends on the validity of the assumptions required for employing a particular type of model for the analysis.

For the sake of algebraic simplicity, we first focus on the situation in which \( E \) is dichotomous and no interaction terms involving either \( E \) or \( A \) are included in the model. This model implies homogenous age effects, in that the age slope of the transformed risk or rate \( f(R) \) is not modified by the exposure \( E \) or some covariable \( C_i \). The adjusted coefficients \( \beta_1 \) and \( \beta_2 \) represent the change of disease risk or rate associated with the presence of the exposure and a one-unit change in the age variable, respectively.

According to model (1), exposed individuals \((E = 1)\) who reach a certain level of disease risk or rate \( R \) at Age \( A = x \) would be expected to have reached the same level \( R \) only at Age \( A = y \) \((y > x)\) had they not been exposed \((E = 0)\), given that all other covariables are identical, that is:

\[ f(R \mid E = 1, A = x, C_i = c_i, i = 1, \ldots, n) = f(R \mid E = 0, A = y, C_i = c_i, i = 1, \ldots, n) \]

\[ \iff \quad \beta_1 + \beta_2 \cdot x = \beta_2 \cdot y \]

\[ \iff \quad y - x = \frac{\beta_1}{\beta_2} \]

From (2) it follows that the ratio \( \beta_1 / \beta_2 \) equals the RAP parameter for the presence of the exposure \( E \). Conversely, RAP can be interpreted as the average time period by which any (critical) risk or rate level could potentially be postponed in exposed subjects through an elimination of the exposure (in the absence of competing causes of death). These interpretations hold also for continuously measured exposures by considering one-unit changes in the exposure variable.

A straightforward estimate for RAP is given by the ratio of the estimates for the model parameters \( \beta_1 \) and \( \beta_2 \), respectively. Consistency, (asymptotic) unbiasedness, and (asymptotic) normality of the estimator \( \widehat{\text{RAP}} \) follow immediately from the corresponding properties of the vector \((\widehat{\beta}_1, \widehat{\beta}_2)'\). Applying the delta–method [9], an estimator for the asymptotic variance of the \( \widehat{\text{RAP}} \) can be derived as

\[ \widehat{\text{Var}} \, \widehat{\text{RAP}} = \frac{1}{\beta_2^2} \cdot \left[ \widehat{\text{Var}} \, \widehat{\beta}_1 - 2 \cdot \frac{\widehat{\beta}_1}{\beta_2} \cdot \widehat{\text{Cov}}(\widehat{\beta}_1, \widehat{\beta}_2) + \left( \frac{\widehat{\beta}_1}{\beta_2} \right)^2 \cdot \widehat{\text{Var}} \, \widehat{\beta}_2 \right]. \tag{3} \]
An asymptotic \((1 - \alpha)\)% confidence interval for RAP based on the asymptotic normality of \(\hat{\text{RAP}}\) can be easily calculated using (3) as follows

\[
\left[ \hat{\text{RAP}} - u_{1-\alpha/2} \cdot \sqrt{\text{Var} \hat{\text{RAP}}}, \hat{\text{RAP}} + u_{1-\alpha/2} \cdot \sqrt{\text{Var} \hat{\text{RAP}}} \right],
\]

where \(u_{1-\alpha/2}\) refers to the \((1 - \alpha/2)\) quantile of the standard normal distribution.

Inclusion of the exposure of interest \(E\) or age \(A\) in interaction terms allows for more flexible modelling of exposure and age impact on the disease risk or rate which may often be more appropriate from a biologic point of view. In these situations, derivation of the RAP parameters, their estimates and corresponding asymptotic variances becomes slightly more complex. For the three most common situations the resulting expressions are summarized below:

I) Models including an interaction term between \(E\) and \(A\):

\[
f(D) = \beta_0 + \beta_1 \cdot E + \beta_2 \cdot A + \beta_3 \cdot E \cdot A + \sum_{i=1}^{n} \beta_{i+3} \cdot C_i
\]

\[
\hat{\text{RAP}}_A = \frac{\beta_1 + \beta_3 \cdot A}{\beta_2}
\]

\[
\text{Var} \hat{\text{RAP}}_A = \frac{1}{(\beta_2)^2} \cdot \left[ \text{Var} \beta_1 + 2 \cdot A \cdot \text{Cov}(\beta_1, \beta_3) + A^2 \cdot \text{Var} \beta_3 \right] + \frac{\beta_1 + A \cdot \beta_3}{(\beta_2)^3} \cdot \left[ \frac{\beta_1 + A \cdot \beta_3}{\beta_2} \cdot \text{Var} \beta_2 - 2 \cdot (\text{Cov}(\beta_1, \beta_2) + A \cdot \text{Cov}(\beta_2, \beta_3)) \right]
\]

II) Models including an interaction term between \(E\) and some covariable \(C_j\):

\[
f(D) = \beta_0 + \beta_1 \cdot E + \beta_2 \cdot A + \beta_3 \cdot E \cdot C_j + \sum_{i=1}^{n} \beta_{i+3} \cdot C_i
\]

\[
\hat{\text{RAP}}_{C_j} = \frac{\beta_1 + \beta_3 \cdot C_j}{\beta_2}
\]

\[
\text{Var} \hat{\text{RAP}}_{C_j} = \frac{1}{(\beta_2)^2} \cdot \left[ \text{Var} \beta_1 + 2 \cdot C_j \cdot \text{Cov}(\beta_1, \beta_3) + C_j^2 \cdot \text{Var} \beta_3 \right] + \frac{\beta_1 + C_j \cdot \beta_3}{(\beta_2)^3} \cdot \left[ \frac{\beta_1 + C_j \cdot \beta_3}{\beta_2} \cdot \text{Var} \beta_2 - 2 \cdot (\text{Cov}(\beta_1, \beta_2) + C_j \cdot \text{Cov}(\beta_2, \beta_3)) \right]
\]

III) Models including an interaction term between \(A\) and some covariable \(C_j\):

\[
f(D) = \beta_0 + \beta_1 \cdot E + \beta_2 \cdot A + \beta_3 \cdot A \cdot C_j + \sum_{i=1}^{n} \beta_{i+3} \cdot C_i
\]

\[
\hat{\text{RAP}}_{C_j} = \frac{\beta_1}{\beta_2 + \beta_3 \cdot C_j}
\]

\[
\text{Var} \hat{\text{RAP}}_{C_j} = \frac{(\beta_1)^2}{(\beta_2 + C_j \cdot \beta_3)^4} \cdot \left[ \text{Var} \beta_2 + 2 \cdot C_j \cdot \text{Cov}(\beta_2, \beta_3) + C_j^2 \cdot \text{Var} \beta_3 \right] + \frac{\beta_1}{(\beta_2 + C_j \cdot \beta_3)^3} \cdot \left[ \frac{\beta_2 + C_j \cdot \beta_3}{\beta_1} \cdot \text{Var} \beta_1 - 2 \cdot (\text{Cov}(\beta_1, \beta_2) + C_j \cdot \text{Cov}(\beta_1, \beta_3)) \right]
\]

1065
4. Computational Realization in SAS: The RAPESTIM Macro

In this section, a detailed description of the computational realization of the RAP approach within the SAS software is provided. A flexible SAS macro called RAPESTIM has been developed in order to compute the estimator $\hat{R}_A$ and its corresponding asymptotic $(1 - \alpha)\%$ confidence interval in models with and without interaction terms between $E$, $A$ and some co-variable $C_j$. Optionally, the results of the RAP calculations are graphically displayed (using the SAS/GRAPH® software) as an useful supplement in situations where the RAP varies with age or some covariable.

The general prerequisite for the application of the RAPESTIM macro is that, prior to running the macro, the analysis of model (1) has been performed using some SAS/STAT procedure in combination with the OUTTEST–option. The specification of this option leads to the creation of a structured SAS data set comprising estimated model parameters and covariance terms. Such an OUTTEST–option is available in many SAS/STAT procedures analyzing generalized linear models, for example, in PROC LOGISTIC, PROC PHREG, PROC CATMOD etc., and leads to a simplified computational realization of estimating RAP and corresponding confidence intervals as it makes the results of the statistical modelling procedure accessible for the necessary calculations to derive $\hat{R}_A$ and $\text{Var} \hat{R}_A$.

The general form of the macro call of RAPESTIM is given by:

```sas
%RAPESTIM(<DATA=dataset>,
    <MODEL = type of model>,
    E = exposure variable,
    A = age variable
    <, INT = interaction variable>
    <, CILEVEL = confidence level>
    <, OUT = output dataset>
    <, PLOT = plot indicator>
    <, MIN = lower bound for RAP calculation>
    <, MAX = upper bound for RAP calculation>
    <, INCR = width of the grid for RAP calculation>
    <, DEV = device name>
    <, DEVPLOT = targetdevice name>
    <, HS = horizontal size of the plot>
    <, VS = vertical size of the plot>
    <, TITLE_F = font for the titles>
    <, TEXT_F = font for the rest of the text>
    <, TITLE_H = height for the titles>
    <, TEXT_H = height for the rest of the text>
    <, C1 = colour for the RAP curve>
    <, C2 = colour for the confidence curves>);```

where

dataset: names the SAS data set created by the OUTTEST–option (default: _LAST_ as in SAS)
type of model: distinguishes between different types of analysis (default: 0):
MODEL = 0: no interaction term included in the model
MODEL = 1: model includes interaction between exposure and age
MODEL = 2: model includes interaction between exposure and a covariable
MODEL = 3: model includes interaction between age and a covariable
exposure variable: specifies the exposure variable for the analysis
age variable: defines the age variable
interaction variable: specifies the variable containing the interaction term (default: no interaction variable)
confidence level: defines the confidence level of the asymptotic RAP confidence interval (default: 0.95)
output dataset: offers the opportunity to direct the output to a data set (default: output is printed)
plot indicator: option to request graphical display of the RAP curve including confidence interval limits (PLOT = 1, default: 0)
lower/upper bound and width of the grid for RAP calculation: specify the grid for the RAP calculation

All other macro variables of the RAPESTIM call refer to potential modifications of the graphical display and are thus only relevant in combination with PLOT = 1:

device name: specifies the device name for the terminal or for the plotter (default: win)
target device name: specifies the target device name to preview the graphics to be plotted on a white background (default: hpljs2)
horizontal, vertical size of the plot: defines the plot area (default: HS = 26cm, VS = 18cm)
font for the titles/text: names the SAS font to be used for the titles and the text, respectively (default for both: swissl)
height for the titles/text: defines the height to be used for the titles (default: 1.5) and the rest of the test (default: 1.0), respectively
colour for the RAP/confidence curves: specifies the colour of the curves (default for both: black)

Further important local variables within the RAPESTIM macro are:
b1, b2, b3: denote the estimates of the model parameters \( \beta_1, \beta_2 \) and \( \beta_3 \)
varb1, varb2, varb3, covb1b2, covb1b3 covb2b3: refer to the estimates of the particular elements of the variance-covariance matrix
rap: contains the estimated risk/rate advancement period(s)
varrap: comprises the estimated variance(s) of RAP
cil, ciu: gives the lower and upper \((1 - \alpha)\%\) confidence limits of RAP

Listing of the SAS program code:

```
%macro rapestim(data= &sysdsn,model=0,e=,a=,int=none,cilevel=0.95,
out=PRINT,plot=0,
min=0,max=0,incr=0,
dev=win,devplot=hpljs2,hs=26cm,vs=18cm,

1067
```
data prep; /*preparation*/
set &data;
retain b1 b2 b3 varb1 varb2 varb3 covb1b2 covb1b3 covb2b3;
i=1;
if upcase(_name_) = "ESTIMATE"
   then do;
      b1=&e;
      b2=&a;
      b3=&int;
   end;
if upcase(_name_) = upcase("&e")
   then do;
      varb1=&e;
      covb1b2=&a;
   end;
if upcase(_name_) = upcase("&a")
   then varb2=&a;
if upcase(_name_) = upcase("&int")
   then do;
      varb3=&int;
      covb1b3=&e;
      covb2b3=&a;
   end;
keep i b1 b2 b3 varb1 varb2 varb3 covb1b2 covb1b3 covb2b3;
run;

data final;
set prep;
by i;
if last.i;
file &out notitle;
u = probit(0.5 + &cilevel / 2);
/* model 0: Models without an interaction term involving E or A */
if &model = 0
   then do;
      rap = b1/b2;
varrap = (1/(b2**2))*(varb1 - 2*(b1/b2)*covb1b2 + varb2*(b1/b2)**2));
ciu = rap - u*sqrt(varrap);
cio = rap + u*sqrt(varrap);
end;

/* model I: Models including an interaction term between E und A */
if &model = 1 then do;
  do k = &min to &max by &incr;
  rap = (b1 + b3*k)/b2;
  varrap = (1/(b2**2))*(varb1 + 2*k*covb1b3 + (k**2)*varb3) + 
  ((b1+k*b3)/b2**3)*((b1+k*b3)/b2*varb2 - 2*(covb1b2 + k*covb2b3));
  ciu = rap - u*sqrt(varrap);
  cio = rap + u*sqrt(varrap);
  output;
end;
end;

/* model II: Models including an interaction term between E and C */
if &model = 2 then do;
  do k = &min to &max by &incr;
  rap = (b1 + b3*k)/b2;
  varrap = (1/(b2**2))*(varb1 + 2*k*covb1b3 + (k**2)*varb3) + 
  ((b1+k*b3)/b2**3)*((b1+k*b3)/b2*varb2 - 2*(covb1b2 + k*covb2b3));
  ciu = rap - u*sqrt(varrap);
  cio = rap + u*sqrt(varrap);
  output;
end;
end;

/* model III: Models including an interaction term between A and C */
if &model = 3 then do;
  do k = &min to &max by &incr;
  rap = b1 / (b2 + b3*k);
  varrap = (b1/((b2+k*b3)**3)) 
  *(((b2+k*b3)/b1)*varb1) - 2*(covb1b2 + k*covb1b3) 
  +((b1**2)/((b2+k*b3)**4)) 
  *((varb2+2*k*covb2b3+(k**2)*varb3));
  ciu = rap - u*sqrt(varrap);
  cio = rap + u*sqrt(varrap);
  output;
end;
end;

%if &model = 1 %then %let labelvar = %upcase(&a);
%else %let labelvar = %upcase(&int);
%if &model = 0 %then %do;
  %let plot = 0;

1069
%let varlist= rap varrap ciu cio;
%end;
%else %let varlist= k rap varrap ciu cio;
%let ciproz = %substr(&cilevel,3,2);

label k = "&labelvar";

put "Estimation of Risk/Rate Advancement Periods (RAP) for Exposed Subjects";
put;
put "DATASET = &dataset";
put "Exposure variable = &e";
put "Age variable = &a";
put "Effect-modifying variable = &int";
put "Confidence interval level = &ciproz. %";
put;
title1 'Estimation of Risk/Rate Advancement Periods (RAP) for Exposed Subjects';

%if &model = 0
%then %do;
    %let titleva1=
        No interaction between exposure/and other factors assumed;
    %let titleva2=
        Estimated RAP value and &ciproz. % confidence limits;
%end;
%if &model = 1
%then %do;
    %let titleva1=
        Interaction between exposure variable and age variable assumed;
    %let titleva2=
        Estimated RAP values and &ciproz. % confidence limits varying with &a ;
%end;
%if &model = 2
%then %do;
    %let titleva1=
        Interaction between exposure variable and variable &int assumed;
    %let titleva2=
        Estimated RAP values and &ciproz. % confidence limits varying with &int ;
%end;
%if &model = 3
%then %do;
    %let titleva1=
        Interaction between age variable and variable &int assumed;
    %let titleva2=
        Estimated RAP values and &ciproz. % confidence limits varying with &int ;
%end;

title2 "&titleva1";
title3 "&titleva2";
It should be noted that the options FORMDLIM and TARGETDEVICE are only available in SAS installations beginning with version 6.06. If an earlier SAS version is used, these options have to be deleted and other defaults for the macro variables C1 and C2 defining the colours of the RAP and confidence curves should be specified. In addition, modifications of the options responsible for details of the graphical display may be performed within the programming code to tailor the graphical output to the needs of the user.

4. Discussion

The concept of the risk and rate advancement periods provides an attractive alternative to traditional measures of risk factor impact in certain types of epidemiologic studies on chronic diseases. Unlike the various concepts of the attributable risk, it conveys the information on the time period by which the risk or rate of disease onset is advanced due to the risk factor of interest. With this simple interpretation, the RAP measures may facilitate communication of risks of hazardous behaviour such as cigarette smoking to exposed population groups [10] and hence be a valuable tool in health education.

Risk and rate advancement periods must be interpreted with reference to the strength of the “age effect”, i.e., the underlying age-incidence curve, because the magnitude of the advancement period will vary inversely with the strength of the age effect. In special situations, this age effect can dominate any exposure effect in comparison of RAPs. A detailed discussion of this and several other problematic issues in the practical application of the RAP approach to epidemiologic data, for example, absence of competing causes of death, is provided in Brenner et al. [6]. With careful interpretation, however, risk and rate advancement periods can be a useful tool to quantify and communicate exposure impact on risks or rates of many chronic diseases. The SAS macro RAPESTIM described in this paper provides an easy-to-use access to this new methodology. It offers a convenient solution to the problems of the practical realization concerning the computational aspects of the RAP approach without the necessity to leave the powerful SAS system.

Acknowledgements

The authors gratefully acknowledge the support given by Prof. Dr. S. Greenland in developing the theoretical foundations of the new methodology.
References


Address for correspondence:

Dipl.-Stat. Annette Pfahlberg
Institut für Medizinische Informatik und Biomathematik
Westfälische Wilhelms-Universität Münster
Domagkstraße 9
D-48129 Münster
Germany

SAS, SAS/GRAPH, and SAS/STAT are registered trademarks of SAS Institute Inc., Cary, NC, USA.