Cancer Projections Network (C-Proj)

Canproj: The *R* Package of Cancer Projection Methods Based on Generalized Linear Models for Age, Period, and/or Cohort

Version I

Prepared by

Alberta Health Services for

the

Canadian Partnership Against Cancer

February 2013

Document prepared by:

Dr. Zhenguo Qiu, Senior Biostatistician

And the Cancer Projection Analytical Network Working Team:

Dr. Juanita Hatcher, Principle Investigator

Surveillance and Reporting, C-MORE, Cancer Care, Alberta Health Services

Production of this document has been made possible through a financial contribution from Health Canada, through the Canadian Partnership Against Cancer.

Preferred Citation: Canproj — The *R* package of cancer projection methods based on generalized linear models for age, period, and/or cohort. Alberta Health Services: 2011-12-16

Contents

Chapter 1. Introduction to Methods 4

1.1. Age-Drift-Period-Cohort Models — Nordpred 5

1.2. Age-Cohort Models 6

1.3. Age-Only, Common Trend and Age-Specific Trend Models — Hybrid 7

1.4. Selection of Methods 9

1.5. Other Considerations 11

1.5.1. Linear Interpolation 11

1.5.2. Changes due to Risk, Population Growth and Aging 12

References 14

Chapter 2. Usages of Canproj Package 15

2.1. Data Formatting 15

2.2. How to Use canproj 18

2.3. How to Choose Projection Methods 26

Chapter 3. The Canproj Manual 34

3.1. Canproj Functions 34

3.2. Adpcproj Functions 38

3.3. Acproj Functions 46

3.4. Hybdproj Functions 53

3.5. Ave5proj Functions 61

3.6. Miscellaneous Functions 63

Chapter 1. Introduction to Methods

Generalized linear models (GLM, McCullagh & Nelder, 1989) can be used to analyze age, period, and/or cohort effects in historical cancer incidence/mortality data, and the corresponding fitted model can be utilized to project the future cancer incidence/mortality in terms of both rates and counts at age-specific level.

Most cancers are age-dependent. In the Canproj *R*-package, three models are available: the Nordpred model incorporating age, drift, period and cohort effects, the Age-Cohort model, and the Hybrid models incorporating age and potentially period effects. The Hybrid method considers the age and period (can be 1-, 2-, 3- and 4-year time period) factors as the predictors and includes the common trend model and age-specific trend model. When there is cohort effect, the age-cohort method (AC-model) or the age-drift-period-cohort method (the Nordpred *R*-package) can be used, depending on the significance of the drift-term in the Nordpred. The Canproj package is designed to determine which model is appropriate using a defined decision tree, but contains 3 sub-packages which can be run independently: adpcproj (Nordpred), acproj (AC-model) and hybdproj (Hybrid) which also contains the age-only model.

The following assumptions apply to all models:

- Age is grouped into 5-year age groups indexed by $a = 1, 2, \ldots, A$ ($A = 18$);
- Period is indexed by 5-year period index $p = 1, 2, \ldots, P$, where p represents a 5-year period in the Age-Cohort and Nordpred models, and in the Hybrid method represents any-year aggregated period index or annual year index as appropriate;
- The 10-year overlapped cohort in the Age-Cohort and Nordpred models is indexed by $c = 1, 2, \ldots, C$, where $C = A + P - 1$ and $c = A - a + p$;
- The "power 5" link function, $x^{0.2}$, is used as the default link function, together with other choices: The logarithmic (log), the square-root (sqrt) and the identity functions;
- Attenuating parameters are used to degenerate any linear trend;
- Recent 10 year average age-specific rates are used for age groups with small numbers;

- The number of age-specific cases, in the p^{th} period, Y_{ap}, follows a Poisson distribution with offset from the age-specific population size, n_{ap}, $Y_{ap} \sim Poi(\mu_{ap}, n_{ap})$;
- When there is "over-dispersion", the Poisson distribution will be replaced by the negative-binomial distribution (*NB*), i.e., $Y_{ap} \sim NB(\mu_{ap}, n_{ap})$.

All 3 methods provide the projected age-specific rates. The annual rates are calculated by the segmental linear interpolation method. The projected age-specific counts are obtained by multiplying the projected age-specific rates with corresponding age-specific population sizes. Age-standardized rate and total annual number are summarized from the age-specific rates and counts.

1.1. Age-Drift-Period-Cohort Models — Nordpred

Møller (2004) and his co-authors proposed the Nordpred approach (Møller *et al*, 2003) based on the age-drift-period-cohort analysis using the GLM suggested by Clayton & Schifflers (1987). The age-drift-period-cohort model can be written as

$$h(\mu_{ap} / n_{ap}) = \alpha_a + D \times p + \beta_p + \gamma_c, p = 1,2,\ldots,P, \tag{1}$$

where μ_{ap} is the Poisson mean parameter (expected count) in age group a and period p, D is the common drift parameter, α_a, β_p and γ_c are the non-linear components for age group a, period p and cohort c, respectively. Throughout this document, h represents the link function in the GLM.

The trend (the drift parameter) in the most recent 10 years is generally used to capture the most recent linear trend. In Nordpred, the choice between using the trend over the entire historical data or the recent trend is based on whether the quadratic term in the log-linear model is significant:

$$\log(\mu_{ap}) = \alpha_a + S_0 \times p + S \times p^2 + \gamma_c + \log(n_{ap}) \tag{2}$$

If the estimated $\hat{S}$ was significant, the recent trend $\hat{D}_{last} = \hat{D} - \hat{\beta}_{P-1}$ is used; otherwise, $\hat{D}$ is used.

Due to the linear dependence among the three time variables: age, period and cohort, restrictions must be placed on the parameters. In the full Nordpred model, the first (β_1) and the last non-linear period effects (β_P) are constrained to be equal, and the first (γ_1) and the last non-linear cohort effects (γ_C) equal. Any net period- and cohort-specific linear change is accounted for by a common component: the drift parameter D in (1).

For the projections of the future N periods (The maximum of N is 5), the estimated drift parameter ($\hat{D}$ or $\hat{D}_{last}$) is attenuated from 100%, to 75%, 50%, 25% and 0% in the future first, second, third, fourth and fifth 5-year period of the projection years, respectively. The future non-linear period effects are assumed equal to the last estimated effects (β_P) in the model (1). Due to the constraint, the estimate of β_P is assumed to be zero. Thus, the future age-specific rate:

$$\hat{r}_{ap} = h^{-1}(\hat{\alpha}_a + \hat{D} \times p + \hat{D} \times P + \hat{\gamma}_c) \tag{3}$$

or

$$\hat{r}_{ap} = h^{-1}(\hat{\alpha}_a + \hat{D}_{last} \times p + \hat{D} \times P + \hat{\gamma}_c)^5 \tag{4}$$

for p =1, 2, …, N, where $c = A - a + (P + p)$. The estimated age effects are assumed to be invariant in the future periods. The new cohort effects in the future periods are assumed to be zero, or the estimate of the second last cohort effect. It is noted that the Poisson distribution may be replaced by the negative-binomial distribution if there is over dispersion and the data for at least the most recent 15 years (3 5-year periods) are required to apply the adpcproj sub-package. Users can change the parameter settings such as the attenuation parameter and the way to estimate the new cohort effect in the sub-package. Full details are in Qiu & Hatcher (2011).

1.2. Age-Cohort Models

Breslow & Clayton (1993) proposed the log-linear mixed effect model for the age-cohort analysis using the data of the breast cancer incidence in Iceland as an example. Under the Bayesian framework, this example had been re-analyzed by non-parametric smoothing method using Markov chain Monte Carlo technique. The age-cohort model can be written as

$$h(\mu_{ap} / n_{ap}) = \alpha_a + \gamma_c, c = p + A - a, p = 1,2,\cdots,P. \tag{5}$$

Due to the paucity of data (one point only) in the both extreme cohorts (the youngest and the eldest birth cohorts), the age and cohort effects are estimated by setting the middle cohort (with complete observations) as the reference. The age-specific rates for young age groups with small number (say, < 10 per 5-year period) of cases/deaths are estimated by the average method based on the most recent two periods (recent 10 years); and the rest of age groups with sufficient numbers are used for the age-cohort modeling. Again, when there is "over-dispersion", the Poisson distribution will be replaced by the negative-binomial distribution and the data for the most recent 15 years at least should be available to apply the acproj sub-package.

As the estimates of both extreme cohorts are not robust owing to the paucity of data, the estimations of $\hat{\gamma}_3, \hat{\gamma}_4, \cdots, \hat{\gamma}_{C-2}$ are used in the linear regression to estimate the future cohort effects. The estimated slope from the linear model is utilized to extrapolate the future cohort effects, including the last two cohort effects (C – 1, C) from the GLM (5). Similar to the extrapolation of drift effect in Nordpred, the estimated slope is attenuated gradually from 100%, to 75%, 50%, 25% and 0% in the estimation of the future first, second, third, fourth and fifth cohort parameters, respectively; and the attenuation parameter can be changed by user.

For the projections of the future N periods (The maximum of N is 5), the projected age-specific rate is:

$$\hat{r}_{ap} = h^{-1}(\hat{\alpha}_a + \hat{\gamma}_{p+A-a}), p = P+1, P+2, \cdots, P+N \tag{6}$$

It is noted that the estimated age effects are unchanged in the future but the cohort effects are evolved over future periods as the cohort is aging.

1.3. Age-Only, Common Trend and Age-Specific Trend Models — Hybrid

In situations where the cohort effects are not significant, the Hybrid method takes the advantages of the average method, the joinpoint type methods (Kim, Fay & Feuer *et al*, 2000) and the Poisson regression method (Dyba, Hakulinen & Päivärinta, 1997), to fit the *best* regression model for the *best* defined projection base, and uses the estimated parameters to extrapolate future age-specific rates (Qiu, Jiang & Hatcher, 2010).

For annual or aggregated data, year (or period) is regarded as a continuous variable and age group as an ordinal categorical variable. When the number of annual cases is small the data are usually aggregated over a number of years to stabilize the rate. The historical data is split by age into two datasets: one for age groups where there are small numbers (say, < 2 per year) and the average method (using data from the most recent 5-years) is used, and the other where statistical regression modeling can be applied.

In the age groups where regression models can be used, first compare the common trend model for the entire historical period:

$$h(\mu_{ap} / n_{ap}) = \alpha_a + \beta \times p, p = 1,2,\cdots,P \tag{7}$$

with the age-specific trend model:

$$h(\mu_{ap} / n_{ap}) = \alpha_a + \beta_a \times p, p = 1,2,\cdots,P \tag{8}$$

by the chi-square test. If there is no significant difference between the two models, the common trend model should be used; otherwise, the age-specific model should be used. Based on the selected model, the likelihood searching method (Qiu, Jiang & Hatcher, 2010) is used to determine the projection base.

Figure 1 illustrates the model selection diagram. In order to determine the final model to be used in the projections, the model (7) and (8) are fitted using only data from the projection base. The final model is selected based on the following criteria:

- If there is no significant difference between the common trend model and the age-specific model, and the estimated common trend parameter is not significant, the final model should be the age-only model;
- If there is no significant difference between the common trend model and the age-specific model, and the estimated common trend parameter is significant, the final model should be the common trend model;
- If there is a significant difference between the common trend model and the age-specific model the final model should be the age-specific trend model;
- When over-dispersion exists in the age-specific trend model, the negative binomial distribution will replace the Poisson distribution.

Similar to Nordpred, the default attenuation parameter for the estimated slope(s) ($\hat{\beta}$ or $\hat{\beta}_a$) is 5% per year after the first 5 projection years. The attenuation parameter in the first 5 projection years can be set by user. The data for the most recent 6 years at least should be available to apply the hybdproj sub-package. It is noted that the Hybrid method can work for annual data, which allows us to obtain better estimates of the trends for major cancer sites; for rare cancer sites, however, aggregation is required to stabilize the trend.

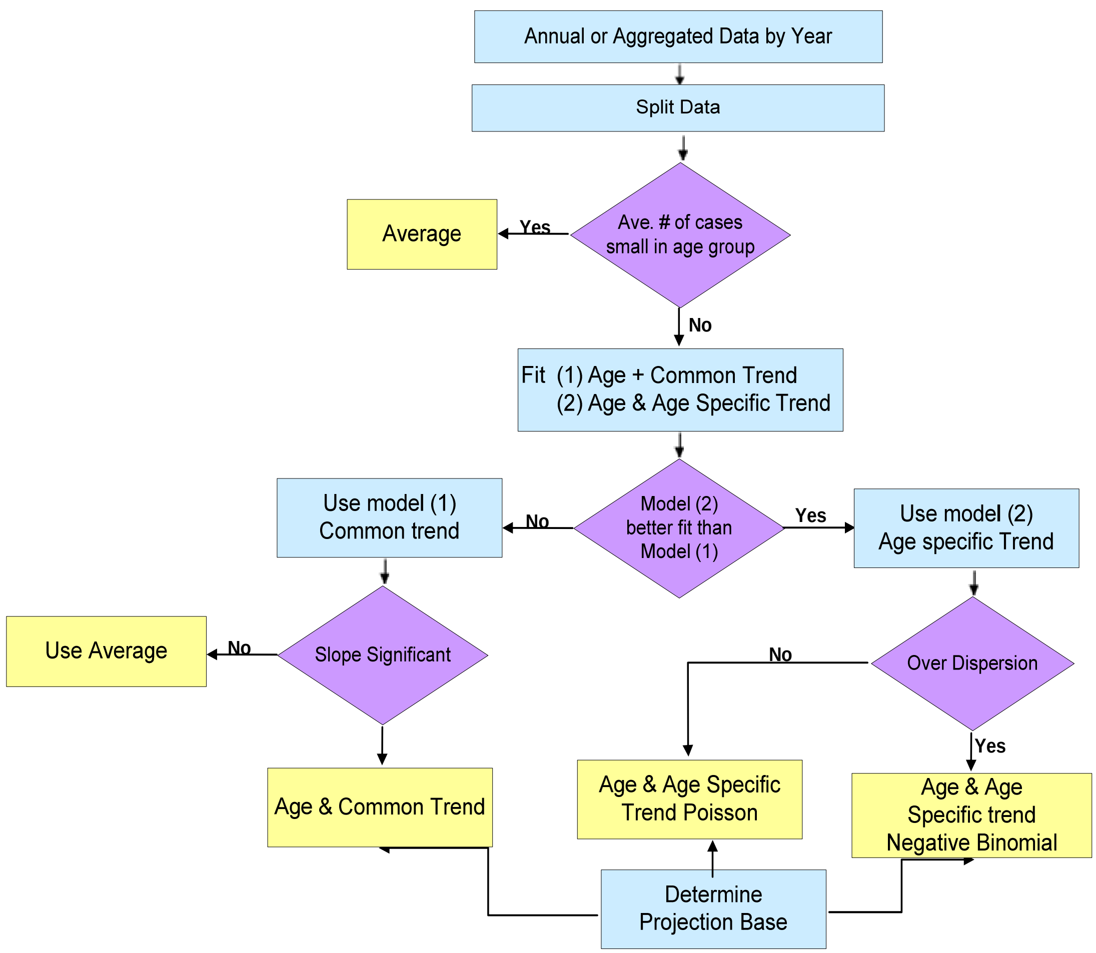

Figure 1. The diagram of model decision in the Hybrid method.

1.4. Selection of Methods

The full age-drift-period-cohort model may over-parameterize the data and thus dilute the significance of the drift estimator. For certain cancer sites cohort and/or period effects may not be present irrespective of the numbers of cancers or deaths. The model is based on 5-year aggregated data, whatever the number of cancer cases/deaths in each age group, which may obscure the detailed changes in trends for major cancer sites. When the full model over-parameterizes historical data, the best model could be the age-cohort model, age-specific trend model, common trend model, or the age-only model. The age-only model is the baseline of the set of candidate models assuming the cancer is age-dependent.

Cancer projection using statistical regression models is a statistical inference process. In general, if the models are nested, sequential statistical testing techniques can be used to select the best model. However, the data sets used in Nordpred, AC model and Hybrid are different unless the same aggregated dataset (5-year aggregation) is used in the Hybrid method. Thus model selection based on sequential statistical testing is not appropriate to determine to whether the Nordpred, the AC model or the Hybrid method should be used for cancer projections. To determine which method is more appropriate, a decision tree has been developed (see Fig. 2) and the canproj package uses this decision tree to select the method to be used.

Statistically, the difference between Nordpred or AC-model and the Hybrid methods is whether the projection model includes the non-linear cohort effects. The Nordpred or AC model is used if the cohort effects are significant (using a Type III error analysis and an *F*-test), and the Hybrid method otherwise. In Nordpred, the estimated drift plays the most important role in the projections. If the estimated drift in the full age-drift-period-cohort model is not significant then the projections would be mainly based on the estimates of the cohort effects. However, the cohort estimates may not be robust in the extreme cohorts. The choice of the Nordpred method or the AC model method in this decision tree is based on the *p*-value of the drift estimator. If the drift estimator is not significant the AC model method is used, otherwise the Nordpred method is used.

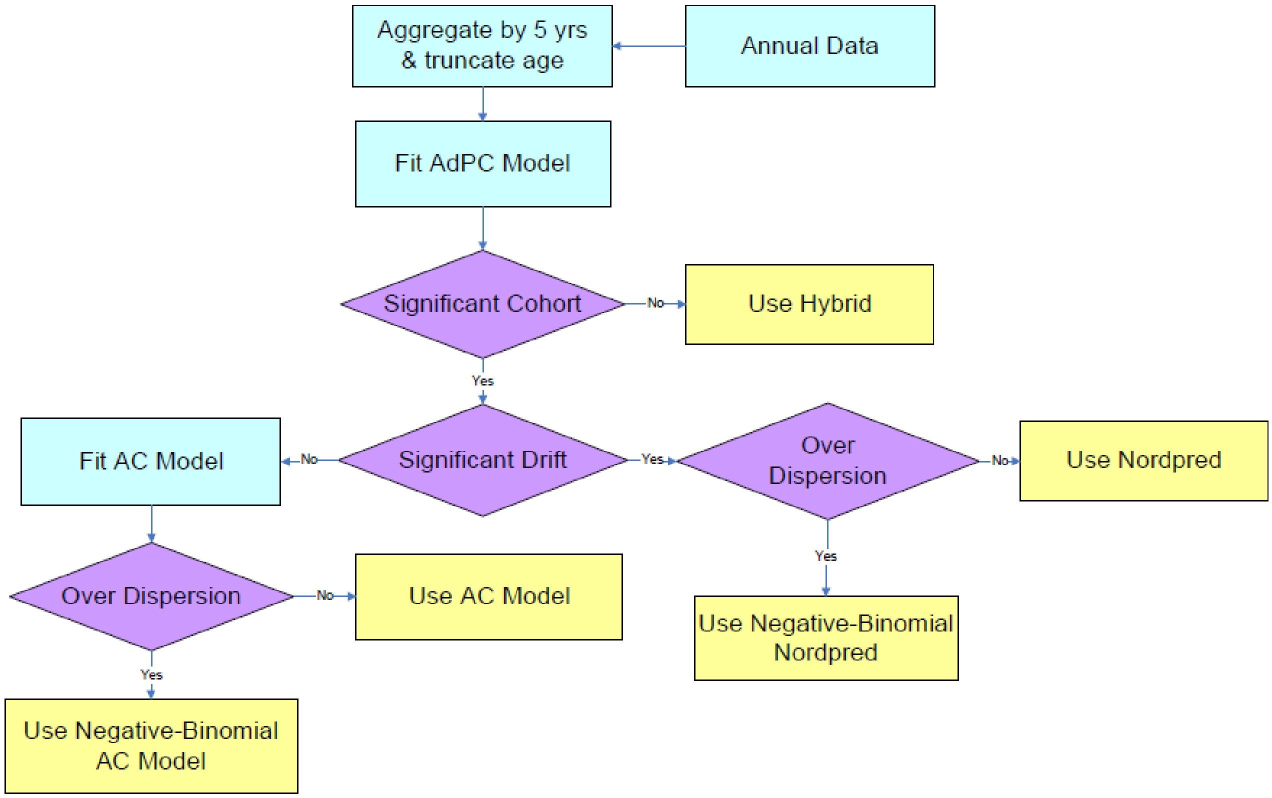

Figure 2. The diagram of method selection in canproj package.

1.5. Other Considerations

1.5.1. Linear Interpolation

The Nordpred method and the AC-model method provide the estimated rates and numbers in five yearly periods. When the Hybrid method is used, the data may be aggregated by 2, 3, or 4 years. The results represent the rates and numbers in the middle year of the period. The segmental linear interpolation method can be used to expand the estimated age-specific rates into annual rates. For example, when data is aggregated by 5 years, for age-group a, the t^{th} year rate between the p^{th} and the $(p + 1)^{th}$ period is

$$r_{at} = \frac{5-t}{5} r_{ap} + \frac{t}{5} r_{a(p+1)}, \; t = 1, 2, 3, 4, 5 \tag{9}$$

where r_{ap} and $r_{a(p+1)}$ are the age-specific rates at age-group a in period p and $p + 1$, respectively. Essentially, for aggregation by odd number of years, the estimated rate expresses the rate in the middle year of the period, and thus, the period is split by equal-part and each part represents one year. Therefore, the 5-year period is split by 4 parts and the 3-year period is split by 2 parts.

When data is aggregated by even number of years, the middle doesn't represent any fully calendar year. However, the similar way can be adopted to expand the estimated age-specific rates into annual rates. For example, when data is aggregated by 2 years, the annual age-specific rates will be

$$r_{at} = \frac{4-t}{4} r_{ap} + \frac{t}{4} r_{a(p+1)}, t = 1,2,3,4$$

where t represents every half year. That is, the 2-year period is split by 4 equal parts. Similarly, the 4-year period is split by 8 equal parts.

To facility the linear interpolation, the function "asrpy" is developed in the canproj package.

1.5.2. Changes due to Risk, Population Growth and Aging

Health planners may be interested in how much of the change in the future number of cases is due to the change in risk, how much is due to population growth, and how much is due to population aging, compared to a baseline or reference year. This may be calculated as follows.

Let r_{as} be the age-specific rates in year s (the reference year), r_{at} be the age-specific rates in year t (the comparison year), n_{as} be the a^{th} age population sizes in year s, and n_{at} be the a^{th} age population sizes in year t. Then, the total number of cases in the reference year is $Y_s = \sum_{a=1}^{18} n_{as} \times r_{as}$;

And the total number of cases in the comparison year is $Y_t = \sum_{a=1}^{18} n_{at} \times r_{at}$.

Overall, the percentage change in annual number (C_o) is defined as the percentage of the difference between the total number of cases in the comparison year and that in the reference year:

$$C_o = 100 \times \frac{Y_t - Y_s}{Y_s}. \tag{10}$$

Applying the population in the comparison year to the age-specific rates in the reference year, i.e. assuming there is no rate change in the future, (11) gives the change by population growth and aging (C_p):

$$C_p = 100 \times \frac{Y'_t - Y_s}{Y_s}, \tag{11}$$

where $Y'_t = \sum_{a=1}^{18} n_{at} \times r_{as}$. And thus, the rest from the overall change is due to the change in risk (C_r):

$$C_r = C_o - C_p = 100 \times \frac{Y_t - Y'_t}{Y_s}. \tag{12}$$

Eliminating the impact from population growth and risk change, the change due to population growth is (C_g):

$$C_g = 100 \times \frac{Y^*_t - Y_s}{Y_s} = 100 \times \frac{N_t - N_s}{N_s}, \tag{13}$$

where $Y^*_t = R_s \times N_t, N_t = \sum_{a=1}^{18} n_{at}, N_s = \sum_{a=1}^{18} n_{as}$, and R_s is the crude rate in the reference year. Then the rest in total change due to population is the change due to population aging (C_a):

$$C_a = C_p - C_g = 100 \times \frac{Y'_t - Y^*_t}{Y_s}. \tag{14}$$

It is noted that the change due to population growth is only related to the population size in the reference and the comparison year and should be a constant for any cancer sites for a target population. The function "chper" is developed in the canproj package to calculate the percentage changes.

References

Breslow NE & Clayton DG. Approximate Inference in Generalized Linear Mixed Models. *JASA*, 1993, **88**(421): 9 – 25.

Clayton D & Schifflers E. Models for temporal variation in cancer rates. I. Age-period and age-cohort models. *Stat Med*, 1987a, **6**: 449 - 467.

Clayton D & Schifflers E. Models for temporal variation in cancer rates. II. Age-period-cohort models. *Stat Med*, 1987b, **6**: 469-481.

Dyba T, Hakulinen T & Päivärinta L. A simple non-linear model in incidence prediction. *Stat Med*, 1997, **16**: 2297 - 309.

Kim HJ, Fay MP & Feuer EJ *et al*. Permutation tests for joinpoint regression with applications to cancer rates. *Stat Med*, 2000, **19**: 335 - 51 (correction: 2001, **20**:655).

McCullagh P & Nelder JA. *Generalized Linear Models*, 2nd ed. London: Chapman and Hall, 1989.

Møller B. Prediction of cancer incidence — Methodological considerations and trends in the Nordic countries 1958 - 2022. Dissertation, Faculty of Medicine, University of Oslo, Unipub AS, Oslo, 2004.

Møller B, Fekjær H & Hakulinen T *et al*. Prediction of cancer incidence in the Nordic countries: empirical comparison of difference approaches. *Stat Med*, 2003, **22**: 2751 - 66.

Qiu Z and Hatcher J. An Experimental Comparison between Nordpred and Hybrid Approaches — Introducing the Canproj Method. Technique Report for Cancer Projections Network (C-Proj), 2011.

Qiu Z, Jiang Z, Wang M & Hatcher J. Long-Term Projection Methods: Comparison of Age-Period-Cohort Model-Based Approaches. Technique Report for Cancer Projections Network (C-Proj), 2011.

Qiu Z, Jiang Z & Hatcher J. Comparison of Short-Term Projection Methods: Proposing the Hybrid Approach and Bayesian Models. Technique Report for Cancer Projections Network (C-Proj), 2010.

Qiu Z, Hatcher J and Wang M. Review Cancer Projection Methods for Canadian Partnership Against Cancer Analytical Network. Technique Report for Cancer Projections Network (C-Proj), 2010.

Chapter 2. Usages of Canproj Package

The "canproj" package was written in R 2.12.0 and contains 3 sub-packages: "adpcproj", "acproj" and "hybdproj" using the methods introduced in Chapter 1. Each has the control functions: canproj, adpcproj, acproj and hybdproj. canproj selects the projected results produced from one of the 3 sub-packages based on the decision tree (Figure 2, Chapter 1); and each sub-package can be used directly and independently to obtain projections if the "canproj" decision tree is not appropriate as it may be the cases for some cancer sites. The 3 sub-packages have their own model fitting function (.estimate), extrapolation function (.prediction), result producing functions (.getpred and .getproj), estimation summarizing function (glm.), results plotting function (plot.) and projection method summarizing function (summary.). With other complimentary functions and datasets, the file names in the canproj package are given below:

```
"acproj"              "acproj.estimate"      "acproj.getpred"       "acproj.getproj"
"acproj.prediction"   "adpcproj"             "adpcproj.estimate"    "adpcproj.getpred"
"adpcproj.getproj"    "adpcproj.prediction"  "asrpy"      "asrsd"          "asry"
"ave5proj"            "ave5proj.getproj"     "ca91"                 "ca96"
"canproj"             "canproj.getproj"      "chper"                "datagg"
"glm.acproj"          "glm.adpcproj"         "glm.canproj"          "glm.hybdproj"
"hybdproj"            "hybdproj.estimate"    "hybdproj.getpred"     "hybdproj.getproj"
"hybdproj.prediction" "obasr"                "plot.acproj"          "plot.adpcproj"
"plot.canproj"        "plot.hybdproj"        "projplot"             "summary.acproj"
"summary.adpcproj"    "summary.canproj"      "summary.hybdproj"     "wdsd"
```

Please refer to Chapter 3 for details.

2.1. Data Formatting

For both cancer data and population data, the data frame format is required, with age-specific annual numbers of cases/deaths or population sizes with the rows representing the 18 age groups in ascending ordered and the columns representing the calendar years in ascending ordered. For example, the age-specific population size from age group 1 - 9 and calendar year 1984 – 1993 is

```
> poyf[1:9, 1:10]
   y1984  y1985  y1986  y1987  y1988  y1989  y1990  y1991  y1992  y1993
1 101026 100621 104035 103110 102626 103306 104162 103910 103942 105030
2  91083  89628  93587  94676  97125  99619 102380 104355 105195 105303
3  88760  85456  87003  87154  87591  89165  91280  93604  95827  98822
4  94369  90721  93271  91984  91433  90536  90059  89664  90389  90929
5 128181 118074 118042 111600 106758 102495 101230 100514 98949 97698
6 131007 127694 132003 130024 128522 126968 124950 121054 116506 111166
7 107304 108556 115754 118500 121370 124054 127290 129247 130639 131015
```

```
8 87318 89699 95545 95824 98279 102600 108037 113443 118154 121937
9  64822  66205  70458  76143  80434  84470  89367  93916  95624  98392
```

and the female breast cancer incidence data in the corresponding age groups and calendar years is

```
> iBrstF[1:9, 1:10]
   1984 1985 1986 1987 1988 1989 1990 1991 1992 1993
325    0    0    0    0    0    0    0    0    0    0
326    0    0    0    0    0    0    0    0    0    0
327    0    0    0    0    0    0    0    0    0    0
328    0    1    0    0    0    0    0    0    0    0
329    1    1    0    0    1    0    0    1    2    0
330    8   14    7   10   11   10    6    6    6    5
331   25   30   32   21   27   21   20   27   32   28
332   55   39   42   63   45   42   51   59   62   65
333   55   71   60   67   91   78  109  100   97  106
```

The column names in poyf start with the character “y”:

```
> colnames(poyf)
 [1] "y1984" "y1985" "y1986" "y1987" "y1988" "y1989" "y1990" "y1991" "y1992" "y1993"
[11] "y1994" "y1995" "y1996" "y1997" "y1998" "y1999" "y2000" "y2001" "y2002" "y2003"
[21] "y2004" "y2005" "y2006" "y2007" "y2008" "y2009" "y2010" "y2011" "y2012" "y2013"
[31] "y2014" "y2015" "y2016" "y2017" "y2018" "y2019" "y2020" "y2021" "y2022" "y2023"
[41] "y2024" "y2025" "y2026" "y2027" "y2028" "y2029" "y2030" "y2031" "y2032" "y2033"
```

If we want to remove the y’s, we can assign the numeric calendar year by

```
> colnames(poyf) <- 1984:2033
```

resulting in more easily reading format:

```
> poyf[1:9, 1:10]
    1984   1985   1986   1987   1988   1989   1990   1991   1992   1993
1 101026 100621 104035 103110 102626 103306 104162 103910 103942 105030
2  91083  89628  93587  94676  97125  99619 102380 104355 105195 105303
3  88760  85456  87003  87154  87591  89165  91280  93604  95827  98822
4  94369  90721  93271  91984  91433  90536  90059  89664  90389  90929
5 128181 118074 118042 111600 106758 102495 101230 100514  98949  97698
6 131007 127694 132003 130024 128522 126968 124950 121054 116506 111166
7 107304 108556 115754 118500 121370 124054 127290 129247 130639 131015
8  87318  89699  95545  95824  98279 102600 108037 113443 118154 121937
9  64822  66205  70458  76143  80434  84470  89367  93916  95624  98392
```

The names in row in the cancer case dataset “iBrstF” do not reflect the actual age group:

```
> row.names(iBrstF)
 [1] "325" "326" "327" "328" "329" "330" "331" "332" "333" "334" "335" "336" "337"
[14] "338" "339" "340" "341" "342"
```

This can be changed by

```
> row.names(iBrstF) <- 1:18
```

and thus producing the more easily understood:

```
> row.names(iBrstF)
 [1] "1"  "2"  "3"  "4"  "5"  "6"  "7"  "8"  "9"  "10" "11" "12" "13" "14" "15" "16"
[17] "17" "18"
```

In general, we can produce the appropriate format for the cancer and population datasets by SAS using PROC TRANSPOSE or by SEER*STAT, and then use read.table or read.csv to read the txt file or csv file into R with the same format. For example,

```
> read.table("H://CPACproj/Longterm/Data/Population/popf.txt",header =T,sep="")
```

However, sometimes we may have dataset in rectangle format with index variables for age groups and calendar years. For example, if `candat' is a data frame sorted by age and then by year in the rectangle format:

```
>  candat[c(1:5, 32:36), ]
    Age Year   Cases
1     1   1984    0
2     2   1984    0
3     3   1984    0
4     4   1984    0
5     5   1984    1
32  14   1985   109
33  15   1985    90
34  16   1985    63
35  17   1985    42
36  18   1985    31
```

we can apply

```
> cdat <- trandat(candat, 1984)
```

where the “trandat” function is

```
trandat <- function(dat, starty) {
        n <- dim(dat)[1]/18
        outdat <- as.data.frame(matrix(dat$case, ncol=n))
        row.names(outdat) <- 1:18
        colnames(outdate) <-  starty : (starty + n -1)
        return(outdate)
  }
```

to gain the required data format.

The data requirements are listed below:

- Cancer dataset and population dataset must have data for 18 age groups by row;
- Population dataset must include information for all years in the cancer dataset;
- Population dataset must also include information for the projection years, and the number of projection years must be less than or equal to 25;
- The number of years in cancer dataset must be greater or equal to 15;
- The averaged annual number of cases/deaths in cancer dataset must be greater than 2.

For examples, the iBrstF contains breast cancer incidence in 18 age groups and 25 years (using the dimension function for matrix or data frame "dim" to figure them out)

```
> dim(iBrstF)
[1] 18 25
```

And the poyf contains the population information in 18 age groups and 50 years including the first 25 years matched with those in iBrstF

```
> dim(poyf)
[1] 18 50
```

The averaged annual number of cases/deaths in iBrstF is 1340.5:

```
> mean(apply(iBrstF, 2, sum))
[1] 1340.52
```

Thus, we can apply the canproj functions for iBrstF and poyf.

2.2. How to Use canproj

First install the canproj package by

```
> source("T://surveillance/Projection/Canproj/canproj-v1.R")
```

The folder directory may be changed according to the location of the package. The canproj is the control function to produce the projection results, with arguments addressed below:

```
canproj(cdat, pdat, startp, projfor="incidence", nagg=NULL, ncase=NULL,
        startestage=NULL, newcohort=NULL, Ave5=NULL, sum5=NULL, methods=NULL,
        linkfunc="power5", cuttrd=0.05, shortp=0, pD=0.05, pGOF=0.05, standpop=ca91))
```

where cdat asks the cancer dataset, pdat asks the corresponding population dataset, startp asks the calendar year when the projection start; those 3 arguments must have inputs; For others, defaults can be used. In details,

projfor - "projection for incidence or mortality",

nagg – "the number of years for data aggregation", ncase – "the minimum number of cases/deaths in a age group",

startestage – "the youngest age group used in GLM for adpcproj and acproj sub-packages",

newcohort – "count new cohort effect as zero or not for adpcproj sub-package",

Ave5 – "replace age-only model by the 5-year average method for hybdproj sub-package",

sum5 – "the way of `average' over the recent 5 years for ave5proj sub-package",

linkfunc – "the link function to be used in the GLM",

cuttrd – "the attenuating percent for drift/slope per year after first 5 years",

shortp – "the attenuating percent for drift/slope in the first 5 years",

pD – "the critical value of the *p*-value for testing drift or slope",

pGOF – "the critical value of the *p*-value for testing over-dispersion or model comparison",

standpop – "the proportions in a standard population to get ASR".

Please refer to the canproj manual (Chapter 3, Section 3.1) for more details.

Example 1. Projections for female lung cancer incidence in 2009 – 2033

Creating the projection object by

```
> iLungF.obj0 <- canproj(iLungF, poyf, 2009)  # or canproj(cdat=iLungF, pdat=poyf, startp=2009)
```

From the method summary table below, we determine that, for age 0 – 29, the average method was used using data in 1999 – 2008; for age 30 – 85+ (The starting age group for regression model based projection: startuseage=7), the original Nordpred method (Method: Age-drift-Period-Cohort Model) was used for future 25 years projections (nopred=5); the trend in drift term was attenuated additional 25% in 5-year period after the first 5 years (cuttrend=0, 0.25, 0.5, 0.75, 1) and the trend is based on the entire data in 1984 – 2008 (recent=FALSE). The *p*-value of goodness-of-fit is .0797 > .05, indicating the Poisson distribution was used for the GLM.

```
> summary.canproj(iLungF.obj0)
Prediction done with:
Method:                                          Age-drift-Period-Cohort Model
Number of periods predicted (nopred):            5
Trend used in predictions (cuttrend):            0 , 0.25 , 0.5 , 0.75 , 1
Number of periods used in estimate (noperiod):   5
Distribution function of regression:             Poisson
P-value for goodness of fit:                     0.0797
Used recent (recent):                            FALSE
P-value for recent:                              0.3277
First age group used (startuseage):              7
First age group estimated (startestage):         7
```

The estimate of drift ($coef["Period",]) is a positive value, which is significant (Pr(>|z| = 3.281912e-09):

```
> glm.canproj(iLungF.obj0)$coef["Period", ]
   Estimate      Std. Error       z-value          Pr(>|z|)
4.607290e-03  7.786744e-04  5.916837e+00   3.281912e-09
```

The recent and future cohorts were estimated to be at a lower relative risk than previous cohorts (Fig. 3), thus age-standardized incidence rate (ASR) is projected to have a turning point at 2011, after which the ASRs are projected to decrease over time (Fig. 4).

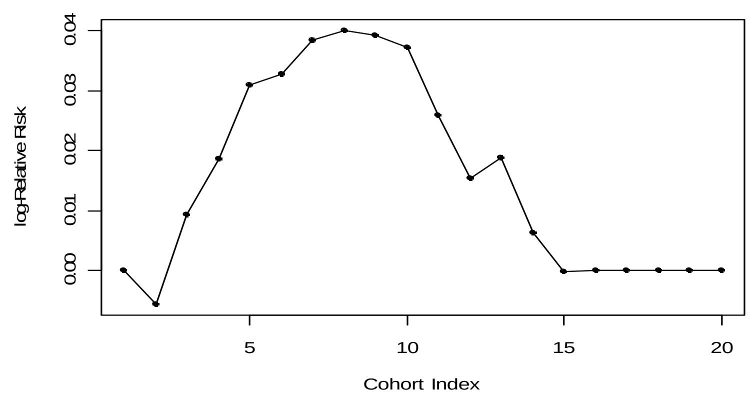

Fig. 3. The estimated and predicted cohort effects in the AdPC model for female lung cancer incidence.

```
# Produce Figure 4:
iLungF.asr <- canproj.getproj(iLungF.obj0, ca91)  # Get annual ASRs and total numbers
projplot(iLungF.asr, starty=1984, startp=2009)      # produce Fig. 4
```

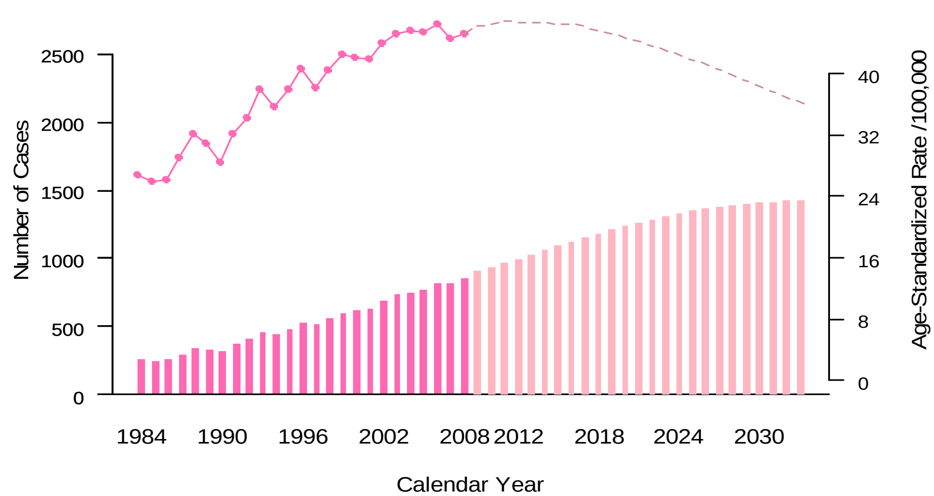

Fig. 4. ASRs and total numbers for female lung cancer incidence in 1984 –2033.

power 5 *vs* log *vs* sqrt

If other link functions are used, what will happen?

```
iLungF.obj1 <- canproj(iLungF, poyf, 2009, linkfunc="log")      # log-link
iLungF.obj2 <- canproj(iLungF, poyf, 2009, linkfunc="sqrt")     # sqrt-link
```

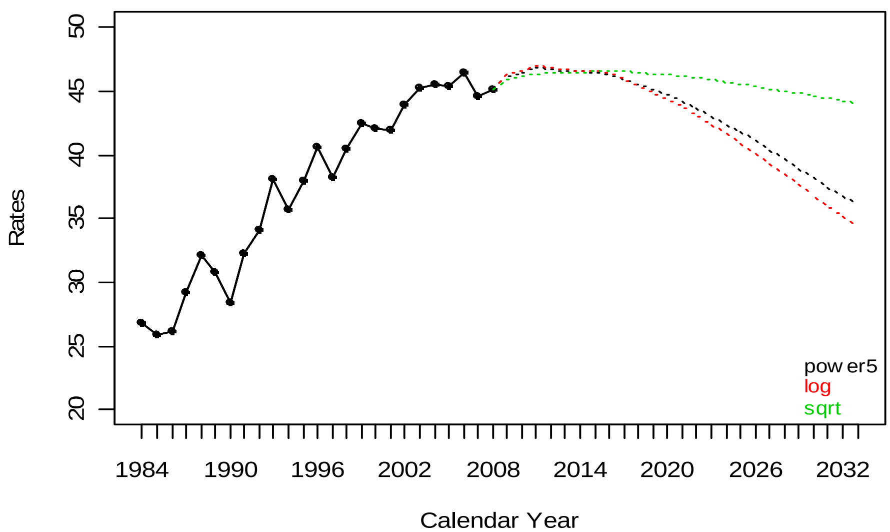

Fig. 5. Observed and projected ASRs by power5, log and sqrt link functions for female lung cancer incidence 1984-2033.

```
# Produce Fig. 5:
plot.canproj(iLungF.obj0, ca91, ylim=c(20, 50), col=c(1,1)) # setting the y-axis scale by ylim and
plot.canproj(iLungF.obj1, ca91, new=F, col=c(1,2))           # defining the colors by col
plot.canproj(iLungF.obj2, ca91, new=F, col=c(1,3))
legend("bottomright", text.col=1:3, legend=c("power5", "log", "sqrt"), bty="n", cex=0.8)
```

The downward trend is fastest if the log-link is used as the inverse function of logarithm gives the exponentially decreasing curve. When the sqrt-link is used, the canproj determined that the AC-model is the appropriate projection method, the change in age-specific rates is only due to the cohort effects evolving in the future periods, and is not dependent on any linear trend over time. The inverse function of sqrt, x^2, comparing with the inverse functions of “log” and “power5”: $\exp(x)$ and x^5, produced slowly changing in projections when $x \geq 1$.

```
> summary.canproj(iLungF.obj2)
Prediction done with:
Method:                                              Age-Cohort Model
Number of periods predicted (nopred):                5
Trend used in new cohort estimation (cuttrend):  0 , 0.25 , 0.5 , 0.75 , 1
Number of periods used in estimate (noperiod):   5
Distribution function of regression:                 Poisson
```

P-value for goodness of fit: 0.9971
First age group estimated (startestage): 7

newcohort=NULL *vs* newcohort=T

When the default of the argument newcohort is used (NULL), the new cohort effects in the next 5 period are assumed to be the cohort effect of the last observed one, i.e., cohort (C) 15 in Fig. 3, which is zero as one of the reference cohorts. Setting the newcohort=T, the new cohort effects are assumed to be that of the second last cohort, i.e. C14 in Fig. 3, which is non-zero but the difference between C14 and C15 is very small, so not too much change in projections was observed (Fig. 6).

```
iLungF.obj3 <- canproj(iLungF, poyf, 2009, newcohort=T)  # setting non-zero new cohort
#Produce Fig. 6:
plot.canproj(iLungF.obj0, ca91, ylim=c(20, 50), col=c(1,1))
plot.canproj(iLungF.obj3, ca91, new=F, col=c(1,2))
legend("bottomright", text.col=1:2, legend=c("new-c=c15=0", "new-c=c14"), bty="n", cex=0.8)
```

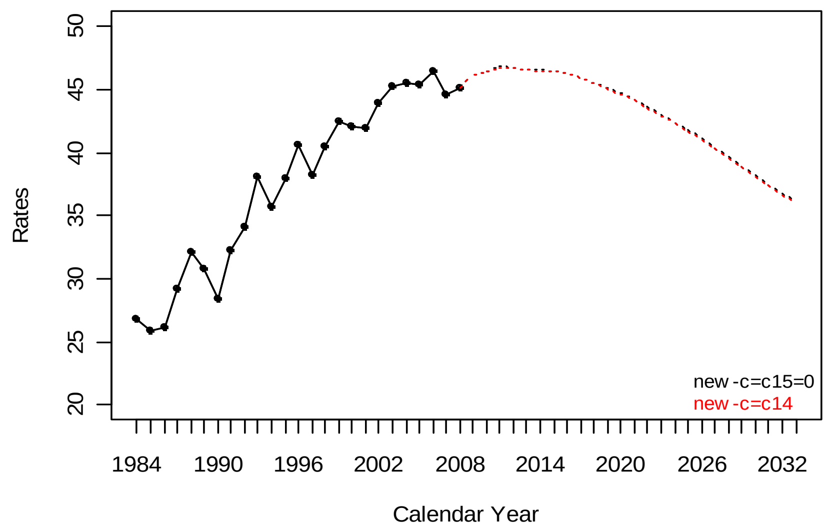

Fig. 6. Comparison of projections between zero and non-zero new cohort effects for female lung cancer incidence 2009-2033.

ncase=1 *vs* ncase = 5 (or setting startestage)

When the argument, ncase, increases from the default (1) to ncase=5, the average method was used for two more age groups: 30 – 39, and thus the trend based method was used for age 40 – 85+. Since the previously observed downward trend in age standardized rates is driven by the more recent cohort effect (i.e. among the younger population), omitting 2 of these age groups from the modeling results in the projections using ncase=5 being slightly higher than those in the projections using ncase=1 (Fig. 7). It is noted that the argument "nagg" is used when hybdproj is selected.

```
iLungF.obj4 <- canproj(iLungF, poyf, 2009, ncase=5) # set ncase=5
```
or setting
```
iLungF.obj4 <- canproj(iLungF, poyf, 2009, startestage=9) # set the youngest age as 40
summary(iLungF.obj4) # identify the smallest age groups in regression
# Produce Fig. 7:
plot.canproj(iLungF.obj0, ca91, ylim=c(20, 50), col=c(1,1))
plot.canproj(iLungF.obj4, ca91, new=F, col=c(1,2))
legend("bottomright", text.col=1:2, legend=c("ncase=1", "ncase=5"), bty="n", cex=0.8)
```

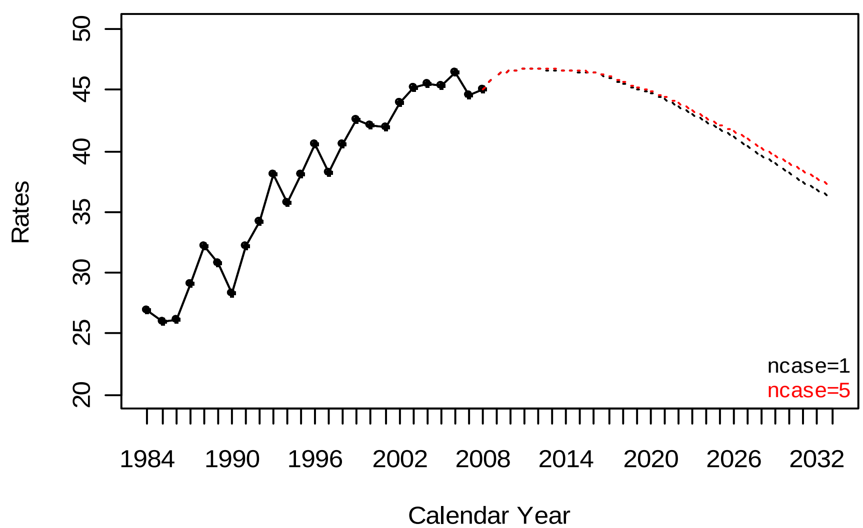

Fig. 7. Comparison of projections between ncase=1 and ncase=5 for female lung cancer incidence 2009-2033.

Changing cuttrd and / or shortp

The change of the attenuating parameter may lead to big differences in projected rates. Let the attenuating parameter be 50% in 5-year period (cuttrd=0.1) after the first 5-year period and the attenuating parameter in the first 5-year period be 25% (shortp=.05) (noting that cuttrd and shortp are defined in 1-year, instead of 5-year period):

```
iLungF.obj5 <- canproj(iLungF, poyf, 2009, cuttrd = 0.1, shortp=.05)
```

Because the increasing trend from the drift term is attenuated more than if the defaults were used, a rapid downward trend in projections would be expected (Fig. 8).

```
# Produce Fig. 8:
plot.canproj(iLungF.obj0, ca91, ylim=c(20, 50), col=c(1,1))
plot.canproj(iLungF.obj5, ca91, new=F, col=c(1,2))
legend("bottomright", text.col=1:2, legend=c("cuttrd=.05, shortp=0", "cuttrd=.1, shortp=.05"))
```

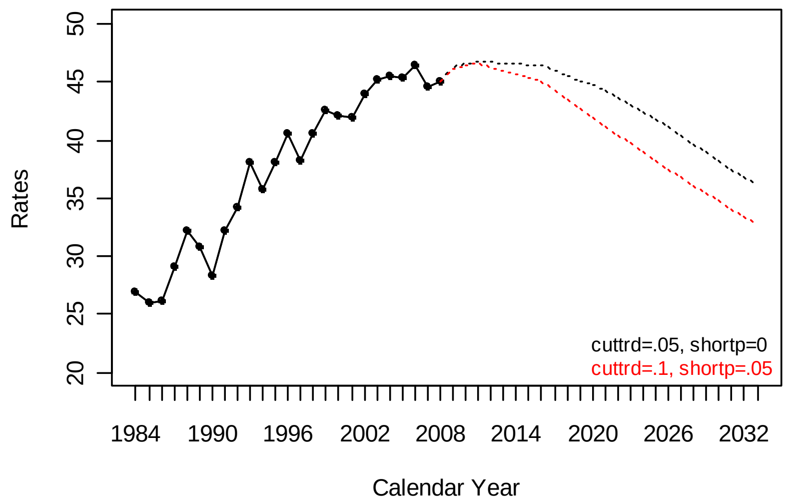

Fig. 8. Comparison of projections between cuttrd=.05 / shortp=0 and cuttrd=.1 / shortp=.05 for female lung cancer incidence 2009-2033.

Standard population

If a different standard population is used, the observed and projected ASRs will be different. However, the pattern would not be changed, because the canproj provides the projections based on the age-specific rates. See Figure 9 for the comparison between the 1991 Canadian Standard Population and The World Standard Population, which is a much younger population.

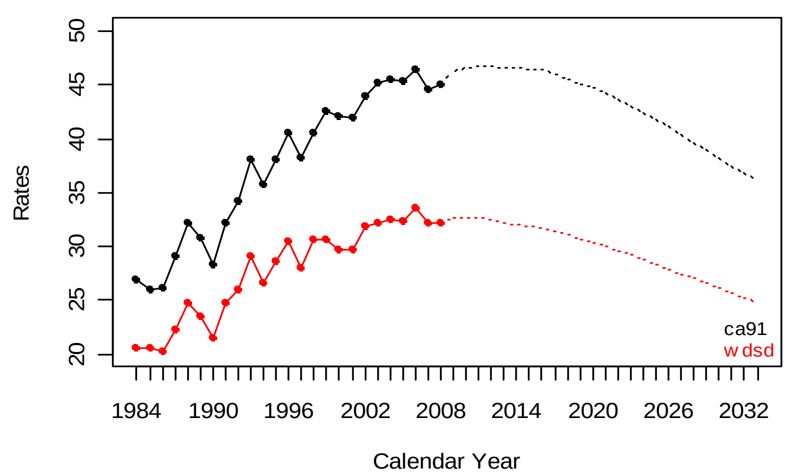

Fig. 9. Comparison of observations and projections between the use of ca91 and wdsd for female lung cancer incidence 2009-2033.

```
# Produce the object using the world standard population:
iLungF.obj6 <- canproj(iLungF, poyf, 2009, standpop=wdsd)  # replace ca91 by wdsd
# Produce Fig. 9:
```

```
plot.canproj(iLungF.obj0, ca91, ylim=c(20, 50), col=c(1,3))
plot.canproj(iLungF.obj6, wdsd, new=F, col=c(2,4))
legend("bottomright", text.col=1:2, legend=c("ca91", "wdsd"), bty="n", cex=0.8)
```

Usage of pD and pGOF

pGOF is the significance level of *p*-value for model comparison, testing over-dispersion, and the non-linear cohort effects, and pD is the significance level of the *p*-value for testing the estimated drift or slope. Changing the settings for pGOF and pD may lead to different projection methods being selected by canproj. For example, there is a higher chance of selecting the Nordpred method if the level of pGOF is increased; in particular, the adpcproj results are always used if pGOF=1; while the Hybrid method will be selected if pGOF=0, as no significant cohort effects would be identified:

```
> iLungF.obj7 <- canproj(iLungF, poyf, 2009, pD=0, pGOF=0)  #  produce the object
> summary(iLungF.obj7)                                        # summarize the method
Method:                         Hybrid approach
Number of prediction years:     25
First period cutting trend:      0
Degenerating trend per year:     0.05
Projection base (years):         19
Aggregating years (nagg):        1
Age-cases per year (ncase):     1
Model for regression:            age-specific
Link function for GLM:           power5
P-value for goodness of fit:     0.0068
Age group for regression:        7,8,9,10,11,12,13,14,15,16,17,18
Age group for average method:  1,2,3,4,5,6
```

On the other hand, a small pD will lead to a non-significant drift-term so there is a higher chance that the AC-model will be selected. For example,

```
> iLungF.obj8 <- canproj(iLungF, poyf, 2009, pD=0)   #  produce the object
> summary(iLungF.obj8)                                 # summarize the method
Method:                                          Age-Cohort Model
Number of periods predicted (nopred):            5
Trend used in new cohort estimation (cuttrend):  0 , 0.25 , 0.5 , 0.75 , 1
Number of periods used in estimate (noperiod):   5
Distribution function of regression:             Poisson
P-value for goodness of fit:                      0.1246
First age group estimated (startestage):          7
```

The components of “pdPC” in a canproj object list are the *p*-values of the drift, the non-linear period effects and the non-linear cohort effects from the age-drift-period-cohort (AdPC) analysis. For example,

```
> iLungF.obj8$pdPC
[1] 3.281912e-09 8.601364e-01 3.091421e-10
```

In the AdPC analysis for the female lung cancer incidence in 1984-2008, the *p*-value of estimated drift-term is 3.281912e-09 (<.0001), the *p*-value of estimated non-linear period effects (Type III error) is 0.8601, and the *p*-value of estimated non-linear cohort effects is 3.091421e-10 (<.0001).

Thus adjusting the combination of the settings for pGOF and pD may change the path through the Decision Tree (Fig 2) and allows the user to determine the preferred projection method for specific sites, such as prostate cancer. In general, however, the significance level 0.05 is suggested for both pGOF and pD, which are set as the defaults in the canproj package.

In this version, without bothering the pD and pGOF, the argument “methods” allows user to select anyone of the following 9 projection methods: the 5-year average (ave5), the age-only (age-only), the common trend (com-trd), the Poisson based age-specific trend (age-trd-poi), the negative-binomial based age-specific trend (age-trd-nb), the Poisson based age-cohort (ac-poi), the negative-binomial based age-cohort (ac-nb), the Poisson based age-drift-period-cohort (nordpred), the negative-binomial based age-drift-period-cohort (adpc-nb). For examples,

```
> iLungF.obj7 <- canproj(iLungF, poyf, 2009, methods=”age-trd-poi”)
> iLungF.obj8 <- canproj(iLungF, poyf, 2009, methods=”ac-poi”)
```

2.3. How to Choose Projection Methods

Users may define the method to be used by specifying methods in canproj function, or apply the sub-package directly in certain circumstance, as the 3 sub-packages are independent each other. Automatically, the canproj selects the sub-package to be used based on the AdPC analysis. For cancer site with very small numbers of cases (the average number of cases in recent 10 years is less than 3), canproj will use the 5-year average method (ave5proj) directly. In the other side, without AdPC analysis, user can define the “sound” reasonable projection method.

Usage of adpcproj

The adpcproj is the control function in the sub-package for the Nordpred method:

```
adpcproj(cdat, pdat, projfor="incidence", n5case=NULL, startestage=NULL, noperiods=NULL, recent=NULL, newcohort=NULL, pGOF=0.05, cuttrend=c(0,.25,.5,.75,1), linkfunc="power5")
```

where, n5case = 5 × ncase represents "the minimum number of cases/deaths in a age group per 5-year period"; startestage is "the youngest age group used in GLM"; noperiods is "the number of 5-year periods to be used for fitting the age-drift-period-cohort model"; recent is "to force to use the recent trend (T) or not (F) without testing the curvature"; and cuttrend is "a vector to attenuate the drift-term in projection periods" (This can be calculated by cuttrd in the canproj control function). adpcproj can be used independently of the canproj package to provide estimates based on the AdPC model even if one of the components, drift, period and cohort effect, are not significant.

Example 2. Projections for female breast cancer incidence in 2009 – 2033

Creating the projection object by

```
> iBrstF.obj0 <- adpcproj(iBrstF, poyf)
```

which is equivalent to

```
>iBrstF.obj0 <- canproj(iBrstF, poyf, 2009, methods="nordpred")
```

Although the drift estimate is not significant (p-value=0.3461):

```
> glm.adpcproj(iBrstF.obj0)$coef["Period", ]
    Estimate        Std. Error       z value          Pr(>|z|)
  0.0004474130   0.0004749289   0.9420632451   0.3461602608
```

and the non-linear cohort effects are not significant (p-value=0.0775):

```
> drop1(iBrstF.obj0$glm, test="F")["as.factor(Cohort)",  ]
                    Df    Deviance    AIC      F-value    Pr(F)
as.factor(Cohort)   15    83.008      625.21   1.8042     0.07754
```

we can use the projected results from the adpcproj function and plot the ASRs as in Fig. 10:

```
> par(cex=0.8, mar=c(4.5,4.5,2,2))   # adjust the font and margins in graph
> plot.adpcproj(iBrstF, poyf, 2009, iBrstF.obj0, ca91, ylim=c(60, 120))  # Produce Fig. 10
```

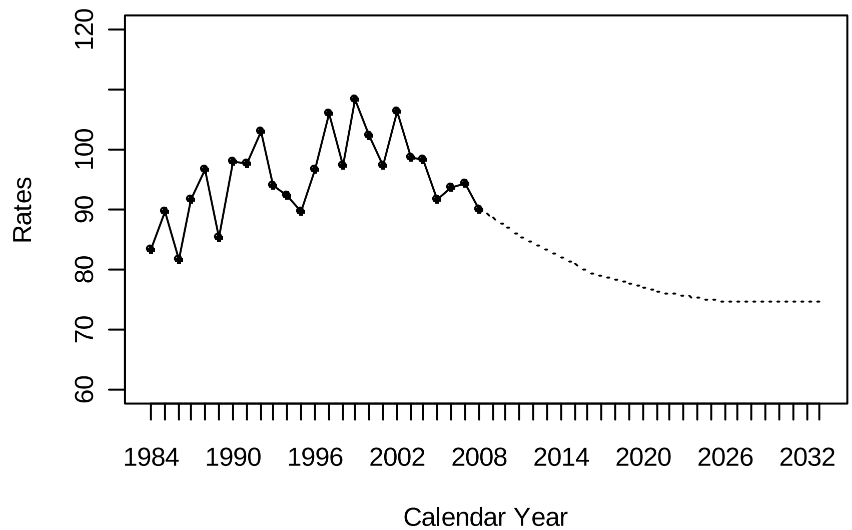

Fig. 10. Observed and projected ASRs of female breast cancer incidence in 1984 - 2033 based on AdPC model.

Noting that the non-linear cohort effect is marginally insignificant, the same results can be obtained from the canproj function by setting pD=0.5, pGOF=0.1. In the adpcproj sub-package, however, the setting of the pGOF is for the choice of distribution function (Users can specify methods in canproj for the choice), e.g, if the setting pGOF=1, the negative-binomial distribution will be used:

```
> iBrstF.obj1<- adpcproj(iBrstF, poyf, pGOF=1) # or canproj(iBrstF,poyf,2009,methods="adpc-nb")
> summary.adpcproj(iBrstF.obj1)
Method:                                      Age-drift-Period-Cohort Model
Number of periods predicted (nopred):        5
Trend used in predictions (cuttrend):         0 , 0.25 , 0.5 , 0.75 , 1
Number of periods used in estimate (noperiod): 5
Distribution function of regression:          Negative-binomial
P-value for goodness of fit:                  0.2503
Used recent (recent):                         TRUE
P-value for recent:                           0
First age group used (startuseage):           6
First age group estimated (startestage):      6
```

On the other hand, setting pGOF = 0 leads to the selection of Poisson distribution, and thus, the adpcproj package becomes the original nordpred package. Other arguments, such as recent, cuttrend and linkfunc, can be changed by user. Please refer to the canproj manual in Chapter 3, Section 3.2 for details.

Usage of acproj

The acproj is the control function in the sub-package for the AC-model method:

```
acproj(cdat,pdat, projfor="incidence", n5case=NULL, startestage=NULL, noperiods=NULL,
                    cuttrend=c(0, .25, .5, .75, 1), pGOF=0.05, linkfunc="power5")
```

This can be used directly, when the drift-term (the linear trend) and the non-linear period effects are either not reliable or not significant but cohort effects are significant in the AdPC analysis, or where previous experience indicates that the AC-model is appropriate. The projection object is created by

```
> iBrstF.obj2 <- acproj(iBrstF, poyf)  # or canproj(iBrstF, poyf, 2009, methods="ac-nb")
```

Comparing with the results from adpcproj sub-package, the AC-model method gave a more stable trend in projected rates (see Fig. 11) due to the absence of the decreasing drift (trend):

```
# Produce Fig. 11:
plot.adpcproj(iBrstF, poyf, 2009, iBrstF.obj0, ca91, ylim=c(60, 120), col=c(1,1))
plot.acproj(iBrstF, poyf, 2009, iBrstF.obj2, ca91, new=F, col=c(1,2))
legend("bottomright", text.col=1:2, legend=c("adpcproj", "acproj"), bty="n", cex=0.8)

# If user defined canproj methods:
#plot.canproj(iBrstF.obj0, ca91, ylim=c(60, 120), col=c(1,1))
#plot.canproj(iBrstF.obj2, ca91, new=F, col=c(1,2))
```

Other arguments, such as cuttrend and linkfunc, can be changed by user and produce different projections, noting that there is no "recent" argument in the acproj sub-package. Please refer to the canproj manual in Chapter 3, Section 3.3 for details.

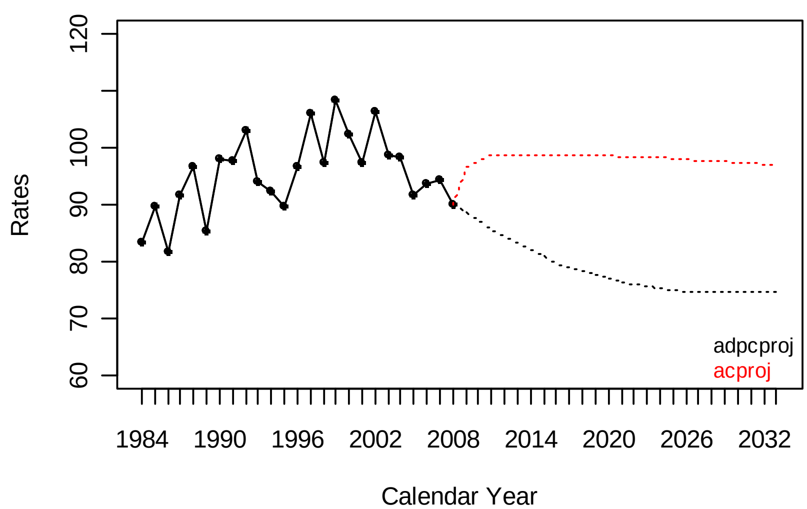

Fig. 11. Comparison between adpcproj and acproj sub-packages for projections of breast cancer incidence.

Usage of hybdproj and ave5proj

The hybdproj is the control function in the sub-package for the Hybrid method:

```
hybdproj(cdat, pdat, projfor="incidence", nagg=NULL, ncase=NULL,
                        cuttrd=0.05, shortp=0, linkfunc="power5", pD=0.05, pGOF=0.05)
```

where the argument "nagg" determines the number of years for data aggregation. The sub-package can be used directly, when the cohort effects are absent in the AdPC analysis, or previous experience indicate that this is the appropriate model. There are 4 candidate models in the hybdproj and statistical testing approaches are adopted automatically for the model selection. Users may select their preferred model by setting appropriate critical values for pGOF and pD or specify methods in canproj.

Creating the projection object by

```
> iBrstF.obj3 <- hybdproj(iBrstF, poyf)  # Produce hybdproj object
```

Or # iBrstF.obj3 <- canproj(iBrstF, poyf, 2009, methods="age-only")

The hybdproj function selected the average method using data in 1989-2008, and gave the constant ASRs in projection. These results are compared to those from the adpcproj and the acproj sub-packages (Fig. 12):

```
> summary.hybdproj(iBrstF.obj3)
Method:                        Hybrid approach
Number of prediction years:    25
First period cutting trend:     0
Degenerating trend per year:    0.05
Projection base (years):        20
Aggregating years (nagg):       1
Age-cases per year (ncase):     1
Model for regression:           average
Link function for GLM:          power5
P-value for goodness of fit:    0
Age group for regression:       6,7,8,9,10,11,12,13,14,15,16,17,18
Age group for average method:   1,2,3,4,5
```

```
# Produce Fig. 12:
plot.adpcproj(iBrstF, poyf, 2009, iBrstF.obj0, ca91, ylim=c(60, 120), col=c(1,1))
plot.acproj(iBrstF, poyf, 2009, iBrstF.obj2, ca91, new=F, col=c(1,2))
plot.hybdproj(iBrstF, poyf, 2009, iBrstF.obj3, ca91, new=F, col=c(1,3))
legend("bottomright", text.col=1:3, legend=c("adpcproj", "acproj", "hybdproj"), bty="n", cex=0.8)

# If user defined canproj methods:
# plot.canproj(iBrstF.obj0, ca91, ylim=c(60, 120), col=c(1,1))
# plot.canproj(iBrstF.obj2, ca91, new=F, col=c(1,2))
```

plot.canproj(iBrstF.obj3, ca91, new=F, col=c(1,3))

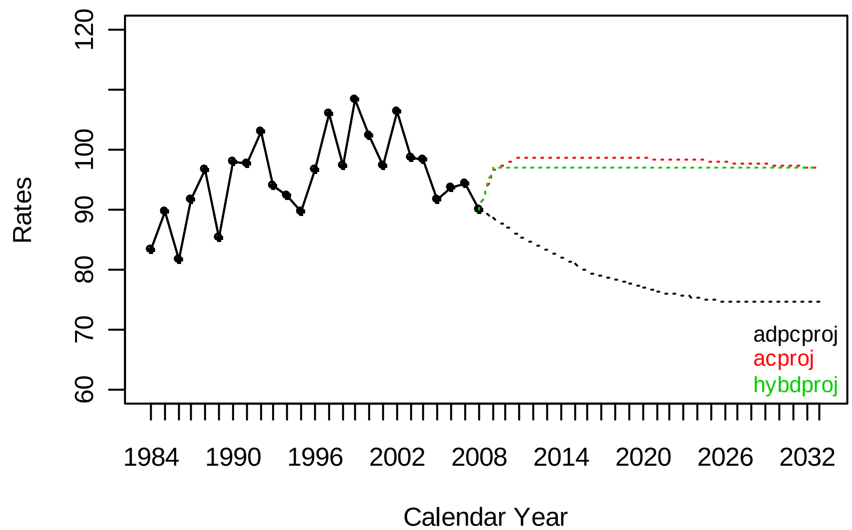

Fig. 12. Comparison between adpcproj, acproj and hybdproj sub-packages for projections of breast cancer incidence in 2009-2033.

The results by the average method in hybdproj would be different from those from the 5-year average method "ave5proj" (ave5proj(cdat, pdat, startp, sum5=NULL). refer to Chapter 3, Section 3.5) because the average from hybdproj is based on the full projection base (in this instance 20 years) and the one from ave5proj is based on recent 5 years (see Fig. 13). In addition, the ave5proj allows user select the method to do "average" over the recent 5 years: sum5=NULL (default) calls the age-specific rates over the 5-year period; sum5=T averages the yearly age-specific rates over the 5 years. Again, users can specify methods in canproj to conduct the sub-package "ave5proj".

```
# Produce ave5proj object:
# iBrstF.obj4 <- ave5proj(iBrstF, poyf, startp=2009, sum=NULL)
# iBrstF.obj5 <- ave5proj(iBrstF, poyf, startp=2009, sum5=T)
# or
> iBrstF.obj4 <- canproj(iBrstF, poyf, 2009, methods="ave5")
> iBrstF.obj5 <- canproj(iBrstF, poyf, 2009, methods="ave5", sum5=T)

# Produce Fig. 13:
plot.canproj(iBrstF.obj3, ca91, ylim=c(60, 120), col=c(1,1))
plot.canproj(iBrstF.obj4, ca91, new=F, col=c(1,2))
plot.canproj(iBrstF.obj5, ca91, new=F, col=c(1,3))
legend("bottomright", text.col=1:3, legend=c("age-only", "ave5-period", "ave5-yearly"), bty="n",
cex=0.8)
```

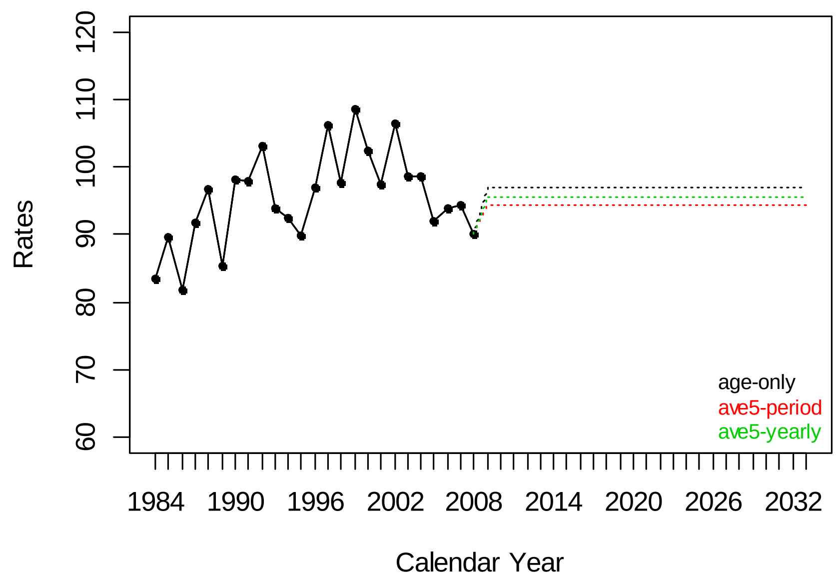

Fig. 13. Comparison between hybdproj and ave5proj sub-packages for projections of breast cancer incidence in 2009-2033.

If setting Ave5=T in canproj or hybdproj.getproj, the projection results will be obtained from the ave5proj, when the age-only model is selected:

```
> iBrstF.obj3 <- hybdproj(iBrstF, poyf)
> iBrstF.obj3.asr <- hybdproj.getproj(iBrstF, poyf, 2009, iBrstF.obj3, ca91, Ave5=T)
> iBrstF.obj6 <- canproj(iBrstF, poyf, 2009, Ave5=T)
> iBrstF.obj7 <- canproj(iBrstF, poyf, 2009, Ave5=T, sum5=T)
```

Setting different arguments in hybdproj from the defaults may lead to a different selection of the final model. For example, if the data is aggregated by 2 years, instead of the default (nagg = 1 based on the average of ASRs), the common-trend model will be chosen:

```
> iBrstF.obj8 <- hybdproj(iBrstF, poyf, nagg=2)   # Produce the object by setting nagg=2
> summary.hybdproj(iBrstF.obj5)
Method:                          Hybrid approach
Number of prediction years:      24
First period cutting trend:       0
Degenerating trend per year:      0.05
Projection base (years):          14
Aggregating years (nagg):          2
Age-cases per year (ncase):        1
Model for regression:              common-trend
Link function for GLM:              power5
P-value for goodness of fit:       0
Age group for regression:          6,7,8,9,10,11,12,13,14,15,16,17,18
Age group for average method:  1,2,3,4,5
```

This can be done by

```
> iBrstF.obj8 <- canproj(iBrstF, poyf, 2009, nagg=2, methods="com-trd")
> summary(iBrstF.obj8)
```

```
#Produce Fig. 14:
plot.hybdproj(iBrstF, poyf, 2009, iBrstF.obj3, ca91, ylim=c(60, 120), col=c(1,1))
plot.canproj(iBrstF.obj8, ca91, new=F, col=c(1,2))
legend("bottomright", text.col=1:2, legend=c("nagg=1", "nagg=2"), bty="n", cex=0.8)
```

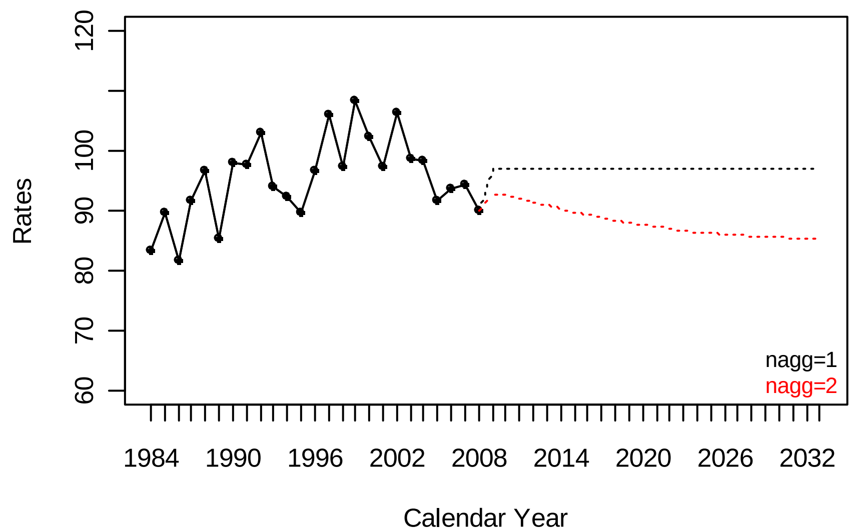

Fig. 14. Comparison between nagg=1 and nagg=2 for projections of breast cancer incidence in 2009-2033.

Other arguments, such as ncase, cuttrd, linkfunc, pD and pGOF, can be changed by user and produce different projections. Please refer to the canproj manual in Chapter 3, Section 3.4 for details.

Chapter 3. The Canproj Manual

3.1. Canproj Functions

canproj *Projection by fitting age, period and/or cohort generalized linear models*

Description

Produce annual projection object by selecting the candidate method from the 3 sub-packages: adpcproj, acproj and hybdproj by testing whether the estimated drift term and non-linear cohort effects in age-drift-period-cohort Poisson regression model are statistically significant. The hybdproj is used if insignificant cohort effect; the acproj is used if significant cohort effect but insignificant drift term; and the adpcproj is used when both cohort effect and drift term are significant.

Usage

```
canproj(cdat, pdat, startp,
        projfor="incidence", nagg=NULL, ncase=NULL, startestage=NULL, newcohort=NULL,
        Ave5=NULL, sum5=NULL, methods=NULL,
        linkfunc="power5", cuttrd=0.05, shortp=0, pD=0.05, pGOF=0.05, standpop=ca91)
```

Arguments

cdat — A data frame for annual numbers of cancer cases/deaths with the rows representing the 18 5-year age groups in ascending order and the N columns representing the calendar years in ascending order where $15 \le N \le 125$.

pdat — A data frame for annual observed and projected population with the rows representing the 18 5-year age groups in ascending ordered and the $N+M$ columns representing the calendar years in ascending order where M is the number of projection years, $5 \le M \le 25$.

startp — Specify the starting calendar year of projection to format appropriate outputs, e.g. startp = 2009.

projfor — Specify "incidence" or "mortality" projection to determine the default settings of "nagg" and "ncase" (see below). Default is projfor="incidence".

nagg — Number of years for data aggregation to stabilize rates, nagg=5 is fixed for adpcproj and acproj. If nagg is not specified (NULL), when projfor="incidence", nagg=1 if the averaged asir > 15, nagg=2 if 10 < asir ≤ 15, nagg=3 if 5 < asir ≤ 10 and nagg=4 if asir ≤ 5; when projfor="mortality", nagg=1 if asmr > 9, nagg=2 if 6 < asmr ≤ 9, nagg=3 if 3 < asmr ≤ 6 and nagg=4 if asmr ≤ 3.

ncase	Minimum number of cancer cases/deaths per year for splitting data by age group into two datasets: one for average method, the other for regression modeling. When ncase is not specified (NULL), 1 is used for projfor="incidence", and 3/5 is used for projfor = "mortality".
startestage	The youngest age group which is included in the GLM for adpcproj or acproj; if startestage=NULL (default), the startestage will be determined based on ncase.
newcohort	Used in adpcproj only to determine the setting of "startuseage" in the original Nordpred, Specifying newcohort=T gives startuseage=startestage+1 and hence the new cohort effects is the estimation of the second last cohort; the default (NULL) assumes startuseage=startestage and thus the new cohort effects is zero.
Ave5	Used in hybdproj only to determine if the 5-year average method is used when age-only model is selected. The default (NULL) chose the projection results from the age-only model, otherwise (Ave5=T) gives the projection results from the 5-year average method.
sum5	Used in ave5proj only to determine how to average the 5-year rates. The default (NULL) calculates the age-specific rates using the aggregated numbers and population sizes, respectively, in the recent 5 years; otherwise (sum5=T), the yearly age-specific rates will be averaged over the 5 years.
methods	Used for user selected projection method. The default (NULL) calls the decision tree to select the projection model; otherwise, methods="nordpred" selects the nordpred model; methods ="adpc-nb" selects the negative-binomial distribution based nordpred model; methods="ac-poi" selects the Poisson distribution based age-cohort model; methods ="ac-nb" selects the negative-binomial distribution based age-cohort model; methods ="age-trd-nb" selects the negative-binomial distribution based age-specific trend model; methods ="age-trd-poi" selects the Poisson distribution based age-specific trend model; methods ="com-trd" selects the common trend model; methods ="age-only" selects the age-only model; and methods ="ave5" selects the 5-year average method.
linkfunc	Specify the link function to use in the GLM. Default is "power5", can be specified as "log", "sqrt" and "identity".
cuttrd	Attenuation percent for drift term or slope per year, default is 0.05 per year after the first 5 projection years.
shortp	Attenuation percent of drift term or slope for the first 5 projection year, default is zero.
pD	Significance level of *p*-value for estimated drift term in adpcproj or slope in hybdproj, default is 0.05.
pGOF	Significance level of *p*-value for model selection based on their goodness-of-fits, default is 0.05. User can call the original Nordpred by setting pGOF=0 and pD=1.

standpop — A vector with the weights (proportions) of 18 age groups in a standard population. Default is based on the 1991 Canadian standard population (ca91). Alternatively, user can choose `ca96' (the 1996 Canadian standard population), `wdsd' (World standard population), or define a vector for other standard population.

Value

An object of class "canproj" — a list with components:

annproj — A data frame with annual observed and projected age-standardized rates and numbers of cancer incidence or mortality, ordered by calendar year in row.

agsproj — A data frame with annual observed and projected age-specific rates, ordered by age group (in row) and calendar year (in column).

method — sub-package used for projection based on the decision tree, "ADPC" represents the modified Nordpred package, "AC" represents the AC-model package and "Hybrid" is the Hybrid package; If user define the projection method, the output is one of the 9 projection methods. Detailed information can be found by summary.canproj.

out — A list with detailed information on model fitting, projection base or recent trend, age group splitting and projection methods; when the 5 year average method is selected, the output is "NA".

Obsy — Number of years for observed cancer data.

pdPC — A vector listed *p*-values of the estimated drift-term, the non-linear period effects and the non-linear cohort effects from the age-drift-period-cohort model using historical data; when user selects the projection methods, the output is "NA".

See Also

canproj.getproj, glm.canproj, plot.canproj and summary.canproj

---

canproj.getproj — *Extract the projection results*

Description

Uses a canproj.object to extract the observed and projected annual age-specific rates, or age-standardized rates and annual numbers

Usage

```
canproj.getproj(canproj.object, standpop=NULL)
```

Arguments

canproj.object	An object based on canproj() function
standpop	Default (NULL) gives the age-specific rates. Otherwise, annual age-standardized rates and total numbers will be produced.

Value

A data frame containing annual observed and projected age-specific rates, or annual age-standardized rates and numbers

---

glm.canproj *Summarize the estimations from the final model*

Description

Uses a canproj.object to summarize the estimations from the final selected model (except the 5-year average method)

Usage

glm.canproj(canproj.object)

Arguments

canproj.object	An object based on canproj() function

Value

A summary table from the glm.object

---

plot.canproj *Produce the graph of the observed and projected age-standardized rates*

Description

Uses a canproj.object to produce the graph of the observed and projected age-standardized rates against the calendar year

Usage

plot.canproj(canproj.object, standpop, new=T, ...)

Arguments

canproj.object	An object based on canproj() function
standpop	Specify a vector with the weights (proportions) of 18 age groups in a standard population to get the annual age-standardized rates.

New	Produce graph in another graphic window by default, or in the same window by setting new=F.
…	Other "par" parameters such as ylab, ylim, type, col etc. can be applied in this function

---

summary.canproj *Summarize the information on projection method used*

Description
Uses a canproj.object to summarize the information on projection method used

Usage
summary.canproj(canproj.object)

Arguments
canproj.object An object based on canproj() function

Value
An information table describing the projection method used

3.2. Adpcproj Functions

---

adpcproj *Projection by fitting age-drift-period-cohort generalized linear model*

Description
Produce projections by the age-drift-period-cohort generalized linear model. This is a revision of the original nordpred package by introducing negative binomial distribution when over-dispersion occurs in the Poisson regression model, and additional link functions of \`sqrt' and \`identity' beside the \`power5' and \`log' link functions.

Usage
```
adpcproj (cdat, pdat, projfor="incidence",
        n5case=NULL, noperiods=NULL, recent=NULL, startestage=NULL, newcohort=NULL,
        pGOF=0.05, cuttrend=c(0,.25,.5,.75,1), linkfunc="power5")
```

Arguments

cdat — A data frame for annual numbers of cancer cases/deaths with the rows representing the 18 5-year age groups in ascending order and the N columns representing the calendar years in ascending order where $15 \le N \le 125$.

pdat — A data frame for annual observed and projected population with the rows representing the 18 5-year age groups in ascending ordered and the $N+M$ columns representing the calendar years in ascending order where M is the number of projection years, $5 \le M \le 25$.

projfor — Specify "incidence" or "mortality" projection to determine the default settings of "n5case" (see below). Default is projfor="incidence".

n5case — Minimum number of cancer cases/deaths per 5 years for splitting data by age group into two datasets: one for average method, the other for regression modeling (n5case = 5*ncase in canproj). Default is n5case=5 for projfor="incidence", and n5case=3 for projfor="mortality".

noperiods — A list of candidate number of periods (in 5-year) in projection base, use a fixed number to force a specific projection base; otherwise (the default), goodness-of-fit test is used to determine if the ancient periods are removed from projection base.

recent — Specify the estimated drift term from recent trend (recent=T) or whole trend (recent=F). If recent=NULL (default), the choice is based on the comparison between the age-drift-cohort model and the model with additional quadratic term of the drift.

startestage — The youngest age group which is included in the GLM; if startestage=NULL (default), the startestage will be determined based on n5case.

newcohort — Determine "startuseage". Specifying newcohort=T gives startuseage=startestage+1 and hence the new cohort effects is the estimation of the second last cohort; the default (NULL) assumes startuseage=startestage and thus the new cohort effects is zero.

pGOF — Significance level of p-value to select the distribution between Poisson and negative-binomial, default is 0.05.

cuttrend — A vector to attenuate drift term, default is 0, 25%, 50%, 75% and 100% for future 5-year periods.

linkfunc — The link function to be used in the GLM, default is "power5", can be specified as "log", "sqrt" and "identity".

Value

An object of class "adpcproj" — a list with components:

predictions — A data frame with projected number of cases in 5-year period

pyr	A data frame with observed and projected population
nopred	Number of 5-year periods in projection years
noperiod	Number of 5-year periods used in historical data
gofpvalue	The *p*-value of goodness of fit for the age-drift-period-cohort model
recent	The indicator of the estimated drift term from recent trend (recent=T) or whole trend (recent=F)
pvaluerecent	The *p*-value of the test for the use of recent trend
cuttrend	A vector to attenuate drift term
distribution	Distribution used in the GLM
startuseage	The youngest age group which uses the estimates from the GLM to do projection
startestage	The youngest age group which have been included in the GLM
glm	The glm.object of the fitted GLM

See Also
adpcproj.estimate, adpcproj.prediction, adpcproj.getpred, adpcproj.getproj, glm.adpcproj, plot.adpcproj and summary.adpcproj

---

adpcproj.estimate *Fit age-drift-period-cohort models*

Description
Estimate parameters in the age-drift-period-cohort GLM based on appropriate distribution and test if "recent" trend should be used.

Usage
```
adpcproj.estimate(cases, pyr, noperiod, startestage, pGOF=0.05, linkfunc="power5")
```

Arguments

cases	A data frame with number of cases in 5-year period by ascending ordered 18 age groups in row
pyr	A data frame with observed and projection population size in 5-year period by ascending ordered 18 age groups in row
noperiod	Number of 5-year periods used in historical data

startestage	The youngest age group which is included in the GLM
pGOF	Significance level of *p*-value to select the distribution between Poisson and negative-binomial, default is 0.05.
linkfunc	The link function to be used in the GLM, default is "power5", can be specified as "log", "sqrt" and "identity".

Value

An object of class "adpcproj.estimate" — a list with components:

glm	The glm.object of the fitted GLM
cases	A data frame with number of cases in 5-year period by ascending ordered 18 age groups in row
pyr	A data frame with observed and projection population size in 5-year period by ascending ordered 18 age groups in row
noperiod	Number of 5-year periods used in historical data
linkfunc	The link function used in the GLM.
startestage	The youngest age group which have been included in the GLM
distribution	Indicate Poisson or negative-binomial distribution used in the GLM.
gofpvalue	The *p*-value of the test for the goodness-of-fit of the GLM
suggestionrecent	Indicate the suggestion on the use of the recent trend for the drift term, if suggestionrecent=F, the recommendation is to use whole trend; otherwise, the recommendation is to use the drift based on the recent 10 years.
pvaluerecent	The *p*-value of the test for the use of recent trend

---

adpcproj.prediction *Extrapolate estimated trend from age-drift-period-cohort model*

Description

Calculating age-specific projections using an `adpcproj.estimate' object which is based on an age-drift-period-cohort model

Usage

adpcproj.prediction(adpcproj.estimate.object, startuseage, recent, cuttrend=c(0,.25,.5,.75, 1))

Arguments

adpcproj.estimate.object	A glm.object based on 'adpcproj.estimate()'
startuseage	The youngest age group which uses the estimates from the GLM for projection
recent	The indicator of the estimated drift term from recent trend (recent=T) or whole trend (recent=F)
cuttrend	A vector to attenuate drift term

Value

An object of class "adpcproj" — a list with components:

predictions	A data frame with projected number of cases in 5-year period
pyr	A data frame with observed and projected population in 5-year period
nopred	Number of 5-year periods in projection years
noperiod	Number of 5-year periods used in historical data
gofpvalue	The *p*-value of goodness of fit for the age-drift-period-cohort model
recent	The indicator of the estimated drift term from recent trend (recent=T) or whole trend (recent=F)
pvaluerecent	The *p*-value of the test for the use of recent trend
cuttrend	A vector to attenuate drift term
distribution	Distribution used in the GLM
startuseage	The youngest age group which uses the estimates from the GLM to do projection
startestage	The youngest age group which have been included in the GLM
glm	The glm.object of the fitted GLM

---

adpcproj.getpred *Extract the projection results by 5-year period*

Description

Uses an adpcproj.object to extract the observed and projected annual age-specific rates, or age-standardized rates and annual numbers in 5-year period

Usage

```
adpcproj.getpred(adpcproj.object,
                 incidence=T, standpop=NULL, excludeobs=F,
                 byage, agegroups="all")
```

Arguments

adpcproj.object	An object based on the 'adpcproj()' or 'adpcproj.prediction()' function
incidence	Indicating whether to give rates or numbers of cases
standpop	A vector with the weights (proportions) of 18 age groups in a standard population, default (NULL) gives the age-specific rates. Age-standardized rates and total numbers are given by specifying standpop, e.g. standpop=ca91.
excludeobs	Indicates whether to output observed values, default (F) does not output observed values.
byage	Report numbers by age groups. If byage=F, crude or age-standardized rates are given.
agegroups	Specify age groups to include. E.g. agegroups=7:18 include age groups seven to eighteen. The default ('All') includes all 18 age groups.

Value

A data frame containing observed and / or projected rates and / or numbers in 5-year period

---

adpcproj.getproj *Extract the annual projection results*

Description

Apply the segmental linear interpolation function "asrpy" on an adpcproj.object to extract the observed and projected annual age-specific rates, or annual age-standardized rates and annual numbers

Usage

```
adpcproj.getproj(cdat, pdat, startp, adpcproj.object, standpop=NULL)
```

Arguments

cdat	A data frame for annual numbers of cancer cases/deaths with the rows representing the 18 5-year age groups in ascending order and the N columns representing the calendar years in ascending order where $15 \le N \le 125$.

pdat	A data frame for annual observed and projected population with the rows representing the 18 5-year age groups in ascending ordered and the $N + M$ columns representing the calendar years in ascending order where M is the number of projection years, $5 \le M \le 25$.
startp	Specify the starting calendar year of projection to format appropriate outputs, e.g. startp=2009.
adpcproj.object	An object based on the 'adpcproj()' or 'adpcproj.prediction()' function
standpop	A vector with the weights (proportions) of 18 age groups in a standard population, default (NULL) gives the age-specific rates. Age-standardized rates and total numbers are given by specifying standpop, e.g. standpop=ca91.

Value
A data frame containing annual observed and projected age-specific rates, or annual age-standardized rates and numbers

---

glm.adpcproj *Summarize estimations from the age-drift-period-cohort model*

Description
Using an adpcproj.object to summarize the estimations from the age-drift-period-cohort model

Usage
glm.adpcproj(adpcproj.object)

Arguments
adpcproj.object An object based on the 'adpcproj()' or 'adpcproj.prediction()' function

Value
A summary table from the glm.object

---

plot.adpcproj *Produce the graph of the observed and projected age-standardized rates*

Description
Using an adpcproj.object to plot the observed and projected age-standardized rates against the calendar year

Usage
plot.adpcproj(cdat, pdat, startp, adpcproj.object, standpop, startplot=1, new=T...)

Arguments

cdat	A data frame for annual numbers of cancer cases/deaths with the rows representing the 18 5-year age groups in ascending order and the N columns representing the calendar years in ascending order where $15 \leq N \leq 125$.
pdat	A data frame for annual observed and projected population with the rows representing the 18 5-year age groups in ascending ordered and the $N + M$ columns representing the calendar years in ascending order where M is the number of projection years, $5 \leq M \leq 25$.
startp	Specify the starting calendar year of projection to format appropriate outputs, e.g. startp=2009.
adpcproj.object	An object based on the 'adpcproj()' or 'adpcproj.prediction()' function
standpop	A vector with the weights (proportions) of 18 age groups in a standard population
startplot	Indicating the starting year for the plot, default (1) gives the plot starting from the first observed year to the end of projection year.
new	Produce graph in another graphic window by default, or in the same window by setting new=F.
…	Other "par" parameters such as ylab, ylim, type, col etc. can be applied in this function.

---

summary.adpcproj *Summarize the information on projection method used*

Description

Uses an adpcproj.object to summarize the information on projection method used

Usage

```
summary.adpcproj(adpcproj.object, printpred=F, printcall=F, digits=0)
```

Arguments

adpcproj.object	An object based on the 'adpcproj()' or 'adpcproj.prediction()' function
printpred	Indicating whether to print the observed and predicted number of cases, default (F) does not print them out.
printcall	Indicating whether to print the function `Call' for the adpcproj.object, default (F) does not print it out.
digits	Specify the number of digits in the tabulation, default (0) keeps integer only.

Value
An information table describing the projection method used

3.3. Acproj Functions

---

acproj *Projection by fitting age-cohort generalized linear model*

Description
Produce projections by the age-cohort generalized linear model. The new cohort effects are estimated from the linear predictor of the estimated cohort effects from the GLM with at least 3 observations.

Usage
```
acproj (cdat, pdat, projfor="incidence",
        n5case=NULL, startestage=NULL, noperiods=NULL,
        cuttrend=c(0,.25,.5,.75,1), pGOF=0.05, linkfunc="power5")
```

Arguments

cdat	A data frame for annual numbers of cancer cases/deaths with the rows representing the 18 5-year age groups in ascending order and the N columns representing the calendar years in ascending order where $15 \leq N \leq 125$.
pdat	A data frame for annual observed and projected population with the rows representing the 18 5-year age groups in ascending ordered and the $N+M$ columns representing the calendar years in ascending order where M is the number of projection years, $5 \leq M \leq 25$.
projfor	Specify "incidence" or "mortality" projection to determine the default settings of "n5case" (see below). Default is projfor="incidence".
n5case	Minimum number of cancer cases/deaths per 5 years for splitting data by age group into two datasets: one for average method, the other for regression modeling (n5case = 5*ncase in canproj). Default is n5case=5 for projfor="incidence", and n5case=3 for projfor="mortality".
startestage	The youngest age group which is included in the GLM; if startestage=NULL (default), the startestage will be determined based on n5case.
noperiods	The number of periods (in 5-year) in projection base

cuttrend	A vector to attenuate the estimated slope from the linear model for estimated cohort effects in historical data, default is 0, 25%, 50%, 75% and 100% for future 5-year periods.
pGOF	Significance level of *p*-value to select the distribution between Poisson and negative-binomial, default is 0.05.
linkfunc	The link function to be used in the GLM, default is "power5", can be specified as "log", "sqrt" and "identity".

Value
An object of class "acproj" — a list with components:

predictions	A data frame with projected number of cases in 5-year period
pyr	A data frame with observed and projected population in 5-year period
nopred	Number of 5-year periods in projection years
noperiod	Number of 5-year periods used in historical data
cuttrend	A vector used for attenuating the estimated slope from the linear model for estimated cohort effects in historical data, default is 0, 25%, 50%, 75% and 100% for future 5-year periods.
gofpvalue	The *p*-value of goodness of fit for the age-cohort model
distribution	Distribution used in the GLM
startestage	The youngest age group which have been included in the GLM
glm	The glm.object of the fitted GLM

See Also
acproj.estimate, acproj.prediction, acproj.getpred, acproj.getproj, glm.acproj, plot.acproj and summary.acproj

---

acproj.estimate *Fit age-cohort models*

Description
Estimate parameters in the age-cohort GLM based on appropriate distribution.

Usage
```
acproj.estimate(cases, pyr, noperiod, startestage, pGOF=0.05, linkfunc="power5")
```

Arguments

cases	A data frame with number of cases in 5-year period by ascending ordered 18 age groups in row
pyr	A data frame with observed and projection population size in 5-year period by ascending ordered 18 age groups in row
noperiod	Number of 5-year periods used in historical data
startestage	The youngest age group which is included in the GLM
pGOF	Significance level of *p*-value to select the distribution between Poisson and negative-binomial, default is 0.05.
linkfunc	The link function to be used in the GLM, default is "power5", can be specified as "log", "sqrt" and "identity".

Value

An object of class "acproj.estimate" — a list with components:

glm	The glm.object of the fitted GLM
cases	A data frame with number of cases in 5-year period by ascending ordered 18 age groups in row
pyr	A data frame with observed and projection population size in 5-year period by ascending ordered 18 age groups in row
maxc	The number of the last cohort
midc	The number of the middle cohort
noperiod	Number of 5-year periods used in historical data
linkfunc	The link function used in the GLM.
startestage	The youngest age group which have been included in the GLM
distribution	Indicate Poisson or negative-binomial distribution used in the GLM.
gofpvalue	The *p*-value of the test for the goodness-of-fit of the GLM

acproj.prediction *Extrapolate estimated trend from age-cohort model*

Description

Calculating age-specific projections in 5-year period using an `acproj.estimate' object which is based on an age-cohort model

Usage
acproj.prediction(acproj.estimate.object, cuttrend=c(0,.25,.5,.75, 1))

Arguments

acproj.estimate.object	A glm.object based on 'acproj.estimate()'
cuttrend	A vector used for attenuating the estimated slope from the linear model for estimated cohort effects in historical data, default is 0, 25%, 50%, 75% and 100% for future 5-year periods.

Value
An object of class "acproj" — a list with components:

predictions	A data frame with projected number of cases in 5-year period
pyr	A data frame with observed and projected population in 5-year period
nopred	Number of 5-year periods in projection years
noperiod	Number of 5-year periods used in historical data
cuttrend	A vector used for attenuating the estimated slope from the linear model for estimated cohort effects in historical data, default is 0, 25%, 50%, 75% and 100% for future 5-year periods.
gofpvalue	The *p*-value of goodness of fit for the age-cohort model
distribution	Distribution used in the GLM
startestage	The youngest age group which have been included in the GLM
glm	The glm.object of the fitted GLM

---

acproj.getpred *Extract the projection results by 5-year period*

Description
Uses an acproj.object to extract the observed and projected annual age-specific rates, or age-standardized rates and annual numbers in 5-year period

Usage
```
acproj.getpred(acproj.object,
               incidence=T, standpop=NULL, excludeobs=F,
               byage, agegroups="all")
```

Arguments

acproj.object An object based on the 'acproj()' or 'acproj.prediction()' function

incidence Indicating whether to give rates or numbers of cases

standpop A vector with the weights (proportions) of 18 age groups in a standard population, default (NULL) gives the age-specific rates. Age-standardized rates and total numbers are given by specifying standpop, e.g. standpop=ca91.

excludeobs Indicates whether to output observed values, default (F) does not output observed values.

byage Report numbers by age groups. If byage=F, crude or age-standardized rates are given.

agegroups Specify age groups to include. E.g. agegroups=7:18 include age groups seven to eighteen. The default ('All') includes all 18 age groups.

Value

A data frame containing observed and / or projected rates and / or numbers in 5-year period

---

acproj.getproj *Extract the annual projection results*

Description

Apply the segmental linear interpolation function "asrpy" on an acproj.object to extract the observed and projected annual age-specific rates, or annual age-standardized rates and annual numbers

Usage

acproj.getproj(cdat, pdat, startp, acproj.object, standpop=NULL)

Arguments

cdat A data frame for annual numbers of cancer cases/deaths with the rows representing the 18 5-year age groups in ascending order and the N columns representing the calendar years in ascending order where $15 \le N \le 125$.

pdat A data frame for annual observed and projected population with the rows representing the 18 5-year age groups in ascending ordered and the $N+M$ columns representing the calendar years in ascending order where M is the number of projection years, $5 \le M \le 25$.

startp Specify the starting calendar year of projection to format appropriate outputs, e.g. startp=2009.

acproj.object	An object based on the 'acproj()' or 'acproj.prediction()' function
standpop	A vector with the weights (proportions) of 18 age groups in a standard population, default (NULL) gives the age-specific rates. Age-standardized rates and total numbers are given by specifying standpop, e.g. standpop=ca91.

Value
A data frame containing annual observed and projected age-specific rates, or annual age-standardized rates and numbers

---

glm.acproj *Summarize estimations from the age-cohort model*

Description
Using an acproj.object to summarize the estimations from the age-cohort model

Usage
```
glm.acproj(acproj.object)
```

Arguments
acproj.object An object based on the 'acproj()' or 'acproj.prediction()' function

Value
A summary table from the glm.object

---

plot.acproj *Produce the graph of the observed and projected age-standardized rates*

Description
Using an acproj.object to plot the observed and projected age-standardized rates against the calendar year

Usage
```
plot.acproj(cdat, pdat, startp,
            acproj.object,
            standpop, startplot=1, new=T...)
```

Arguments

cdat	A data frame for annual numbers of cancer cases/deaths with the rows representing the 18 5-year age groups in ascending order and the N columns representing the calendar years in ascending order where $15 \le N \le 125$.

pdat	A data frame for annual observed and projected population with the rows representing the 18 5-year age groups in ascending ordered and the $N+M$ columns representing the calendar years in ascending order where M is the number of projection years, $5 \leq M \leq 25$.
startp	Specify the starting calendar year of projection to format appropriate outputs, e.g. startp=2009.
acproj.object	An object based on the 'acproj()' or 'acproj.prediction()' function
standpop	A vector with the weights (proportions) of 18 age groups in a standard population
startplot	Indicating the starting year for the plot, default (1) gives the plot starting from the first observed year to the end of projection year.
new	Produce graph in another graphic window by default, or in the same window by setting new=F.
…	Other "par" parameters such as ylab, ylim, type, col etc. can be applied in this function.

---

summary.acproj *Summarize the information on projection method used*

Description
Uses an acproj.object to summarize the information on projection method used

Usage
summary.acproj(acproj.object, printpred=F, printcall=F, digits=0)

Arguments

acproj.object	An object based on the 'acproj()' or 'acproj.prediction()' function
printpred	Indicating whether to print the observed and predicted number of cases, default (F) does not print them out.
printcall	Indicating whether to print the function `Call' for the acproj.object, default (F) does not print it out.
digits	Specify the number of digits in the tabulation, default (0) keeps integer only.

Value
An information table describing the projection method used

hybdproj *Projection by fitting the final generalized linear model in age-period stream*

Description

Produce projections by the selected generalized linear model from age-only, common trend, age-specific trend and negative-binomial distribution based age-specific trend models in age-period stream. Attenuation parameters for the period trend are adopted and `power5', `log', `sqrt' and `identity' link function can be selected.

Usage

```
hybdproj(cdat, pdat,
          projfor="incidence", nagg=NULL, ncase=NULL,
          cuttrd=0.05, shortp=0, linkfunc="power5",
          pD=0.05, pGOF=0.05)
```

Arguments

cdat A data frame for annual numbers of cancer cases/deaths with the rows representing the 18 5-year age groups in ascending order and the N columns representing the calendar years in ascending order where $15 \le N \le 125$.

pdat A data frame for annual observed and projected population with the rows representing the 18 5-year age groups in ascending ordered and the $N+M$ columns representing the calendar years in ascending order where M is the number of projection years, $5 \le M \le 25$.

projfor Specify "incidence" or "mortality" projection to determine the default settings of "nagg" and "ncase" (see below). Default is projfor="incidence".

nagg Number of years for data aggregation to stabilize rates, if nagg is not specified (NULL), when projfor="incidence", nagg=1 if the averaged asir > 15, nagg=2 if 10 < asir ≤ 15, nagg=3 if 5 < asir ≤ 10 and nagg=4 if asir ≤ 5; when projfor="mortality", nagg=1 if asmr > 9, nagg=2 if 6 < asmr ≤ 9, nagg=3 if 3 < asmr ≤ 6 and nagg=4 if asmr ≤ 3.

ncase Minimum number of cancer cases/deaths per year for splitting data by age group into two datasets: one for average method, the other for regression modeling. When ncase is not specified (NULL), 1 is used for projfor="incidence", and 3/5 is used for projfor = "mortality".

cuttrd Attenuation percent for common trend or age-specific trend per year, default is 0.05 per year after the first 5 projection years.

shortp Attenuation percent for common trend or age-specific trend in the first 5 projection year, default is zero.

linkfunc	Specify the link function to use in GLMs. Default is "power5", can be specified as "log", "sqrt" and "identity".
pD	Significance level for the *p*-value of estimated slope parameter in the common trend model, default is 0.05.
pGOF	Significance level for the *p*-value of model comparison based on their goodness-of-fits, default is 0.05.

Value

An object of class "hybdproj" — a list with components:

predictions	A data frame with projected number of cases in `nagg'-year period
pyr	A data frame with observed and projected population in `nagg'-year period
cuttrend	A vector to attenuate the common trend or age-specific trend in projection `nagg'-year period
cuttrd	Used attenuation percent for common trend or age-specific trend per year
shortp	Used attenuation percent for common trend or age-specific trend in the first 4, 5 or 6-year
nopred	Number of `nagg'-year periods in projection years
noperiod	Number of `nagg'-year periods used in historical data
lastperiod	Number of `nagg'-year periods in the determined projection base
noobsper	Number of `nagg'-year periods in whole observed data
nototper	Number of total observation and projection periods in `nagg'-year
noyearagg	Number of years in a period, i.e. the parameter `nagg'
nocaseagp	Minimum number of cancer cases/deaths per year in an age group for GLMs, i.e. the parameter `ncase'
agrpave	The vector of age groups used for average method
agrpmod	The vector of age groups used for GLMs
linkunc	Link function used in GLMs, one of the 4 outputs: power5, log, sqrt and identity
projbase	Number of years in the projection base

finalmod	The final determined model used for projection, one of the 4 outputs: average, common-trend, age-specific and nba-specific (negative-binomial distribution based age-specific trend GLM)
gofpvalue	The *p*-value of goodness of fit for the final determined model
glm	The glm.object of the final determined model

See Also

hybdproj.estimate, hybdproj.prediction, hybdproj.getpred, hybdproj.getproj, glm.hybdproj, plot.hybdproj and summary.hybdproj

---

hybdproj.estimate *Define projection base, model selection and fitting in age-period stream*

Description

Determine the final model and projection base, and estimate parameters for the final determined model using data in the determined projection base, the final model is selected from age-only, common trend, age-specific trend or negative binomial distribution based age-specific trend models.

Usage

```
hybdproj.estimate(cases, pyr, nagg, ncase, linkfunc="power5", pD=0.05, pGOF=0.05)
```

Arguments

cases	A data frame with number of cases in `nagg'-year period by ascending ordered 18 age groups in row
pyr	A data frame with observed and projection population size in `nagg'-year period by ascending ordered 18 age groups in row
nagg	Number of years for data aggregation
ncase	Minimum number of cancer cases/deaths per year in an age group for GLMs
linkfunc	The link function to be used in GLMs, default is "power5", can be specified as "log", "sqrt" and "identity".
pD	Significance level for the *p*-value of estimated slope parameter in the common trend model, default is 0.05.
pGOF	Significance level for the *p*-value of model comparison based on their goodness-of-fits, default is 0.05.

Value

An object of class "hybdproj.estimate" — a list with components:

glm — The glm.object of the final determined model

cases — A data frame with observed number of cases in `nagg'-year period

pyr — A data frame with observed and projected population in `nagg'-year period

agrpave — The vector of age groups used for average method

lastper — Number of `nagg'-year periods in the determined projection base

cuty — The starting period in `nagg'-year for the determined projection base

noperiod — Number of `nagg'-year periods used in historical data

noyearagg — Number of years in a period, i.e. the parameter `nagg'

nocaseagp — Minimum number of cancer cases/deaths per year in an age group for GLMs, i.e. the parameter `ncase'

linkunc — Link function used in GLMs, one of the 4 outputs: power5, log, sqrt and identity

agrpmod — The vector of age groups used for GLMs

projbase — Number of years in the projection base

finalmod — The final determined model used for projection, one of the 4 outputs: average, common-trend, age-specific and nba-specific (negative-binomial distribution based age-specific trend GLM)

gofpvalue — The *p*-value of goodness of fit for the final determined model

---

hybdproj.prediction *Extrapolate estimated trend from the final model in age-period stream*

Description

Calculating age-specific projections using an `hybdproj.estimate' object which is based on one of age-period stream models: age-only, common trend, age-specific trend or negative binomial distribution based age-specific trend models

Usage

```
hybdproj.prediction(hybdproj.estimate.object, cuttrd=0.05, shortp=0)
```

Arguments

hybdproj.estimate.object — A glm.object based on 'hybdproj.estimate()'

cuttrd	Attenuation percent for common trend or age-specific trend per year, default is 0.05 per year after the first 5 projection years.
shortp	Attenuation percent for common trend or age-specific trend in the first 5 projection year, default is zero.

Value

An object of class "hybdproj" — a list with components:

predictions	A data frame with projected number of cases in `nagg'-year period
pyr	A data frame with observed and projected population in `nagg'-year period
cuttrend	A vector to attenuate the common trend or age-specific trend in projection `nagg'-year period
cuttrd	Used attenuation percent for common trend or age-specific trend per year
shortp	Used attenuation percent for common trend or age-specific trend in the first 4, 5 or 6-year
nopred	Number of `nagg'-year periods in projection years
noperiod	Number of `nagg'-year periods used in historical data
lastperiod	Number of `nagg'-year periods in the determined projection base
noobsper	Number of `nagg'-year periods in whole observed data
nototper	Number of total observation and projection periods in `nagg'-year
noyearagg	Number of years in a period, i.e. the parameter `nagg'
nocaseagp	Minimum number of cancer cases/deaths per year in an age group for GLMs, i.e. the parameter `ncase'
agrpave	The vector of age groups used for average method
agrpmod	The vector of age groups used for GLMs
linkunc	Link function used in GLMs, one of the 4 outputs: power5, log, sqrt and identity
projbase	Number of years in the projection base
finalmod	The final determined model used for projection, one of the 4 outputs: average, common-trend, age-specific and nba-specific (negative-binomial distribution based age-specific trend GLM)
gofpvalue	The *p*-value of goodness of fit for the final determined model

glm	The glm.object of the final determined model

---

hybdproj.getpred *Extract the projection results by `nagg'-year period*

Description

Uses a hybdproj.object to extract the observed and projected annual age-specific rates, or age-standardized rates and annual numbers in `nagg'-year period

Usage

```
hybdproj.getpred(hybdproj.object,
                 incidence=T, standpop=NULL, excludeobs=F,
                 byage, agegroups="all")
```

Arguments

hybdproj.object	An object based on the 'hybdproj()' or 'hybdproj.prediction()' function
incidence	Indicating whether to give rates or numbers of cases
standpop	A vector with the weights (proportions) of 18 age groups in a standard population, default (NULL) gives the age-specific rates. Age-standardized rates and total numbers are given by specifying standpop, e.g. standpop=ca91.
excludeobs	Indicates whether to output observed values, default (F) does not output observed values.
byage	Report numbers by age groups. If byage=F, crude or age-standardized rates are given.
agegroups	Specify age groups to include. E.g. agegroups=7:18 include age groups seven to eighteen. The default ('All') includes all 18 age groups.

Value

A data frame containing observed and / or projected rates and / or numbers in `nagg'-year period

---

hybdproj.getproj *Extract the annual projection results*

Description

Apply the segmental linear interpolation function "asrpy" on a hybdproj.object to extract the observed and projected annual age-specific rates, or annual age-standardized rates and annual numbers

Usage

hybdproj.getproj(cdat, pdat, startp, hybdproj.object, standpop=NULL, Ave5=NULL, sum5=NULL)

Arguments

cdat	A data frame for annual numbers of cancer cases/deaths with the rows representing the 18 5-year age groups in ascending order and the N columns representing the calendar years in ascending order where $15 \le N \le 125$.
pdat	A data frame for annual observed and projected population with the rows representing the 18 5-year age groups in ascending ordered and the $N+M$ columns representing the calendar years in ascending order where M is the number of projection years, $5 \le M \le 25$.
startp	Specify the starting calendar year of projection to format appropriate outputs, e.g. startp=2009.
hybdproj.object	An object based on the 'hybdproj()' or 'hybdproj.prediction()' function
standpop	A vector with the weights (proportions) of 18 age groups in a standard population, default (NULL) gives the age-specific rates. Age-standardized rates and total numbers are given by specifying standpop, e.g. standpop=ca91.
Ave5	Specify if the 5-year average method (Ave5=T) is used when the age-only model is selected; the default (NULL) take the results from the age-only model.
sum5	Determine how to average the 5-year rates when Ave5=T. The default (NULL) calculates the age-specific rates using the aggregated numbers and population sizes, respectively, in the recent 5 years; otherwise (sum5=T), the yearly age-specific rates will be averaged over the 5 years.

Value

A data frame containing annual observed and projected age-specific rates, or annual age-standardized rates and numbers

---

glm.hybdproj *Summarize estimations from the final model*

Description

Using a hybdproj.object to summarize the estimations from the final model in age-period stream used for projection

Usage

glm.hybdproj(hybdproj.object)

Arguments

hybdproj.object An object based on the 'hybdproj()' or 'hybdproj.prediction()' function

Value

A summary table from the glm.object

---

plot.hybdproj *Produce the graph of the observed and projected age-standardized rates*

Description

Using a hybdproj.object to plot the observed and projected age-standardized rates against the calendar year

Usage

plot.hybdproj(cdat, pdat, startp, hybdproj.object, standpop, startplot=1, new=T...)

Arguments

cdat	A data frame for annual numbers of cancer cases/deaths with the rows representing the 18 5-year age groups in ascending order and the N columns representing the calendar years in ascending order where $15 \le N \le 125$.
pdat	A data frame for annual observed and projected population with the rows representing the 18 5-year age groups in ascending ordered and the $N+M$ columns representing the calendar years in ascending order where M is the number of projection years, $5 \le M \le 25$.
startp	Specify the starting calendar year of projection to format appropriate outputs, e.g. startp=2009.
hybdproj.object	An object based on the 'hybdproj()' or 'hybdproj.prediction()' function
standpop	A vector with the weights (proportions) of 18 age groups in a standard population
startplot	Indicating the starting year for the plot, default (1) gives the plot starting from the first observed year to the end of projection year.
new	Produce graph in another graphic window by default, or in the same window by setting new=F.
…	Other "par" parameters such as ylab, ylim, col etc. can be applied in this function

---

summary.hybdproj *Summarize the information on projection method used*

Description
Uses a hybdproj.object to summarize the information on projection method used

Usage
summary.hybdproj(hybdproj.object, printpred=F, printcall=F, digits=0)

Arguments

hybdproj.object	An object based on the 'hybdproj()' or 'hybdproj.prediction()' function
printpred	Indicating whether to print the observed and predicted number of cases, default (F) does not print them out.
printcall	Indicating whether to print the function `Call' for the hybdproj.object, default (F) does not print it out.
digits	Specify the number of digits in the tabulation, default (0) keeps integer only.

Value
An information table describing the projection method used

3.5. Ave5proj Functions

ave5proj *Projection by averaged age-specific rates from recent 5 years*

Description
Calculate the averaged age-specific rate in the recent five years and assume these constant rates in the future for projections

Usage
ave5proj(cdat, pdat, startp, sum5=NULL)

Arguments

cdat	A data frame for annual numbers of cancer cases/deaths with the rows representing the 18 5-year age groups in ascending order and the N columns representing the calendar years in ascending order where $15 \le N \le 125$.

pdat — A data frame for annual observed and projected population with the rows representing the 18 5-year age groups in ascending ordered and the $N + M$ columns representing the calendar years in ascending order where M is the number of projection years, $5 \leq M \leq 25$.

startp — Specify the starting calendar year of projection to format appropriate outputs, e.g. startp = 2009.

sum5 — Determine how to average the 5-year rates. The default (NULL) calculates the age-specific rates using the aggregated numbers and population sizes, respectively, in the recent 5 years; otherwise (sum5=T), the yearly age-specific rates will be averaged over the 5 years.

Value

A list of class "ave5proj" with components:

annproj — A data frame with annual observed and projected age-standardized rates and numbers of cancer incidence or mortality, ordered by calendar year in row.

agsproj — A data frame with the averaged age-specific rates in recent 5 years, ordered by age group.

See Also

ave5.getproj

---

ave5proj.getproj *Extract the projection results*

Description

Uses an ave5proj.object to extract the observed and projected annual age-standardized rates and numbers, or the averaged age-specific rates in recent 5 years

Usage

```
ave5proj.getproj(pdat, ave5proj.object, standpop=NULL)
```

Arguments

pdat — A data frame for annual observed and projected population with the rows representing the 18 5-year age groups in ascending ordered and the $N + M$ columns representing the calendar years in ascending order where M is the number of projection years, $5 \leq M \leq 25$.

ave5proj.object — An object based on ave5proj() function

standpop	Default (NULL) gives the age-specific rates. Otherwise, annual age-standardized rates and total numbers will be produced.

Value
A data frame containing the averaged age-specific rates over recent 5 years, or observed and projected annual age-standardized rates and numbers

---

summary.ave5proj *Summarize the information on projection method used*

Description
Uses an ave5proj.object to summarize the information on projection method used

Usage
summary.ave5proj(ave5proj.object, printcall=F, digits=0)

Arguments

ave5proj.object	An object based on the 'ave5proj()' function
printcall	Indicating whether to print the function `Call' for the ave5proj.object, default (F) does not print it out.
digits	Specify the number of digits in the tabulation, default (0) keeps integer only.

Value
An information table describing the projection method used

3.6. Miscellaneous Functions

---

datagg *Aggregation of data by years*

Description
Aggregate annual observed cancer data and observed and projected population data in 18 age groups by years

Usage

datagg(cdat, pdat, nagg)

Arguments

cdat	A data frame for annual numbers of cancer cases/deaths with the rows representing the 18 5-year age groups in ascending order and the N columns representing the calendar years in ascending order where $15 \leq N \leq 125$.
pdat	A data frame for annual observed and projected population with the rows representing the 18 5-year age groups in ascending ordered and the $N + M$ columns representing the calendar years in ascending order where M is the number of projection years, $5 \leq M \leq 25$.
nagg	Number of years in a period for data aggregation

Value

A list with components:

cases	A data frame with number of cases in \`nagg'-year period by ascending ordered 18 age groups in row
pyr	A data frame with observed and projection population size in \`nagg'-year period by ascending ordered 18 age groups in row

See Also

asrpy

---

asrpy *Convert to annual age-specific rate*

Description

Convert age-specific rates in \`nagg'-year period to annual age-specific rates by segmental linear interpolation

Usage

asrpy(rate, cdat, pdat, startp, nagg)

Arguments

rate	Observed and projected age-specific rates in \`nagg'-year period
cdat	A data frame for annual numbers of cancer cases/deaths with the rows representing the 18 5-year age groups in ascending order and the N columns representing the calendar years in ascending order where $15 \leq N \leq 125$.

pdat	A data frame for annual observed and projected population with the rows representing the 18 5-year age groups in ascending ordered and the $N+M$ columns representing the calendar years in ascending order where M is the number of projection years, $5 \le M \le 25$.
startp	Specify the starting calendar year of projection to format appropriate outputs, e.g. startp = 2009.
nagg	Number of years in a period for data aggregation

Value
A data frame with annual observed and projected age-specific rates

See Also
asry

---

asry *Calculate annual age-standardized rates and total numbers*

Description
Calculate the annual age-standardized rates and total numbers using the annual age-specific rates and population sizes

Usage
asry(rr, pdat, standpop=ca91)

Arguments

rr	Observed and projected annual age-specific rates
pdat	A data frame for annual observed and projected population with the rows representing the 18 5-year age groups in ascending ordered and the $N+M$ columns representing the calendar years in ascending order where M is the number of projection years, $5 \le M \le 25$.
standpop	A vector with the weights (proportions) of 18 age groups in a standard population. Default is based on the 1991 Canadian standard population (ca91). Alternatively, user can choose `ca96' (the 1996 Canadian standard population), `wdsd' (World standard population), or define a vector for other standard population.

Value
A data frame with annual observed and projected age-standardized rates and total numbers

See Also
asrpy

obasr *Calculate observed age-standardized rates and total numbers*

Description
Calculate the annual age-standardized rates and total numbers for observed data

Usage
obasr(cdat, pdat, standpop=ca91)

Arguments

cdat — A data frame for annual numbers of cancer cases/deaths with the rows representing the 18 5-year age groups in ascending order and the N columns representing the calendar years in ascending order where $15 \leq N \leq 125$.

pdat — A data frame for annual observed population with the rows representing the 18 5-year age groups in ascending ordered and the N columns representing the calendar years in ascending order where $15 \leq N \leq 125$.

standpop — A vector with the weights (proportions) of 18 age groups in a standard population. Default is based on the 1991 Canadian standard population (ca91). Alternatively, user can choose `ca96' (the 1996 Canadian standard population), `wdsd' (World standard population), or define a vector for other standard population

Value
A data frame with annual observed age-standardized rates and total numbers

See Also
asrsd

asrsd *Calculate age-standardized rates and corresponding standard error*

Description
Calculate the annual observed and/or projected age-standardized rates with corresponding standard error

Usage
asrsd(cdat, pdat, standpop=ca91)

Arguments

cdat	A data frame for annual numbers of cancer cases/deaths with the rows representing the 18 5-year age groups in ascending order and the N columns representing the calendar years in ascending order where $15 \leq N \leq 125$.
pdat	A data frame for annual observed population with the rows representing the 18 5-year age groups in ascending ordered and the N columns representing the calendar years in ascending order where $15 \leq N \leq 125$.
standpop	A vector with the weights (proportions) of 18 age groups in a standard population. Default is based on the 1991 Canadian standard population (ca91). Alternatively, user can choose \`ca96' (the 1996 Canadian standard population), \`wdsd' (World standard population), or define a vector for other standard population

Value
A data frame with annual observed and/or projected age-standardized rates and corresponding standard error

See Also
obasr

---

chper *Calculate the percentage changes due to risk, population growth and aging*

Description
Given reference year and comparison year, split the overall change into the percentage changes due to risk, population growth and aging

Usage
chper(aspr, pdat, byear, cyear, starty=NULL)

Arguments

aspr	A data frame for annual age-specific rates in ascending ordered 18 age groups (by row)
pdat	A data frame for annual age-specific population size with the same starting year and end year as those in the \`aspr' in ascending ordered 18 age groups (by row)
byear	The calendar year of the reference year
cyear	The calendar year of the comparison year
starty	The first calendar year in historical data, if NULL (default) and the colnames of aspr is expressed by numeric calendar year, the calendar year in the first column in aspr will be used.

Value

A data frame with components:

ref.case	The total number in the reference year
comp.case	The total number in the reference year
overall	The overall percentage of change in total number
risk	The percentage change due to risk
p.growth	The percentage change due to population growth
p.aging	The percentage change due to population aging

---

projplot *Produce the graph of the observed and projected age-standardized rates and numbers*

Description

Plot the observed and projected total numbers by barplot and the age-standardized rates by lines against the calendar year

Usage

projplot(site.asr, sex=NULL, ma=2, mr=1.02, starty=1986, startp=2011)

Arguments

site.asr	A data frame with annual observed and projected age-standardized rates and total numbers
sex	Specify sex for defining color used in graph, default (NULL) gives gray colors for both sexes, pink color (sex="F") for female; blue color for male (sex="M").
ma	A coefficient to adjust the scale of the left y-axis. Default (2) takes the 2 times of the maximum of the number of annual cases/deaths.
mr	A coefficient to adjust the scale of the right y-axis. Default (1.02) takes the 1.02 times of the maximum of the age-standardized rates .
starty	Specify the starting calendar year of observation, default is starty=1986.
startp	Specify the starting calendar year of projection, default is startp=2011.

Datasets:

ca91	A vector with the proportions of 18 age groups in 1991 Canadian standard population
ca96	A vector with the proportions of 18 age groups in 1996 Canadian standard population
wdsd	A vector with the proportions of 18 age groups in the World standard population