Python Module to calculate age-adjusted prevalence using the SPSS Complex Sample Survey (15.0, 16.0, and 17)

Raynald Levesque[1]
Juan Albertorio[2]
Art Kendall[3]

1. Aon Consulting, Montreal, Canada
2. International Statistics Program, National Center for Health Statistics, Centers for Disease Control and Prevention.
3 Social Research Consultants

INTRODUCTION

The following document contains a brief explanation of the tailored python module created by Raynald Levesque and Juan Albertorio to acquire age-adjusted prevalence using the SPSS Complex Sample General Lineal Model (CSGLM). The module has been tested on SPSS versions 15, 16 and 17 (beta). The aim of this documentation is to make available for the SPSS user a piece of program that allows the age-adjusted[1] (or standardized) prevalence[2] calculation using the new complex sample survey module in SPSS instead of SAS or SUDAAN[3].

The case presented in this module directly addresses the utilization of age-adjusted (direct method) prevalence technique that is employ in the discipline of Public Health for calculating age-adjusted high blood pressure using data from the National Health and Nutrition Examination Survey (NHANES) produced by the National Center for Health Statistics, Center for Disease Control and Prevention (CDC). However, the same methodology can be applied to any categorical dichotomous variable that needs to be standardized or age-adjusted.

We would like to kindly thanks the enormous help of Damir Spisic, David Nichols and John Peck for assisting us in this endeavor. Particular thanks to Jocelyn Kennedy-Stephenson, and Margaret Carroll from NCHS in helping us with the understanding of the SAS and SUDAAN programming. Thanks also to Frances Silva (2008 HSHPS summer intern from Mailman School of Public Health at Columbia University) for providing valuable editorial comments to the draft version of this documentation.

Any recommendation or suggestion on how to improve the clarity or usefulness of this module can be directed to Raynald Levesque's email at: raynald@spsstools.net, or Juan Albertorio at jna8@cdc.gov.

There is no GUI established routine to run age-adjusted prevalence in the SPSS CSGLM. User has to open a syntax window and type (or copy/paste) the command utilized in this module and tailor to their particular analysis using the information provided here as a guideline. Nonetheless, we think that the created python module to calculate age-adjusted prevalence using the CSGLM is a worth contribution to the SPSS community for the continue exploration of using SPSS in the context of complex samples surveys.

[1] Age adjustment is an ARTIFICIAL estimate that minimizes the effects of different age distribution and allows comparisons between different populations. Its represents what the crude percentage would have been in the study population if that population have the same age distribution as a standard population.
[2] Prevalence- The percentage of individuals in a population having a disease or a condition. Prevalence is a statistical concept referring to the number of cases of a disease or a condition that are present in a particular population at a given time.
[3] SUDAAN provides the standardized estimates directly within a descriptive procedure, but both SPSS and SAS require use of the general linear model to obtain the same.

Instructions to obtain age adjusted prevalence rates using ageAdjusted.py module.

Sections:

1. **BEFORE YOU START**

Dataset

- Please download the public-use dataset to calculate age-adjusted prevalence provided by the National Health and Nutrition Examination Survey (NHANES) at http://www.cdc.gov/nchs/tutorials/Nhanes/Downloads/intro.htm#15
- We highly recommend reading and reviewing the NHANES tutorial for age standardization and population estimates available at: http://www.cdc.gov/nchs/tutorials/Nhanes/NHANESAnalyses/AgeStandardization/age_standardization_intro.htm .
    i. This information will give you a brief, but complete explanation of the GLM age-adjustment procedure utilized in SAS, which is the same that SPSS employ.

Python module

- Go to the SPSS website (SPSS Developer Central website) to download the following python modules:
    - spssaux.py
    - namedtuple.py,
    - and spssdata.py and install these in "Python\Lib\site-packages[4].
- Also download the ageAdjusted.py module from either the SPSS website or www.spsstools.net/python.htm and save it in the same python library.

2. DATA PREPARATION

Age[5] variable: from continuous to categorical.
The data file must contain a numeric variable with values from 1 to the number of distinct age's categories you wish to use in the analysis. The actual name of the age variable may be any legal SPSS name. Say you have three age ranges then values one to three (and ONLY values 1 to 3) must have value labels such as:

VALUE LABELS **Age**
1 "20-39"
2 "40-59"
3 "60+"

For this end, we highly recommend to leave the original variable untouched and use the "recode into a different variables" command or use SPSS syntax to create a new categorical variable.

Special cases of the **age** variable:

- The maximum age may also be specified as 999, in other words, the "60-999" and "60+" are equivalent.
- If an **age** category includes only a single age (for instance age 20), you may specify either "20" or "20-20".
- Age category "0-n" is equivalent to "-n", in other words, "0-20" is the same as "20".

Additional categorical values used in the analysis
Every variable that will be used in the CSGLM procedure needs to be recoded to contain categorical numerical values. This begins from one to the number of categories that will be used in the analysis. For example, a gender variable has to be recode to 1 and 2 with the corresponding value labels- i.e. 1=Male, 2= Female.

[4] The folder "site-packages" is a default space created by Python for storage of third party files like the AgeAdjusted.py program.
[5] For illustrative purpose, SPSS variable's names are bold.

Suppose a categorical variable contains values from one to four but you only need to include values one to three in the analysis, then you need to make a copy of the data file, delete the value labels for the values that are not used (value 4 in this case) and either

- delete these cases, or
- define these values as missing (using the MISSING VALUES command).

Dependent variable under analysis

The dependent variable has to be recoded as a categorical variable. In this example, **hbpx** is derived from two continuous variables into a unique categorical variable (0= no high blood pressure, 100 = high blood pressure)[6][7]. We use the value of 100 instead of 1 to derive the prevalence estimate.

3. REFERENCE POPULATION PREPARATION

The reference population (which is located in a separate "SPSS .sav" file) has to be located in the same folder where the dataset is located. For this exercise, the 2000 US standard population is utilized. However, this can be modified to use other standard population, such as the 2000 World Health Organization (WHO) population as well as other standard population of your choice.

The reference population is utilized to calculate age-adjusted prevalence based on the direct method of age standardization. The age adjustment direct method applies observed age-specific rates to a standard age distribution to eliminate (or control) differences in crude rates in populations of interest that result from differences in the population's age distribution[8] (Klein & Schoenborn, 2001). For more information about the pro and cons of the use and implication of the age standardization direct method in the public health field, please see Klein & Schoenborn (2001), Curtin & Klein (1995), and Krieger &Williams (2001).

Here is a sample.sav (2000 US standard reference population) file:

AgeRange	Population[9]
0	3,795
1	3,759
2-4	11,433
5	3,896
6-8	11,800
9	4,224
10-11	8,258
12-14	11,799
15-17	11,819
18-19	8,001
20-24	18,257
25-29	17,722
30-34	19,511
35-39	22,180
40-44	22,479
45-49	19,806
50-54	17,224

[6] HBPX is a categorical variable with value of 100 for event and 0 for non-event. The values are chosen so that the event results are displayed as percentages rather than proportions.

[7] For more information, please visit NHANES website tutorial at http://www.cdc.gov/nchs/tutorials/Nhanes/index.htm

[8] This adjustment is done when comparing two or more populations at one point in time or one population at two or more points in time. This technique is particularly relevant when populations being compared have different age structures.

[9] Table 1. Master list: 2000 U.S. projected population in thousands Klein & Shoenborn, 2001 page 2.

55-59	13,307
60-64	10,654
65-69	9,410
70-74	8,726
75-79	7,415
80-84	4,900
85+	4,259

The file must contain two columns (variable names do not matter):
- First column (**AgeRange**) must be a string variable describing the age range (or single age) applicable to the case. The same formats as given in first column above need to be used.
- The second column (**Population**) must contain the corresponding population. In the example above, the population is in thousands. In such case, it is **not** necessary to multiply the population numbers by 1000.

4. SAMPLE USE OF AGEADJUSTED.PY MODULE

The file *ageAdjusted.py* (which contains the original SPSS syntax written by Damir Spics, and later automated by Raynald Levesque) has to be saved in a folder searched by python[10]. This file as a rule has to be saved in the Python folder \Lib\site-packages.

In addition to having the ageAdjusted.py file, you must also have: 1) The *SPSS Integration-plug* installed, and the *spssaux.py*, *namedtuple.py*, and *spssdata s*aved in the Python Lib\site-packages\ folder. These files can be found at SPSS Developer Central website. Missing any of the aforementioned files will not allow the proper function of the python program.

In the following example, the SPSS syntax file tests the four specified models:
1st model: hbpx= age (for the total population age > 18 years and older)
2nd model: hbpx =riagendr + age + riagendr*age (by gender)
3rd model: hbpx=race + age + race*age (by ethnicity)
4th model: hbpx =riagendr + race + age + riagendr*race*age (by ethnicity and sex)

Important Notes:
1) Models are defined using the "factors" list parameter
2) You need to list only the relevant factors (with the age factor being the LAST ONE)
3) You do not list the interaction term. The python module generates the interaction term using the variables listed.

Warning: You must adapt file paths and file names to correspond to the specific location of the files in your computer. When executed, the python script presented here runs 4 specified models and creates the file "ageAdjusted.SPS" (Listed in 7. below).

[10] For SPSS 15 the python folder is Python24. However, for SPSS 16 and 17 the Python folder is 25.

5. RUNNING THE PYTHON PROGRAM FOR CALCULATE AGE ADJUSTED PREVALENCE

After you have successfully installed python and saved the aforementioned python modules in the python site-packages folder, copy and paste the following syntax in a SPSS syntax window.

```
BEGIN PROGRAM.
#!/usr/bin/env python
import ageAdjusted as aa
aa.master (dataFile =r'c:\Mes Documents\CDC\Age_Adj Exercise NHANES_labeled.sav',
        planFile='mec4yr_plan.csaplan',
        referencePop =r'c:\Mes Documents\CDC\US Census pop 2000.sav',
        depVar='hbpx',
        domainVar='sel(1)',
        factors=['age',
                'riagendr age',
                'race age',
                'riagendr race age'])
END PROGRAM.
```

6. SPSS SYNTAX EXPLANATION

Statement	Explanation
BEGIN PROGRAM. #!/usr/bin/env python	BEGIN PROGRAM opens a python session from SPSS syntax
import ageAdjusted as aa	Import the ageAdjusted module and name it aa.
aa.master(dataFile=r'c:\Mes Documents\CDC\Age_Adj Exercise NHANES_labeled.sav',	Call the master function of the aa module
planFile='mec4yr_plan.csaplan',	PLAN FILE subcommand allows SPSS to identify the Complex Survey plan to apply to the analysis. (More information of how to create a complex survey plan can be found at the companion Complex Sample documentation that comes with the SPSS software. This file is assumed to be located in the same folder as the data file.
referencePop=r'c:\Mes Documents\CDC\US Census pop 2000.sav',	REFERENCEPOP tells SPSS where the reference population file is located. We strongly suggest saving this file *in the same folder as the data that is being analyzed*.
depVar='**hbpx**',	Use the DEPVAR parameter to identify the dependent variable under analysis.
domainVar='**sel**(1)',	Use DOMAIN as a subpopulation subcommand to select those 20 years and older.
factors=['age', [1] 'riagendr age', [2] 'race age', [3] 'riagendr race age']) [4]	The FACTORS statement starts with a bracket ([) and finishes with a bracket (]). • Each set of factors are enclosed within quotes and separated by commas. • The last variable of each set **has to be** the **age** variable (the actual name of this variable can be any legal SPSS name)**.**

	• We recommend separating each set of factors (model) by line. This simplifies any future reading of the model. Use FACTORS subcommand to produce prevalence of **hbpx** for 4 models under analysis: 1- '**age'** (One combination- total). 2- '**riagendr' age** (gender has two groups (total of 2 groups). 3- '**race age'** (race has four groups and age has three groups, these together equal a total of 12 combinations. 4- '**riagendr race age'** (gender has two groups and race has 4, these total 8 possible combinations).
END PROGRAM.	This command ends the SPSS program.

Soon, an explanation of the SPSS output would be available at Raynald SPSStools website.

7. APPENDIX

The original SPSS syntax was created by Damir Spics, it replicates the SAS code utilized in the NHANES tutorial example.

The code below is generated by the ageAdjusted module (syntax is listed in section 5 above).
Note that highlighted comments have been added afterwards, they are not part of the file created by the python module.

```
SET MPRINT=YES /PRINTBACK=YES.
CD "c:\Mes Documents\CDC".
DATASET CLOSE ALL.
GET FILE='c:\Mes Documents\CDC\Age_Adj Exercise NHANES_labeled.sav'.
OMS SELECT TABLES
  /IF SUBTYPES=['IndividualTestResults' 'TestsofModelEffects']
  /DESTINATION FORMAT=SAV OUTFILE="c:\Mes Documents\CDC/_indivTest.sav".

*Age adjusted prevalence for total population 18+ years
CSGLM  hbpx BY age
  /PLAN FILE='mec4yr_plan.csaplan'
  /MODEL age
  /INTERCEPT INCLUDE=NO SHOW=YES
  /CUSTOM LABEL="Total"
   LMATRIX=age 77670/195850  72816/195850  45364/195850
  /PRINT SUMMARY VARIABLEINFO SAMPLEINFO
  /STATISTICS PARAMETER SE TTEST
  /DOMAIN VARIABLE= sel(1)
*Age adjusted prevalence by gender 18+ years
CSGLM  hbpx BY riagendr AGE
  /PLAN FILE='mec4yr_plan.csaplan'
  /MODEL riagendr AGE RIAGENDR*AGE
  /INTERCEPT INCLUDE=NO SHOW=YES
  /CUSTOM LABEL="Male"
   LMATRIX=riagendr 1 0 AGE 77670/195850  72816/195850  45364/195850  RIAGENDR*AGE
77670/195850  72816/195850  45364/195850  0 0 0
  /CUSTOM LABEL="Female"
   LMATRIX=riagendr 0 1 AGE 77670/195850  72816/195850  45364/195850  RIAGENDR*AGE 0 0 0
77670/195850  72816/195850  45364/195850
  /PRINT SUMMARY VARIABLEINFO SAMPLEINFO
  /STATISTICS PARAMETER SE TTEST
  /DOMAIN VARIABLE= sel(1)
  /TEST TYPE=F PADJUST=LSD
  /MISSING CLASSMISSING=EXCLUDE
  /CRITERIA CILEVEL=95.

*Age adjusted prevalence by race 18+ years
CSGLM  hbpx BY race age
  /PLAN FILE='mec4yr_plan.csaplan'
  /MODEL race age race*age
  /INTERCEPT INCLUDE=NO SHOW=YES
  /CUSTOM LABEL="NH-White"
   LMATRIX=race 1 0 0 0 age 77670/195850  72816/195850  45364/195850  race*age 77670/195850
72816/195850  45364/195850  0 0 0  0 0 0  0 0 0
```

```
  /CUSTOM LABEL="NH-Black"
   LMATRIX=race 0 1 0 0 age 77670/195850  72816/195850  45364/195850  race*age 0 0 0
77670/195850  72816/195850  45364/195850  0 0 0  0 0 0
  /CUSTOM LABEL="Mex-Am"
   LMATRIX=race 0 0 1 0 age 77670/195850  72816/195850  45364/195850  race*age 0 0 0  0 0 0
77670/195850  72816/195850  45364/195850  0 0 0
  /CUSTOM LABEL="Other"
   LMATRIX=race 0 0 0 1 age 77670/195850  72816/195850  45364/195850  race*age 0 0 0  0 0 0  0 0 0
77670/195850  72816/195850  45364/195850
  /PRINT SUMMARY VARIABLEINFO SAMPLEINFO
  /STATISTICS PARAMETER SE TTEST
  /DOMAIN VARIABLE= sel(1)
  /TEST TYPE=F PADJUST=LSD
  /MISSING CLASSMISSING=EXCLUDE
  /CRITERIA CILEVEL=95.

*Age adjusted prevalence by gender and race 18+ years
CSGLM  hbpx BY riagendr race age
  /PLAN FILE='mec4yr_plan.csaplan'
  /MODEL riagendr race age RIAGENDR*race*age
  /INTERCEPT INCLUDE=NO SHOW=YES
  /CUSTOM LABEL="Male NH-White"
   LMATRIX=riagendr 1 0 race 1 0 0 0 age 77670/195850  72816/195850  45364/195850
RIAGENDR*race*age 77670/195850  72816/195850  45364/195850  0 0 0  0 0 0  0 0 0  0 0 0  0 0 0  0 0
0  0 0 0
  /CUSTOM LABEL="Male NH-Black"
   LMATRIX=riagendr 1 0 race 0 1 0 0 age 77670/195850  72816/195850  45364/195850
RIAGENDR*race*age 0 0 0  77670/195850  72816/195850  45364/195850  0 0 0  0 0 0  0 0 0  0 0 0  0 0
0  0 0 0
  /CUSTOM LABEL="Male Mex-Am"
   LMATRIX=riagendr 1 0 race 0 0 1 0 age 77670/195850  72816/195850  45364/195850
RIAGENDR*race*age 0 0 0  0 0 0  77670/195850  72816/195850  45364/195850  0 0 0  0 0 0  0 0 0  0 0
0  0 0 0
  /CUSTOM LABEL="Male Other"
   LMATRIX=riagendr 1 0 race 0 0 0 1 age 77670/195850  72816/195850  45364/195850
RIAGENDR*race*age 0 0 0  0 0 0  0 0 0  77670/195850  72816/195850  45364/195850  0 0 0  0 0 0  0 0
0  0 0 0
  /CUSTOM LABEL="Female NH-White"
   LMATRIX=riagendr 0 1 race 1 0 0 0 age 77670/195850  72816/195850  45364/195850
RIAGENDR*race*age 0 0 0  0 0 0  0 0 0  0 0 0  77670/195850  72816/195850  45364/195850  0 0 0  0 0
0  0 0 0
  /CUSTOM LABEL="Female NH-Black"
   LMATRIX=riagendr 0 1 race 0 1 0 0 age 77670/195850  72816/195850  45364/195850
RIAGENDR*race*age 0 0 0  0 0 0  0 0 0  0 0 0  0 0 0  77670/195850  72816/195850  45364/195850  0 0
0  0 0 0
  /CUSTOM LABEL="Female Mex-Am"
   LMATRIX=riagendr 0 1 race 0 0 1 0 age 77670/195850  72816/195850  45364/195850
RIAGENDR*race*age 0 0 0  0 0 0  0 0 0  0 0 0  0 0 0  0 0 0  77670/195850  72816/195850
45364/195850  0 0 0
  /CUSTOM LABEL="Female Other"
```

```
   LMATRIX=riagendr 0 1 race 0 0 0 1 age 77670/195850  72816/195850  45364/195850
RIAGENDR*race*age 0 0 0  0 0 0  0 0 0  0 0 0  0 0 0  0 0 0  0 0 0  77670/195850  72816/195850
45364/195850
  /PRINT SUMMARY VARIABLEINFO SAMPLEINFO
  /STATISTICS PARAMETER SE TTEST
  /DOMAIN VARIABLE= sel(1)
  /TEST TYPE=F PADJUST=LSD
  /MISSING CLASSMISSING=EXCLUDE
  /CRITERIA CILEVEL=95.

CSGLM  hbpx BY riagendr AGE
  /PLAN FILE='mec4yr_plan.csaplan'
  /MODEL riagendr AGE RIAGENDR*AGE
  /INTERCEPT INCLUDE=NO SHOW=YES
  /CUSTOM LABEL="Male"
   LMATRIX=riagendr 1 0 AGE 77670/195850  72816/195850  45364/195850  RIAGENDR*AGE
77670/195850  72816/195850  45364/195850  0 0 0
  /CUSTOM LABEL="Female"
   LMATRIX=riagendr 0 1 AGE 77670/195850  72816/195850  45364/195850  RIAGENDR*AGE 0 0 0
77670/195850  72816/195850  45364/195850
  /PRINT SUMMARY VARIABLEINFO SAMPLEINFO
  /STATISTICS PARAMETER SE TTEST
  /DOMAIN VARIABLE= sel(1)
  /TEST TYPE=F PADJUST=LSD
  /MISSING CLASSMISSING=EXCLUDE
  /CRITERIA CILEVEL=95.

OMSEND.
GET FILE='_indivTest.sav'.
COMPUTE customID=$CASENUM.

VALUE LABELS customID
   1 "(Model)"
   2 "race"
   3 "age"
   4 "race*age"
   5 "NH-White"
   6 "NH-Black"
   7 "Mex-Am"
   8 "Other"
   9 "(Model)"
   10 "age"
   11 "Total"
   12 "(Model)"
   13 "riagendr"
   14 "race"
   15 "age"
   16 "RIAGENDR*race*age"
   17 "Male NH-White"
   18 "Male NH-Black"
   19 "Male Mex-Am"
   20 "Male Other"
```

```
  21 "Female NH-White"
  22 "Female NH-Black"
  23 "Female Mex-Am"
  24 "Female Other"
  25 "(Model)"
  26 "riagendr"
  27 "AGE"
  28 "RIAGENDR*AGE"
  29 "Male"
  30 "Female"
.
FORMATS ContrastEstimate WaldF(F20.3) Std.Error Sig(F20.8)df1 df2(F4) .
SUMMARIZE
  /TABLES=Label_ customID ContrastEstimate Std.Error df1 df2 WaldF Sig
  /FORMAT=VALIDLIST NOCASENUM TOTAL
  /TITLE='Summary of Tests of Model Effects and of Individual Test Results'
  /MISSING=VARIABLE
  /CELLS=NONE.
```

8. WARNING ISSUES AND CAVEATS OF USING SPSS AGEADJUSTED.PY MODULE

Since CSGLM does not have a native routine to calculate age-adjusted prevalence, the module presented here has been developed as an alternative way to calculate age-adjusted prevalence using SPSS instead of using SAS or SUDAAN. Selected procedures and statistics do vary among SPSS and the aforementioned main statistical package. The following section contains differences that we noted in our initial work. Future notes will be added as soon as they are found in the continuation of our work or reported to us by other SPSS users. We invite you to share any relevant findings with us.

Issue: “Difference found between SAS, SUDAAN, and SPSS Wald F statistics utilized in this example”.

(Damir Spisic’s reply, SPSS)

> “The reason for this difference is that SAS and SPSS do not use the same formulas for the "Wald F" statistic. The basic test developed for complex samples is the Wald chi-square test. There are various adjustments to make this test more robust. The simplest ones consist in using an appropriate F test. While SPSS uses Fellegi formula for this purpose, SAS uses the Shah formula. The computed p-values should be similar in most situations. SPSS offers total of four variations for the Wald test. SUDAAN implements the same four plus the Shah F test. SAS apparently offers the Shah F test only. The "Adjusted Wald F" test is generally the best one to use, but it requires additional computations for the simple random sampling (SRS) covariance matrix that sometimes turns singular. Therefore, Wald F is offered in SPSS as the default”.

Issue: “Warning: The design-based covariance matrix is singular. The validity of results is uncertain" shows up. However when the same analysis is run with SUDAAN we obtain identical results.

(Damir Spisic’s reply, SPSS)

> “Upon inspection of the covariance matrix itself and comparison of results with the other tests in this example we can conclude that it would be safe to ignore this message. Test of singularity depends on the hard coded tolerance and it appears to be overly conservative in this case”.
>
> The indications that the numbers are trustworthy are the following.

1) The covariance matrix looks fine. Variances range from about 3.5 to 107 and this wide range seems to be causing some instability. However, none of the numbers in the matrix are huge or miniscule and zeros appear only for the redundant parameters.
2) The estimates and standard errors of the adjusted estimates for values of SEX * HSP_NHSP in this example are plausible from the previous example containing estimates for values of HSP_NHSP only.
3) After recoding REC_AGE4 into REC_AGE3 by merging age categories 3 and 4 and running the corresponding example, the results turn very similar to the original one. The differences are due to adjustment for categories 3 and 4 becoming identical (.170271). The range of computed variances is smaller and there is no warning issued when using REC_AGE3.

This situation is rather unfortunate. It can happen when there is an interaction between factors and sample design variables. Using CTables you can cross all the factors with your strata and clusters (and subpopulation) and check the cell sizes. You will likely see a lot of zeros there. This in turn can contribute zero variances to the total covariance matrix. Also, correlation between factors and design variables seem to cause problems at times. A way to alleviate the problem would be to merge some of the clusters or get more data. Alternatively, merging category 3 and 4 for the age could also make a difference.

This illustrates some problems of using GLM to compute the tabular data. Still, the parameter and contrast standard errors seem fairly sensible in this example. They could be reasonably accurate in spite of the warning in this case. Also useful could be to print out the covariance matrix in CSGLM. It could reveal which parameters are causing the problem. How? That is easy; you can just add subcommand /PRINT COVB.

9. REFERENCES

Curtin LR, Klein RJ (1995). Direct standardization (age-adjusted death rates). Healthy People Statistical Notes. No.6 (revised).

Klein RJ, Schoenborn CA (2001). Age adjustment using the 2000 projected U.S. population. Healthy People Statistical Notes, no.20. Hyattsville, Maryland: National Center for Health Statistics. Available also at the NCHS website

Further documentation

Anderson RN & Rosenberg HM (1998). Report of the Second Workshop on Age Adjustment. National Center for Health Statistics. Vital Health Stat 4(30).

Feinleb M & Zarate AO (1992). Reconsidering age adjustment procedures: Workshop Procedings. Vital Health Stat 4(29).

Krieger N. Williams D. (2001). Changing to the 2000 Standard Million: Are Declining Racial/Ethnic and Socioeconomic Inequalities in Health Real Progress or Statistical Illusion? American Journal of Public Health. Vol. 91 No 8., 1209-1213.