



THE  
POWER  
TO KNOW.

# **SAS/STAT<sup>®</sup> 12.1 User's Guide The STDRATE Procedure (Chapter)**



This document is an individual chapter from *SAS/STAT® 12.1 User's Guide*.

The correct bibliographic citation for the complete manual is as follows: SAS Institute Inc. 2012. *SAS/STAT® 12.1 User's Guide*. Cary, NC: SAS Institute Inc.

Copyright © 2012, SAS Institute Inc., Cary, NC, USA

All rights reserved. Produced in the United States of America.

**For a Web download or e-book:** Your use of this publication shall be governed by the terms established by the vendor at the time you acquire this publication.

The scanning, uploading, and distribution of this book via the Internet or any other means without the permission of the publisher is illegal and punishable by law. Please purchase only authorized electronic editions and do not participate in or encourage electronic piracy of copyrighted materials. Your support of others' rights is appreciated.

**U.S. Government Restricted Rights Notice:** Use, duplication, or disclosure of this software and related documentation by the U.S. government is subject to the Agreement with SAS Institute and the restrictions set forth in FAR 52.227-19, Commercial Computer Software-Restricted Rights (June 1987).

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

Electronic book 1, August 2012

SAS® Publishing provides a complete selection of books and electronic products to help customers use SAS software to its fullest potential. For more information about our e-books, e-learning products, CDs, and hard-copy books, visit the SAS Publishing Web site at [support.sas.com/publishing](http://support.sas.com/publishing) or call 1-800-727-3228.

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

# Chapter 88

## The STD RATE Procedure

### Contents

Overview: STD RATE Procedure . . . . .	7487
Getting Started: STD RATE Procedure . . . . .	7489
Syntax: STD RATE Procedure . . . . .	7497
PROC STD RATE Statement . . . . .	7498
BY Statement . . . . .	7501
POPULATION Statement . . . . .	7501
REFERENCE Statement . . . . .	7502
STRATA Statement . . . . .	7503
Details: STD RATE Procedure . . . . .	7505
Rate . . . . .	7505
Risk . . . . .	7508
Direct Standardization . . . . .	7510
Mantel-Haenszel Effect Estimation . . . . .	7512
Indirect Standardization and Standardized Morbidity/Mortality Ratio . . . . .	7514
Attributable Fraction and Population Attributable Fraction . . . . .	7517
Applicable Data Sets and Required Variables for Method Specifications . . . . .	7520
Applicable Confidence Limits for Rate and Risk Statistics . . . . .	7521
Table Output . . . . .	7522
ODS Table Names . . . . .	7527
Graphics Output . . . . .	7528
ODS Graphics . . . . .	7528
Examples: STD RATE Procedure . . . . .	7529
Example 88.1: Comparing Directly Standardized Rates . . . . .	7529
Example 88.2: Computing Mantel-Haenszel Risk Estimation . . . . .	7535
Example 88.3: Computing Attributable Fraction Estimates . . . . .	7543
Example 88.4: Displaying SMR Results from BY Groups . . . . .	7550
References . . . . .	7558

### Overview: STD RATE Procedure

Epidemiology is the study of the occurrence and distribution of health-related states or events in specified populations. Epidemiology also includes the study of the determinants that influence these states, and the application of this knowledge to control health problems (Porta 2008). It is a discipline that describes, quantifies, and postulates causal mechanisms for health phenomena in populations (Friss and Sellers 2009).

A common goal is to establish relationships between various factors (such as exposure to a specific chemical) and the event outcomes (such as incidence of disease). But the measure of an association between an exposure and an event outcome can be biased due to confounding. That is, the association of the exposure to some other variables, such as age, influences the occurrence of the event outcome. With confounding, the usual effect between an exposure and an event outcome can be biased because some of the effect might be accounted for by other variables. For example, with an event rate discrepancy among different age groups of a population, the overall crude rate might not provide a useful summary statistic to compare populations.

One strategy to control confounding is stratification. In stratification, a population is divided into several subpopulations according to specific criteria for the confounding variables, such as age and sex groups. The effect of the exposure on the event outcome is estimated within each stratum, and then these stratum-specific effect estimates are combined into an overall estimate.

Two commonly used event frequency measures are rate and risk:

- A rate is a measure of the frequency with which an event occurs in a defined population in a specified period of time. It measures the change in one quantity per unit of another quantity. For example, an event rate measures how fast the events are occurring. That is, an event rate of a population over a specified time period can be defined as the number of new events divided by population-time (Kleinbaum, Kupper, and Morgenstern 1982, p. 100) over the same time period.
- A risk is the probability that an event occurs in a specified time period. It is assumed that only one event can occur in the time period for each subject or item. The overall crude risk of a population over a specified time period is the number of new events in the time period divided by the population size at the beginning of the time period.

Standardized overall rate and risk estimates based on stratum-specific estimates can be derived with the effects of confounding variables removed. These estimates provide useful summary statistics and allow valid comparison of the populations. There are two types of standardization:

- Direct standardization computes the weighted average of stratum-specific estimates in the study population, using the weights from a standard or reference population. This standardization is applicable when the study population is large enough to provide stable stratum-specific estimates. The directly standardized estimate is the overall crude rate in the study population if it has the same strata distribution as the reference population. When standardized estimates for different populations are derived by using the same reference population, the resulting estimates can also be compared by using the estimated difference and estimated ratio statistics.
- Indirect standardization computes the weighted average of stratum-specific estimates in the reference population, using the weights from the study population. The ratio of the overall crude rate or risk in the study population and the corresponding weighted estimate in the reference population is the standardized morbidity ratio (SMR). This ratio is also the standardized mortality ratio if the event is death. SMR is used to compare rates or risks in the study and reference populations. With SMR, the indirectly standardized estimate is then computed as the product of the SMR and the overall crude estimate for the reference population. SMR and indirect standardization are applicable even when the study population is so small that the resulting stratum-specific rates are not stable.

Assuming that an effect, such as the rate difference between two populations, is homogeneous across strata, each stratum provides an estimate of the same effect. A pooled estimate of the effect can then be derived



from these stratum-specific effect estimates. One way to estimate a homogeneous effect is the Mantel-Haenszel method (Greenland and Rothman 2008, p. 271). For a homogeneous rate difference effect between two populations, the Mantel-Haenszel estimate is identical to the difference between two directly standardized rates, but with weights derived from the two populations instead of from an explicitly specified reference population. The Mantel-Haenszel method can also be applied to other homogeneous effects between populations, such as the rate ratio, risk difference, and risk ratio.

The STD RATE procedure computes directly standardized rates and risks for study populations. For two study populations with the same reference population, PROC STD RATE compares directly standardized rates or risks from these two populations. For homogeneous effects across strata, PROC STD RATE computes Mantel-Haenszel estimates. The STD RATE procedure also computes indirectly standardized rates and risks, including SMR.

The attributable fraction measures the excess event rate or risk fraction in the exposed population that can be attributed to the exposure. The rate or risk ratio statistic is required in the attributable fraction computation, and the STD RATE procedure estimates the ratio by using either SMR or the rate ratio statistic in the Mantel-Haenszel estimates.

Although the STD RATE procedure provides useful summary standardized statistics, standardization is not a substitute for individual comparisons of stratum-specific estimates. PROC STD RATE provides summary statistics, such as rate and risk estimates and their confidence limits, in each stratum. In addition, PROC STD RATE also displays these stratum-specific statistics by using ODS Graphics.

Note that the term standardization has different meanings in other statistical applications. For example, the STDIZE procedure standardizes numeric variables in a SAS data set by subtracting a location measure and dividing by a scale measure.

---

## Getting Started: STD RATE Procedure

This example illustrates indirect standardization and uses the standardized mortality ratio to compare the death rate from skin cancer between people who live in the state of Florida and people who live in the United States as a whole.

The Florida\_C43 data set contains the stratum-specific mortality information for skin cancer in year 2000 for the state of Florida (Florida Department of Health 2000, 2012). The variable **Age** is a grouping variable that forms the strata in the standardization, and the variables **Event** and **PYear** identify the number of events and total person-years, respectively. The COMMA11. format is specified in the DATA step to input numerical values that contain commas in **PYear**.

```
data Florida_C43;
  input Age $1-5 Event PYear comma11.;
  datalines;
00-04      0      953,785
05-14      0  1,997,935
15-24      4  1,885,014
25-34     14  1,957,573
35-44     43  2,356,649
45-54     72  2,088,000
55-64     70  1,548,371
65-74    126  1,447,432
```

```

75-84  136  1,087,524
85+    73   335,944
;

```

The US\_C43 data set contains the corresponding stratum-specific mortality information for the United States in year 2000 (Miniño et al. 2002; U.S. Bureau of Census 2011). The variable **Age** is the grouping variable, and the variables **Event** and **PYear** identify the number of events and the total person-years, respectively.

```

data US_C43;
  input Age $1-5 Event comma7. PYear comma12.;
  datalines;
00-04      0  19,175,798
05-14      1  41,077,577
15-24     41  39,183,891
25-34    186  39,892,024
35-44    626  45,148,527
45-54   1,199  37,677,952
55-64   1,303  24,274,684
65-74   1,637  18,390,986
75-84   1,624  12,361,180
85+      803   4,239,587
;

```

The following statements invoke the STD RATE procedure and request indirect standardization to compare death rates between the state of Florida and the United States:

```

ods graphics on;
proc stdrate data=Florida_C43 refdata=US_C43
  method=indirect
  stat=rate(mult=100000)
  plots=all
  ;
  population event=Event total=PYear;
  reference  event=Event total=PYear;
  strata Age / stats smr;
run;
ods graphics off;

```

The **DATA=** and **REFDATA=** options name the study data set and reference data set, respectively. The **METHOD=INDIRECT** option requests indirect standardization. The **STAT=RATE** option specifies the rate as the frequency measure for standardization, and the **MULT=100000** suboption (which is the default) displays the rates per 100,000 person-years in the table output and graphics output. The **PLOTS=ALL** option requests all appropriate plots with indirect standardization.

The POPULATION statement specifies the options that are related to the study population, and the EVENT= and TOTAL= options specify variables for the number of events and person-years in the study population, respectively.

The REFERENCE statement specifies the options related to the reference population, and the EVENT= and TOTAL= options specify variables for the number of events and person-years in the reference population, respectively.

The STRATA statement lists the variable Age that forms the strata. The STATS option requests a strata information table that contains stratum-specific statistics such as rates, and the SMR option requests a table of stratum-specific SMR estimates.

The “Standardization Information” table in [Figure 88.1](#) displays the standardization information.

**Figure 88.1** Standardization Information

The STD RATE Procedure	
Standardization Information	
Data Set	WORK.FLORIDA_C43
Reference Data Set	WORK.US_C43
Method	Indirect Standardization
Statistic	Rate
Number of Strata	10
Rate Multiplier	100000

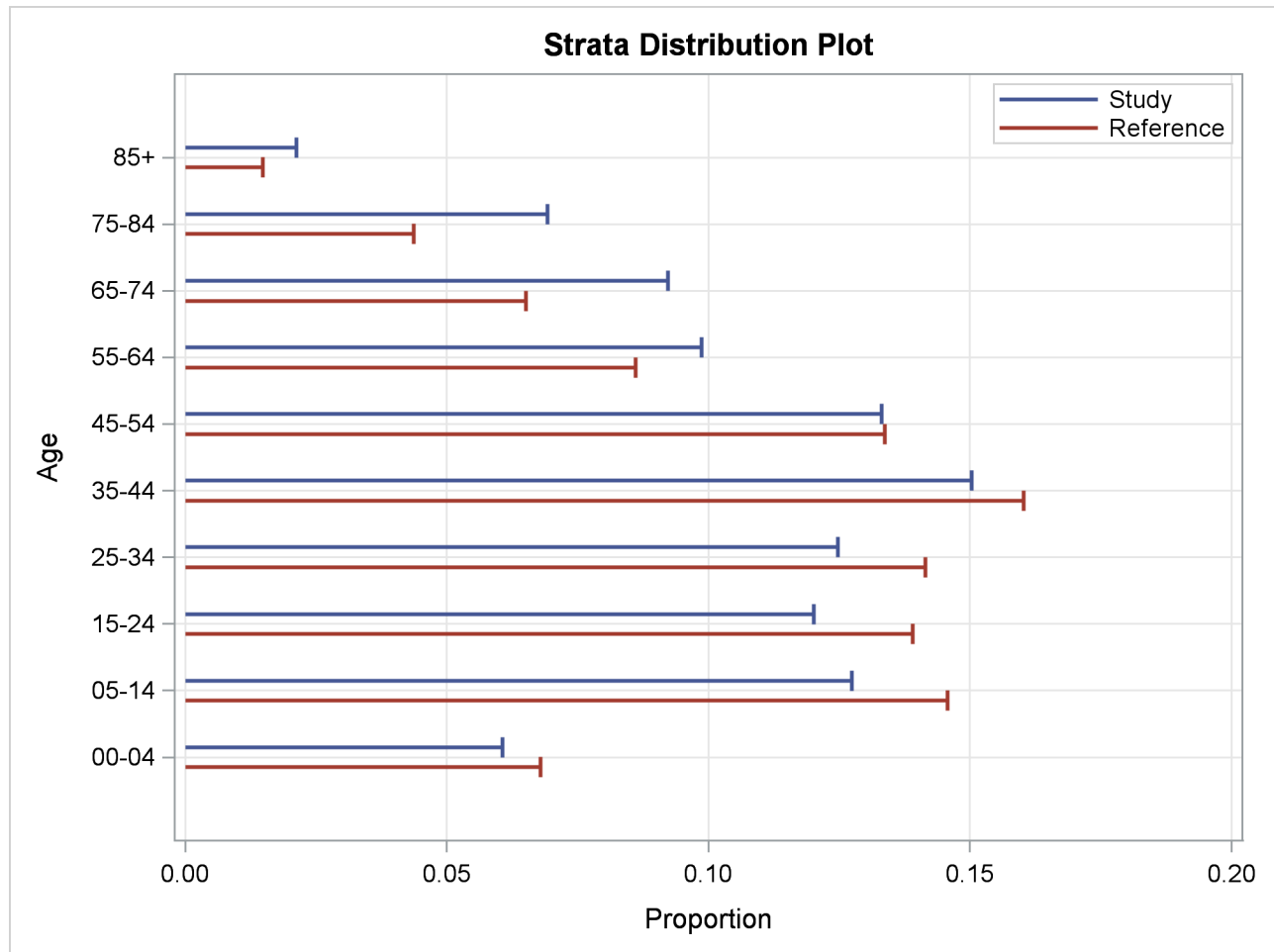
The STATS option in the STRATA statement requests that the “Indirectly Standardized Strata Statistics” table in [Figure 88.2](#) display the strata information and expected number of events at each stratum. The MULT=100000 suboption in the STAT=RATE option requests that crude rates per 100,000 person-years be displayed. The Expected Events column displays the expected number of events when the stratum-specific rates in the reference data set are applied to the corresponding person-years in the study data set.

**Figure 88.2** Strata Information (Indirect Standardization)

The STDRA Procedure						
Indirectly Standardized Strata Statistics						
Rate Multiplier = 100000						
Stratum Index	Age	-----Study Population-----			Crude Rate	Standard Error
		Observed Events	-----Population-Time----- Value	Proportion		
1	00-04	0	953785	0.0609	0.0000	0.00000
2	05-14	0	1997935	0.1276	0.0000	0.00000
3	15-24	4	1885014	0.1204	0.2122	0.10610
4	25-34	14	1957573	0.1250	0.7152	0.19114
5	35-44	43	2356649	0.1505	1.8246	0.27825
6	45-54	72	2088000	0.1333	3.4483	0.40638
7	55-64	70	1548371	0.0989	4.5209	0.54035
8	65-74	126	1447432	0.0924	8.7051	0.77551
9	75-84	136	1087524	0.0695	12.5055	1.07234
10	85+	73	335944	0.0215	21.7298	2.54328
Indirectly Standardized Strata Statistics						
Rate Multiplier = 100000						
Stratum Index	-Study Population-		-----Reference Population-----		Crude Rate	Expected Events
	95% Normal Confidence Limits		-----Population-Time----- Value	Proportion		
1	0.0000	0.0000	19175798	0.0681	0.0000	0.000
2	0.0000	0.0000	41077577	0.1460	0.0024	0.049
3	0.0042	0.4202	39183891	0.1392	0.1046	1.972
4	0.3405	1.0898	39892024	0.1418	0.4663	9.127
5	1.2793	2.3700	45148527	0.1604	1.3865	32.676
6	2.6518	4.2448	37677952	0.1339	3.1822	66.445
7	3.4618	5.5799	24274684	0.0863	5.3677	83.112
8	7.1851	10.2250	18390986	0.0654	8.9011	128.837
9	10.4037	14.6072	12361180	0.0439	13.1379	142.878
10	16.7451	26.7146	4239587	0.0151	18.9405	63.630

With ODS Graphics enabled, the PLOTS=ALL option displays all appropriate plots. With indirect standardization and a rate statistic, these plots include the strata distribution plot, the strata rate plot, and the strata SMR plot. By default, strata levels are displayed on the vertical axis for these plots.

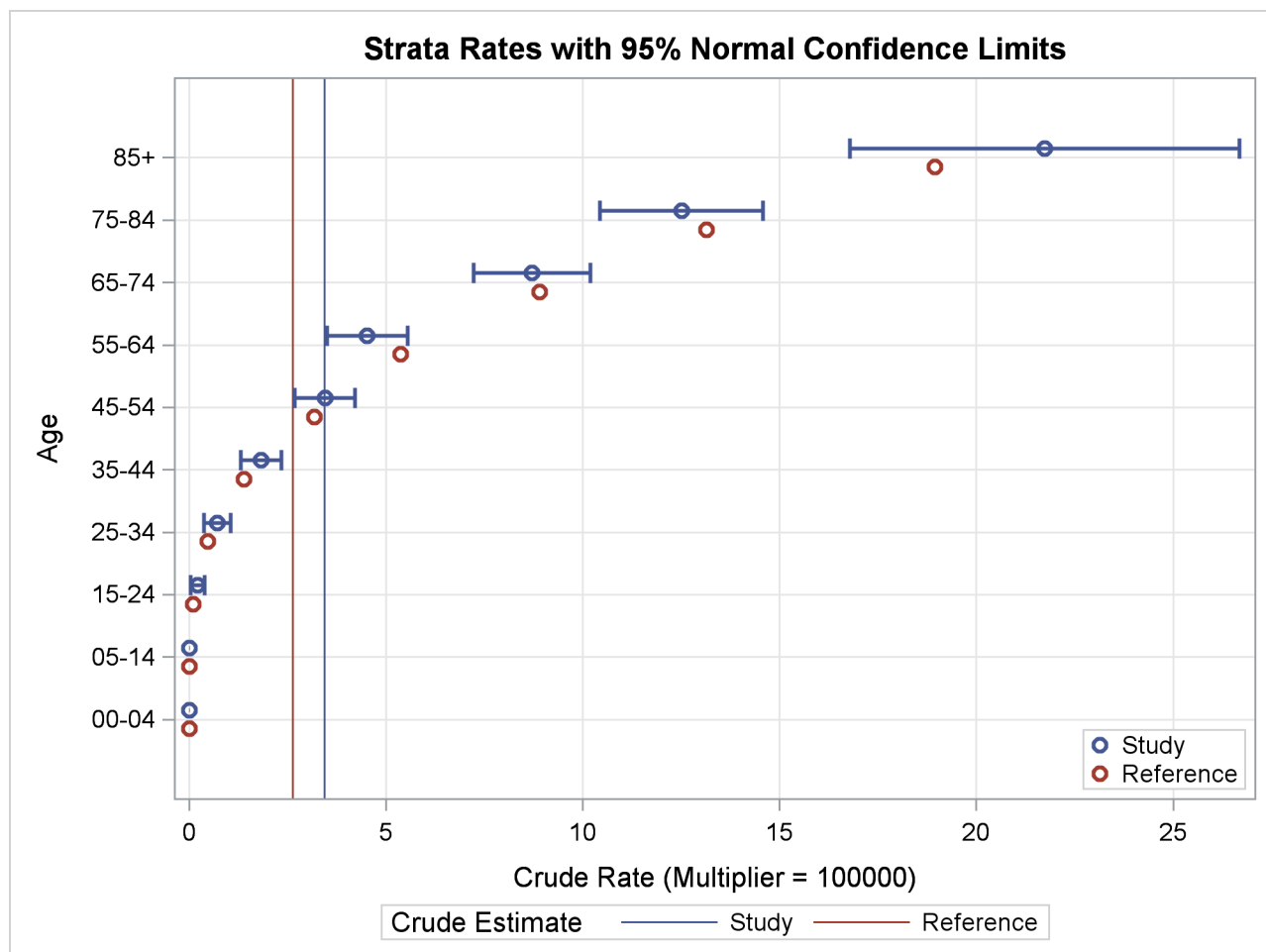
The strata distribution plot displays proportions for stratum-specific person-years in the study and reference populations, as shown in [Figure 88.3](#).

**Figure 88.3** Strata Distribution Plot

The strata distribution plot displays the proportions in the “Indirectly Standardized Strata Statistics” table in [Figure 88.2](#). In the plot, the proportions of the study population are identified by the blue lines, and the proportions of the reference population are identified by the red lines. The plot shows that the study population has higher proportions in older age groups and lower proportions in younger age groups than the reference population.

The strata rate plot displays stratum-specific rate estimates in the study and reference populations, as shown in [Figure 88.4](#). This plot displays the rate estimates in the “Indirectly Standardized Strata Statistics” table in [Figure 88.2](#). In addition, the plot displays the confidence limits for the rate estimates in the study population and the overall crude rates for the two populations.

Figure 88.4 Strata Rate Plot



The SMR option in the STRATA statement requests that the “Strata SMR Estimates” table in Figure 88.5 display the strata SMR at each stratum. The MULT=100000 suboption in the STAT=RATE option requests that the reference rates per 100,000 person-years be displayed.

**Figure 88.5** Strata SMR Information

Strata SMR Estimates							
Rate Multiplier = 100000							
Stratum Index	Age	---Study Observed Events	Population-- Population-- Time	Reference Crude Rate	Expected Events	SMR	Standard Error
1	00-04	0	953785	0.0000	0.000	.	.
2	05-14	0	1997935	0.0024	0.049	0.0000	.
3	15-24	4	1885014	0.1046	1.972	2.0280	1.0140
4	25-34	14	1957573	0.4663	9.127	1.5339	0.4099
5	35-44	43	2356649	1.3865	32.676	1.3160	0.2007
6	45-54	72	2088000	3.1822	66.445	1.0836	0.1277
7	55-64	70	1548371	5.3677	83.112	0.8422	0.1007
8	65-74	126	1447432	8.9011	128.837	0.9780	0.0871
9	75-84	136	1087524	13.1379	142.878	0.9519	0.0816
10	85+	73	335944	18.9405	63.630	1.1473	0.1343

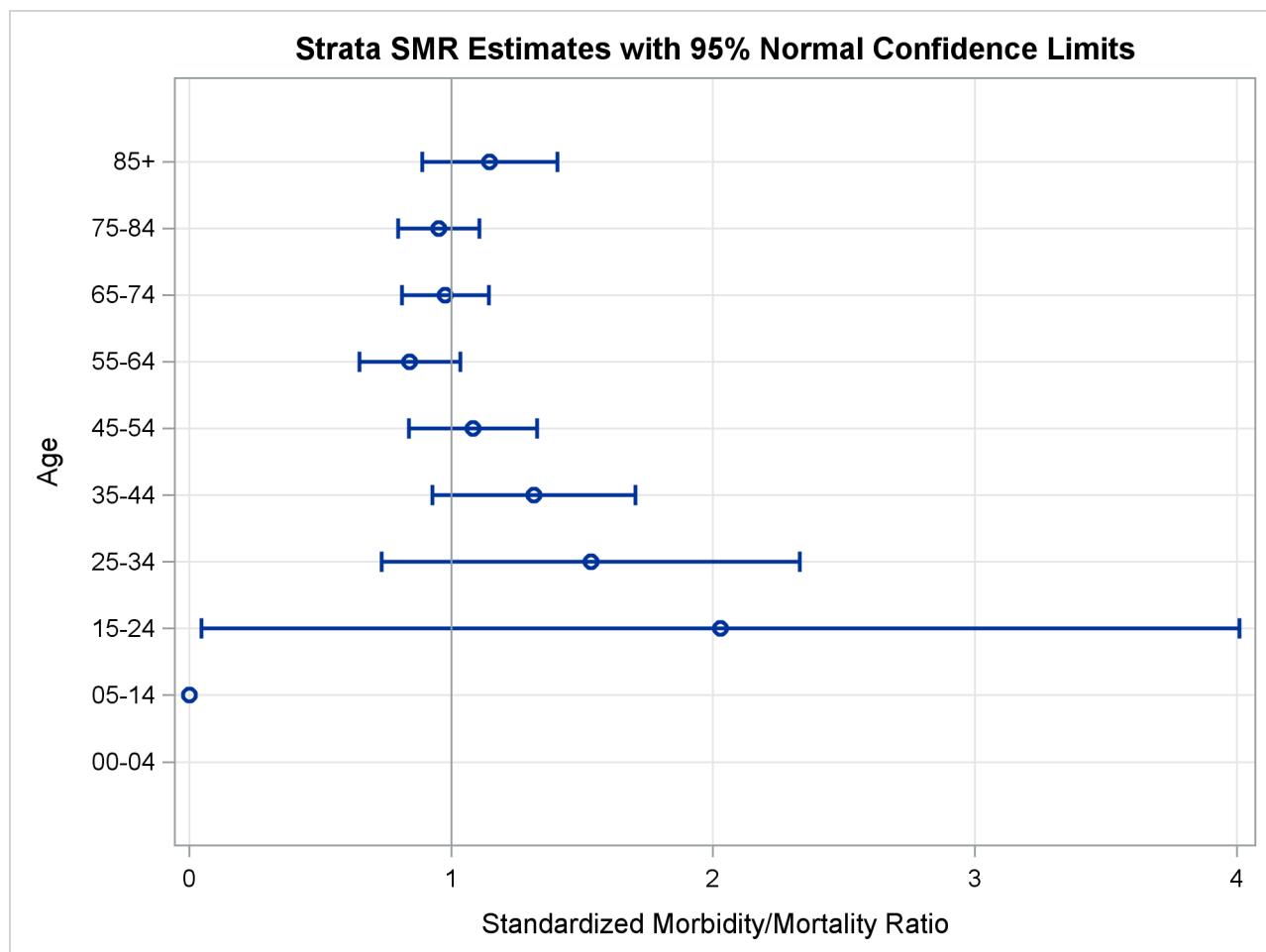
  

Strata SMR Estimates		
Rate Multiplier = 100000		
Stratum Index	95% Normal Confidence Limits	
1	.	.
2	.	.
3	0.0406	4.0154
4	0.7304	2.3373
5	0.9226	1.7093
6	0.8333	1.3339
7	0.6449	1.0395
8	0.8072	1.1487
9	0.7919	1.1118
10	0.8841	1.4104

The “Strata SMR Estimates” table shows that although SMR is less than 1 only at three age strata (55–64, 65–74, and 75–84), these three strata contain about 60% of the total events.

The strata SMR plot displays stratum-specific SMR estimates with confidence limits, as shown in [Figure 88.6](#). The plot displays the SMR estimates in the “Strata SMR Estimates” table in [Figure 88.5](#).

Figure 88.6 Strata SMR Plot



The METHOD=INDIRECT option requests that the “Standardized Morbidity/Mortality Ratio” table in Figure 88.7 be displayed. The table displays the SMR, its confidence limits, and the test for the null hypothesis  $H_0 : \text{SMR} = 1$ . The default ALPHA=0.05 option requests that 95% confidence limits be constructed.

Figure 88.7 Standardized Morbidity/Mortality Ratio

Standardized Morbidity/Mortality Ratio							
Observed Events	Expected Events	SMR	Standard Error	95% Normal Confidence Limits		Z	Pr >  Z
538	528.726	1.0175	0.0439	0.9316	1.1035	0.40	0.6893

The 95% normal confidence limits contain the null hypothesis value  $\text{SMR} = 1$ , and the hypothesis of  $\text{SMR} = 1$  is not rejected at the  $\alpha = 0.05$  level from the normal test.

The “Indirectly Standardized Rate Estimates” table in Figure 88.8 displays the indirectly standardized rate and related statistics.



**Figure 88.8** Standardized Rate Estimates (Indirect Standardization)

Indirectly Standardized Rate Estimates					
Rate Multiplier = 100000					
-----Study Population-----			Reference		
Observed	Population-	Crude	Crude	Expected	
Events	Time	Rate	Rate	Events	SMR
538	15658227	3.4359	2.6366	528.726	1.0175
Indirectly Standardized Rate Estimates					
Rate Multiplier = 100000					
-----Standardized Rate-----					
		Standard	95% Normal		
Estimate		Error	Confidence	Limits	
2.6829		0.1157	2.4562	2.9096	

The indirectly standardized rate estimate is the product of the SMR and the crude rate estimate for the reference population. The table shows that although the crude rate in the state of Florida (3.4359) is 30% higher than the crude rate in the U.S. (2.6366), the resulting standardized rate (2.6829) is close to the crude rate in the U.S.

## Syntax: STD RATE Procedure

The following statements are available in PROC STD RATE:

```
PROC STD RATE < options > ;
  BY variables ;
  POPULATION options ;
  REFERENCE options ;
  STRATA variables < / option > ;
```

The PROC STD RATE statement invokes the procedure, names the data sets, specifies the standardization method, and identifies the statistic for standardization. The BY statement requests separate analyses of groups defined by the BY variables. The required POPULATION statement specifies the rate or risk information in study populations, and the REFERENCE statement specifies the rate or risk information in the reference population. The STRATA statement lists the variables that form the strata.

The following sections describe the PROC STD RATE statement and then describe the other statements in alphabetical order.

## PROC STD RATE Statement

**PROC STD RATE** < options > ;

Table 88.1 summarizes the *options* in the PROC STD RATE statement.

**Table 88.1** Summary of PROC STD RATE Options

Option	Description
<b>Input Data Sets</b>	
<b>DATA=</b>	Names the SAS data set that contains the study populations
<b>REFDATA=</b>	Names the SAS data set that contains the reference population
<b>Standardization Methods</b>	
<b>METHOD=</b>	Specifies the method for standardization
<b>STAT=</b>	Specifies the statistic for standardization
<b>EFFECT</b>	Specifies the test to compare study populations for direct standardization and Mantel-Haenszel estimation
<b>Displayed Output</b>	
<b>ALPHA=</b>	Specifies the significance level for confidence intervals
<b>CL=</b>	Requests the confidence limits for the standardized estimates
<b>PLOTS</b>	Requests stratum-specific plots

You can specify the following *options* in the PROC STD RATE statement to compute standardized rates and risks in the procedure. They are listed in alphabetical order.

### **ALPHA=** $\alpha$

requests that confidence limits be constructed with confidence level  $100(1 - \alpha)\%$ , where  $0 < \alpha < 1$ . The default is ALPHA=0.05. These confidence limits include confidence limits for the stratum-specific rates or risks, standardized rate and risk, standardized morbidity/mortality ratio, and population attributable rate and risk.

### **CL=GAMMA** < (TYPE=AVERAGE | CONSERVATIVE) > | LOGNORMAL | NONE | NORMAL | POISSON

specifies the method to construct confidence limits for SMR and standardized rate and risk. You can specify the following values for this option:

#### **GAMMA**

requests confidence limits based on a gamma distribution for METHOD=DIRECT and METHOD=MH. This value applies only when STAT=RATE. You can specify the TYPE=CONSERVATIVE suboption to request conservative confidence limits that are based on a gamma distribution and were developed by Fay and Feuer (1997), or you can specify the TYPE=AVERAGE suboption to request modified confidence limits proposed by Tiwari, Clegg, and Zou (2006).

**LOGNORMAL**

requests confidence limits based on a lognormal distribution.

**NONE**

suppresses construction of confidence limits.

**NORMAL**

requests confidence limits based on a normal distribution.

**POISSON**

requests confidence limits based on a Poisson distribution. This value applies only when METHOD=INDIRECT.

The default is CL=NORMAL.

**DATA=SAS-data-set**

names the required SAS data set that contains the event information in the study populations.

**EFFECT <=DIFF | RATIO>**

displays a table of the effect estimate and associated confidence limits. This option applies only when METHOD=DIRECT with two study populations and when METHOD=MH, where two study populations are required.

EFFECT and EFFECT=RATIO display a test on the ratio effect of estimates between the study populations, and the EFFECT=DIFF option displays a test on the difference effect.

**METHOD= DIRECT | INDIRECT <(AF)> | MH <(AF)>****M= DIRECT | INDIRECT <(AF)> | MH <(AF)>**

specifies the required method for standardization. METHOD=DIRECT requests direct standardization, METHOD=INDIRECT requests indirect standardization, and METHOD=MH requests Mantel-Haenszel estimation.

The AF suboption requests the attributable fraction that measures the excess event rate or risk fraction in the exposed population that is attributable to the exposure. This suboption also requests the population attributable fraction that measures the excess event rate or risk fraction in the total population that is attributable to the exposure.

**PLOTS <( global-options )> <= plot-request>****PLOTS <( global-options )> <= ( plot-request <...plot-request> )>**

specifies options that control the details of the plots. The default is PLOTS=RATE for STAT=RATE and PLOTS=RISK for STAT=RISK.

You can specify the following *global-options*:

**DISPLAY=INDEX | LEVEL**

specifies tick mark values for the strata axis. DISPLAY=LEVEL displays strata levels on the strata axis, and DISPLAY=INDEX displays strata indices of sequential strata identification numbers on the strata axis. The default is DISPLAY=LEVEL.

**ONLY**

suppresses the default plots and displays only plots that are specifically requested.

**STRATUM=HORIZONTAL | VERTICAL**

controls the orientation of the plots. STRATUM=VERTICAL places the strata information on the vertical axis, and STRATUM=HORIZONTAL places the strata information on the horizontal axis. The default is STRATUM=VERTICAL.

You can specify the following *plot-requests*:

**ALL**

produces all appropriate plots.

**DIST | DISTRIBUTION**

displays a plot of the proportions for stratum-specific exposed time or sample size.

**EFFECT**

displays a plot of the stratum-specific effect estimates and associated confidence limits. This option applies only when METHOD=DIRECT with two study populations and when METHOD=MH, where two study populations are required. If the EFFECT=DIFF option is specified, the stratum-specific rate or risk difference effects are displayed. Otherwise, the stratum-specific rate or risk ratio effects are displayed.

**NONE**

suppresses all plots.

**RATE**

displays a plot of the stratum-specific rates and associated confidence limits. This option applies only when STAT=RATE. If a confidence limits method is specified in the STATS(CL=) option in the STRATA statement, that method is used to compute the confidence limits. Otherwise, the normal approximation is used.

**RISK**

displays a plot of the stratum-specific risks and associated confidence limits. This option applies only when STAT=RISK. If a confidence limits method is specified in the STATS(CL=) option in the STRATA statement, that method is used to compute the confidence limits. Otherwise, the normal approximation is used.

**SMR**

displays a plot of the stratum-specific SMR estimates and associated confidence limits. This option applies only when METHOD=INDIRECT. If a method is specified in the SMR(CL=) option in the STRATA statement, that method is used to compute the confidence limits. Otherwise, the normal approximation is used.

**REFDATA=SAS-data-set**

names the required SAS data set that contains the event information in the reference population.

**STAT=RATE** < ( **MULT** =*c* ) >

**STAT=RISK**

specifies the statistic for standardization. **STAT=RATE** computes standardized rates, and **STAT=RISK** computes standardized risks. The default is **STAT=RATE**.

The **MULT=** suboption in the **STAT=RATE** option specifies a power of 10 constant *c*, and requests that rates per *c* population-time units be displayed in the output tables and graphics. The default is **MULT=100000**, which specifies rates per 100,000 population-time units.

---

## BY Statement

**BY** *variables* ;

You can specify a **BY** statement with **PROC STDRATE** to obtain separate analyses of observations in groups that are defined by the **BY** variables. When a **BY** statement appears, the procedure expects the input data set to be sorted in order of the **BY** variables. If you specify more than one **BY** statement, only the last one specified is used.

If your input data set is not sorted in ascending order, use one of the following alternatives:

- Sort the data by using the **SORT** procedure with a similar **BY** statement.
- Specify the **NOTSORTED** or **DESCENDING** option in the **BY** statement for the **STDRATE** procedure. The **NOTSORTED** option does not mean that the data are unsorted but rather that the data are arranged in groups (according to values of the **BY** variables) and that these groups are not necessarily in alphabetical or increasing numeric order.
- Create an index on the **BY** variables by using the **DATASETS** procedure (in Base SAS software).

For more information about **BY**-group processing, see the discussion in *SAS Language Reference: Concepts*. For more information about the **DATASETS** procedure, see the discussion in the *Base SAS Procedures Guide*.

---

## POPULATION Statement

**POPULATION** < *options* > ;

The required **POPULATION** statement specifies the information in the study data set. You can specify the following *options* in the **POPULATION** statement:

**EVENT=***variable*

specifies the *variable* for the number of events in the study data set.

**GROUP** < ( **ORDER=DATA** | **FORMATTED** | **INTERNAL** ) > =*variable*

specifies the *variable* whose values identify the various populations. This option is required when **METHOD=MH** and also applies when **METHOD=DIRECT** in the **PROC STDRATE** statement.

The optional ORDER= suboption specifies the order in which the values of the *variable* are to be displayed. You can specify the following values for the ORDER= suboption:

<b>DATA</b>	sorts by the order in which the values appear in the input data set.
<b>FORMATTED</b>	sorts by their external formatted values.
<b>INTERNAL</b>	sorts by the unformatted values, which yields the same order that the SORT procedure does.

By default, ORDER=INTERNAL. For ORDER=FORMATTED and ORDER=INTERNAL, the sort order is machine-dependent.

**POPEVENT=number**

specifies the total number of events in the study data set. This option applies only when METHOD=INDIRECT is specified in the PROC STD RATE statement and the total number of events is not available in the study data set.

**RATE < ( MULT=c ) > = variable**

specifies the *variable* for the observed rate in the study data set. This option applies only when STAT=RATE is specified in the PROC STD RATE statement. The MULT=c suboption specifies a power of 10 constant *c* and requests that the rates per *c* population-time units be read from the data set. The default is the value of the MULT= suboption used in the STAT=RATE option in the PROC STD RATE statement.

**RISK=variable**

specifies the *variable* for the observed risk in the study data set. This option applies only when STAT=RISK is specified in the PROC STD RATE statement.

**TOTAL=variable**

specifies the *variable* for either the population-time (STAT=RATE) or the number of observations (STAT=RISK) in the study data set.

---

## REFERENCE Statement

**REFERENCE < options > ;**

The REFERENCE statement specifies the information in the reference data set. This statement is required when METHOD=DIRECT or METHOD=INDIRECT is specified in the PROC STD RATE statement.

You can specify the following *options* in the REFERENCE statement:

**EVENT=variable**

specifies the *variable* for the number of events in the reference data set.

**RATE < ( MULT=c ) > = variable**

specifies the *variable* for the observed rate in the reference data set. This option applies only when STAT=RATE is specified in the PROC STD RATE statement. The MULT=c suboption specifies a power of 10 constant *c* and requests that the rates per *c* population-time units be read from the data set. The default is the value of the MULT= suboption used in the STAT=RATE option in the PROC STD RATE statement.

**RISK=***variable*

specifies the *variable* for the observed risk in the reference data set. This option applies only when STAT=RISK is specified in the PROC STD RATE statement.

**TOTAL=***variable*

specifies the *variable* for either the population-time (STAT=RATE) or the number of observations (STAT=RISK) in the reference data set.

When METHOD=INDIRECT is specified in the PROC STD RATE statement, the overall reference population rate and risk are needed to compute indirect standardized rate and risk, respectively. If the information is not available in the reference data set, you can specify the following *options* for overall reference population rate and risk.

**REFEVENT=***number*

specifies the total *number* of events in the reference data set.

**REFRATE < ( MULT=*c* ) > =** *number*

specifies the crude rate in the reference data set. This option applies only when STAT=RATE is specified in the PROC STD RATE statement. The MULT=*c* suboption specifies a power of 10 constant *c*, and the *number* is the crude rate per *c* population-time units in the data set. The default is the value of the MULT= suboption in the STAT=RATE option in the PROC STD RATE statement.

**REFRISK=***number*

specifies the crude risk in the reference data set. This option applies only when STAT=RISK is specified in the PROC STD RATE statement.

**REFTOTAL=***number*

specifies either the total population-time (STAT=RATE) or the total number of observations (STAT=RISK) in the reference data set.

When STAT=RATE, the REFRATE= option specifies the crude reference rate for the indirect standardized rate. If the REFRATE= option is not specified, the REFEVENT= and REFTOTAL options can be used to compute the crude reference rate. Similarly, when STAT=RISK, the REFRISK= option specifies the crude reference risk for the indirect standardized rate. If the REFRISK= option is not specified, the REFEVENT= and REFTOTAL options can be used to compute the crude reference risk.

---

## STRATA Statement

**STRATA** *variables* </ options> ;

The STRATA statement names *variables* that form the strata in the standardization. The combinations of categories of STRATA variables define the strata in the population.

The STRATA variables are one or more variables in all input data sets. These variables can be either character or numeric. The formatted values of the STRATA variables determine the levels. Thus, you can use formats to group values into levels. See the FORMAT procedure in the *Base SAS Procedures Guide* and the FORMAT statement and SAS formats in *SAS Language Reference: Dictionary* for more information.

When the STRATA statement is not specified or the statement is specified without variables, all observations in a data set are treated as though they are from a single stratum.

You can specify the following *options* in the STRATA statement after a slash (/):

**EFFECT**

displays a table of the stratum-specific effect estimates and associated confidence limits. This option applies only when METHOD=DIRECT with two study populations and when METHOD=MH, where two study populations are required. If the EFFECT=DIFF option in the PROC STD RATE statement is specified, the stratum-specific rate or risk difference effects are displayed. Otherwise, the stratum-specific rate or risk ratio effects are displayed.

**MISSING**

treats missing values as a valid (nonmissing) category for all STRATA variables. When PROC STD RATE determines levels of a STRATA variable, an observation with missing values for that STRATA variable is excluded, unless the MISSING option is specified.

**ORDER=DATA | FORMATTED | INTERNAL**

specifies the order in which the values of the categorical variables are to be displayed. You can specify the following values for the ORDER= option:

<b>DATA</b>	sorts by the order in which the values appear in the input data set.
<b>FORMATTED</b>	sorts by their external formatted values.
<b>INTERNAL</b>	sorts by the unformatted values, which yields the same order that the SORT procedure does.

By default, ORDER=INTERNAL. For ORDER=FORMATTED and ORDER=INTERNAL, the sort order is machine-dependent.

**STATS < ( CL=LOGNORMAL | NONE | NORMAL | POISSON ) >**

displays tables for stratum-specific statistics such as stratum-specific rates and risks. You can specify the following values of the CL= suboption to request confidence limits for the rate or risk estimate in each stratum:

**LOGNORMAL**

requests confidence limits based on a lognormal approximation.

**NONE**

suppresses confidence limits.

**NORMAL**

requests confidence limits based on a normal approximation and also displays the standard error for the rate estimate in each stratum.

**POISSON**

requests confidence limits based on a Poisson distribution for stratum-specific rates. This values applies only when STAT=RATE in the PROC STD RATE statement.

The default is CL=NORMAL.

**SMR < ( CL=LOGNORMAL | NONE | NORMAL | POISSON ) >**

displays tables for stratum-specific SMR estimates. This option applies only when METHOD=INDIRECT is specified in the PROC STD RATE statement. You can specify the following values of the CL= suboption to request confidence limits for the SMR estimate in each stratum:



**LOGNORMAL**

requests confidence limits based on a lognormal approximation.

**NONE**

suppresses confidence limits.

**NORMAL**

requests confidence limits based on a normal approximation and also displays the standard error for the SMR estimate in each stratum.

**POISSON**

requests confidence limits based on a Poisson distribution for stratum-specific SMR estimates. This values applies only when STAT=RATE in the PROC STDRATE statement.

The default is CL=NORMAL.

---

## Details: STDRATE Procedure

---

### Rate

A major task in epidemiology is to compare event frequencies for groups of people. Both rate and risk are commonly used to measure event frequency in the comparison. Rate is a measure of change in one quantity per unit of another quantity. An event rate measures how fast the events are occurring. In contrast, an event risk is the probability that an event occurs over a specified follow-up time period.

An event rate of a population over a specified time period can be defined as the number of new events divided by the population-time of the population over the same time period,

$$\hat{\lambda} = \frac{d}{\mathcal{T}}$$

where  $d$  is the number of events and  $\mathcal{T}$  is the population-time that is computed by adding up the time contributed by each subject in the population over the specified time period.

For a general population, the subsets (strata) might not be homogeneous enough to have a similar rate. Thus, the rate for each stratum should be computed separately to reflect this discrepancy. For a population that consists of  $K$  homogeneous strata (such as different age groups), the stratum-specific rate for the  $j$ th stratum in a population is computed as

$$\hat{\lambda}_j = \frac{d_j}{\mathcal{T}_j}$$

where  $d_j$  is the number of events and  $\mathcal{T}_j$  is the population-time for subjects in the  $j$ th stratum of the population.

Assuming the number of events in the  $j$ th stratum,  $d_j$ , has a Poisson distribution, the variance of  $\hat{\lambda}_j$  is

$$V(\hat{\lambda}_j) = V\left(\frac{d_j}{\mathcal{T}_j}\right) = \frac{1}{\mathcal{T}_j^2} V(d_j) = \frac{d_j}{\mathcal{T}_j^2} = \frac{\hat{\lambda}_j}{\mathcal{T}_j}$$

By using the method of statistical differentials (Elandt-Johnson and Johnson 1980, pp. 70–71), the variance of the logarithm of rate can be estimated by

$$V(\log(\hat{\lambda}_j)) = \frac{1}{\hat{\lambda}_j^2} V(\hat{\lambda}_j) = \frac{1}{\hat{\lambda}_j^2} \frac{\hat{\lambda}_j}{\mathcal{T}_j} = \frac{1}{\hat{\lambda}_j \mathcal{T}_j} = \frac{1}{d_j}$$

Because the rate value can be very small, especially for rare events, it is sometimes expressed in terms of the product of a multiplier and the rate itself. For example, a rate can be expressed as the number of events per 100,000 person-years.

### Normal Distribution Confidence Interval for Rate

A  $(1 - \alpha)$  confidence interval for  $\hat{\lambda}_j$  based on a normal distribution is given by

$$\left( \hat{\lambda}_j - z \sqrt{V(\hat{\lambda}_j)}, \hat{\lambda}_j + z \sqrt{V(\hat{\lambda}_j)} \right)$$

where  $z = \Phi^{-1}(1 - \alpha/2)$  is the  $(1 - \alpha/2)$  quantile of the standard normal distribution.

### Lognormal Distribution Confidence Interval for Rate

A  $(1 - \alpha)$  confidence interval for  $\log(\hat{\lambda}_j)$  based on a normal distribution is given by

$$\left( \log(\hat{\lambda}_j) - z \sqrt{V(\log(\hat{\lambda}_j))}, \log(\hat{\lambda}_j) + z \sqrt{V(\log(\hat{\lambda}_j))} \right)$$

where  $z = \Phi^{-1}(1 - \alpha/2)$  is the  $(1 - \alpha/2)$  quantile of the standard normal distribution and the variance  $V(\log(\hat{\lambda}_j)) = 1/d_j$ .

Thus, a  $(1 - \alpha)$  confidence interval for  $\hat{\lambda}_j$  based on a lognormal distribution is given by

$$\left( \hat{\lambda}_j e^{-\frac{z}{\sqrt{d_j}}}, \hat{\lambda}_j e^{\frac{z}{\sqrt{d_j}}} \right)$$

### Poisson Distribution Confidence Interval for Rate

Denote the  $(\alpha/2)$  quantile for the  $\chi^2$  distribution with  $2d_j$  degrees of freedom by

$$q_{lj} = (\chi_{2d_j}^2)^{-1}(\alpha/2)$$

Denote the  $(1 - \alpha/2)$  quantiles for the  $\chi^2$  distribution with  $2(d_j + 1)$  degrees of freedom by

$$q_{uj} = (\chi_{2(d_j+1)}^2)^{-1}(1 - \alpha/2)$$

Then a  $(1 - \alpha)$  confidence interval for  $\hat{\lambda}_j$  based on the  $\chi^2$  distribution is given by

$$\left( \frac{q_{lj}}{2\mathcal{T}_j}, \frac{q_{uj}}{2\mathcal{T}_j} \right)$$

### Confidence Interval for Rate Difference Statistic

For rate estimates from two independent samples,  $\hat{\lambda}_{1j}$  and  $\hat{\lambda}_{2j}$ , a  $(1 - \alpha)$  confidence interval for the rate difference  $\hat{\lambda}_{dj} = \hat{\lambda}_{1j} - \hat{\lambda}_{2j}$  is

$$\left( \hat{\lambda}_{dj} - z \sqrt{V(\hat{\lambda}_{dj})}, \hat{\lambda}_{dj} + z \sqrt{V(\hat{\lambda}_{dj})} \right)$$

where  $z = \Phi^{-1}(1 - \alpha/2)$  is the  $(1 - \alpha/2)$  quantile of the standard normal distribution and the variance

$$V(\hat{\lambda}_{dj}) = V(\hat{\lambda}_{1j}) + V(\hat{\lambda}_{2j})$$

### Confidence Interval for Rate Ratio Statistic

For rate estimates from two independent samples,  $\hat{\lambda}_{1j}$  and  $\hat{\lambda}_{2j}$ , a  $(1 - \alpha)$  confidence interval for the log rate ratio statistic  $\log(\hat{\lambda}_{rj}) = \log(\hat{\lambda}_{1j}/\hat{\lambda}_{2j})$  is

$$\left( \log(\hat{\lambda}_{rj}) - z \sqrt{V(\log(\hat{\lambda}_{rj}))}, \log(\hat{\lambda}_{rj}) + z \sqrt{V(\log(\hat{\lambda}_{rj}))} \right)$$

where  $z = \Phi^{-1}(1 - \alpha/2)$  is the  $(1 - \alpha/2)$  quantile of the standard normal distribution and the variance

$$V(\log(\hat{\lambda}_{rj})) = V(\log(\hat{\lambda}_{1j})) + V(\log(\hat{\lambda}_{2j}))$$

Thus, a  $(1 - \alpha)$  confidence interval for the rate ratio statistic  $\hat{\lambda}_{rj}$  is given by

$$\left( \frac{\hat{\lambda}_{1j}}{\hat{\lambda}_{2j}} e^{-z \sqrt{V(\log(\hat{\lambda}_{rj}))}}, \frac{\hat{\lambda}_{1j}}{\hat{\lambda}_{2j}} e^{z \sqrt{V(\log(\hat{\lambda}_{rj}))}} \right)$$

### Confidence Interval for Rate SMR

At stratum  $j$ , a stratum-specific standardized morbidity/mortality ratio is

$$\mathcal{R}_j = \frac{d_j}{\mathcal{E}_j}$$

where  $\mathcal{E}_j$  is the expected number of events.

With the rate

$$\hat{\lambda}_j = \frac{d_j}{\mathcal{T}_j}$$

SMR can be expressed as

$$\mathcal{R}_j = \frac{\mathcal{T}_j}{\mathcal{E}_j} \hat{\lambda}_j$$

Thus, a  $(1 - \alpha)$  confidence interval for  $\mathcal{R}_j$  is given by

$$\left( \frac{\mathcal{T}_j}{\mathcal{E}_j} \hat{\lambda}_{jl}, \frac{\mathcal{T}_j}{\mathcal{E}_j} \hat{\lambda}_{ju} \right)$$

where  $(\hat{\lambda}_{jl}, \hat{\lambda}_{ju})$  is a  $(1 - \alpha)$  confidence interval for the rate  $\hat{\lambda}_j$ .

## Risk

An event risk of a population over a specified time period can be defined as the number of new events in the follow-up time period divided by the event-free population size at the beginning of the time period,

$$\hat{\gamma} = \frac{d}{\mathcal{N}}$$

where  $\mathcal{N}$  is the population size.

For a general population, the subsets (strata) might not be homogeneous enough to have a similar risk. Thus, the risk for each stratum should be computed separately to reflect this discrepancy. For a population that consists of  $K$  homogeneous strata (such as different age groups), the stratum-specific risk for the  $j$ th stratum in a population is computed as

$$\hat{\gamma}_j = \frac{d_j}{\mathcal{N}_j}$$

where  $\mathcal{N}_j$  is the population size in the  $j$ th stratum of the population.

Assuming the number of events,  $d_j$ , has a binomial distribution, then a variance estimate of  $\hat{\gamma}_j$  is

$$V(\hat{\gamma}_j) = \frac{\hat{\gamma}_j(1 - \hat{\gamma}_j)}{\mathcal{N}_j}$$

By using the method of statistical differentials (Elandt-Johnson and Johnson 1980, pp. 70–71), the variance of the logarithm of risk can be estimated by

$$V(\log(\hat{\gamma}_j)) = \frac{1}{\hat{\gamma}_j^2} V(\hat{\gamma}_j) = \frac{1}{\hat{\gamma}_j^2} \frac{\hat{\gamma}_j(1 - \hat{\gamma}_j)}{\mathcal{N}_j} = \frac{1 - \hat{\gamma}_j}{\hat{\gamma}_j \mathcal{N}_j} = \frac{1}{d_j} - \frac{1}{\mathcal{N}_j}$$

## Normal Distribution Confidence Interval for Risk

A  $(1 - \alpha)$  confidence interval for  $\hat{\gamma}_j$  based on a normal distribution is given by

$$\left( \hat{\gamma}_j - z \sqrt{V(\hat{\gamma}_j)}, \hat{\gamma}_j + z \sqrt{V(\hat{\gamma}_j)} \right)$$

where  $z = \Phi^{-1}(1 - \alpha/2)$  is the  $(1 - \alpha/2)$  quantile of the standard normal distribution.

## Lognormal Distribution Confidence Interval for Risk

A  $(1 - \alpha)$  confidence interval for  $\log(\hat{\gamma}_j)$  based on a normal distribution is given by

$$\left( \log(\hat{\gamma}_j) - z \sqrt{V(\log(\hat{\gamma}_j))}, \log(\hat{\gamma}_j) + z \sqrt{V(\log(\hat{\gamma}_j))} \right)$$

where  $z = \Phi^{-1}(1 - \alpha/2)$  is the  $(1 - \alpha/2)$  quantile of the standard normal distribution and the variance  $V(\log(\hat{\gamma}_j)) = 1/d_j - 1/\mathcal{N}_j$ .

Thus, a  $(1 - \alpha)$  confidence interval for  $\hat{\gamma}_j$  based on a lognormal distribution is given by

$$\left( \hat{\gamma}_j e^{-z \sqrt{\frac{1}{d_j} - \frac{1}{\mathcal{N}_j}}}, \hat{\gamma}_j e^{z \sqrt{\frac{1}{d_j} - \frac{1}{\mathcal{N}_j}}} \right)$$

### Confidence Interval for Risk Difference Statistic

For rate estimates from two independent samples,  $\hat{\gamma}_{1j}$  and  $\hat{\gamma}_{2j}$ , a  $(1 - \alpha)$  confidence interval for the risk difference  $\hat{\gamma}_{dj} = \hat{\gamma}_{1j} - \hat{\gamma}_{2j}$  is

$$\left( \hat{\gamma}_{dj} - z \sqrt{V(\hat{\gamma}_{dj})}, \hat{\gamma}_{dj} + z \sqrt{V(\hat{\gamma}_{dj})} \right)$$

where  $z = \Phi^{-1}(1 - \alpha/2)$  is the  $(1 - \alpha/2)$  quantile of the standard normal distribution and the variance

$$V(\hat{\gamma}_{dj}) = V(\hat{\gamma}_{1j}) + V(\hat{\gamma}_{2j})$$

### Confidence Interval for Risk Ratio Statistic

For rate estimates from two independent samples,  $\hat{\gamma}_{1j}$  and  $\hat{\gamma}_{2j}$ , a  $(1 - \alpha)$  confidence interval for the log risk ratio statistic  $\log(\hat{\gamma}_{rj}) = \log(\hat{\gamma}_{1j}/\hat{\gamma}_{2j})$  is

$$\left( \log(\hat{\gamma}_{rj}) - z \sqrt{V(\log(\hat{\gamma}_{rj}))}, \log(\hat{\gamma}_{rj}) + z \sqrt{V(\log(\hat{\gamma}_{rj}))} \right)$$

where  $z = \Phi^{-1}(1 - \alpha/2)$  is the  $(1 - \alpha/2)$  quantile of the standard normal distribution and the variance

$$V(\log(\hat{\gamma}_{rj})) = V(\log(\hat{\gamma}_{1j})) + V(\log(\hat{\gamma}_{2j}))$$

Thus, a  $(1 - \alpha)$  confidence interval for the risk ratio statistic  $\hat{\gamma}_{rj}$  is given by

$$\left( \frac{\hat{\gamma}_{1j}}{\hat{\gamma}_{2j}} e^{-z \sqrt{V(\log(\hat{\gamma}_{rj}))}}, \frac{\hat{\gamma}_{1j}}{\hat{\gamma}_{2j}} e^{z \sqrt{V(\log(\hat{\gamma}_{rj}))}} \right)$$

### Confidence Interval for Risk SMR

At stratum  $j$ , a stratum-specific standardized morbidity/mortality ratio is

$$\mathcal{R}_j = \frac{d_j}{\mathcal{E}_j}$$

where  $\mathcal{E}_j$  is the expected number of events.

With the risk

$$\hat{\gamma}_j = \frac{d_j}{\mathcal{N}_j}$$

SMR can be expressed as

$$\mathcal{R}_j = \frac{\mathcal{N}_j}{\mathcal{E}_j} \hat{\gamma}_j$$

Thus, a  $(1 - \alpha)$  confidence interval for  $\mathcal{R}_j$  is given by

$$\left( \frac{\mathcal{N}_j}{\mathcal{E}_j} \hat{\gamma}_{jl}, \frac{\mathcal{N}_j}{\mathcal{E}_j} \hat{\gamma}_{ju} \right)$$

where  $(\hat{\gamma}_{jl}, \hat{\gamma}_{ju})$  is a  $(1 - \alpha)$  confidence interval for the risk  $\hat{\gamma}_j$ .

## Direct Standardization

Direct standardization uses the weights from a reference population to compute the standardized rate of a study group as the weighted average of stratum-specific rates in the study population. The standardized rate is computed as

$$\hat{\lambda}_{ds} = \frac{\sum_j \mathcal{T}_{rj} \hat{\lambda}_{sj}}{\mathcal{T}_r}$$

where  $\hat{\lambda}_{sj}$  is the rate in the  $j$ th stratum of the study population,  $\mathcal{T}_{rj}$  is the population-time in the  $j$ th stratum of the reference population, and  $\mathcal{T}_r = \sum_k \mathcal{T}_{rk}$  is the population-time in the reference population.

Similarly, direct standardization uses the weights from a reference population to compute the standardized risk of a study group as the weighted average of stratum-specific risks in the study population. The standardized risk is computed as

$$\hat{\gamma}_{ds} = \frac{\sum_j \mathcal{N}_{rj} \hat{\gamma}_{sj}}{\mathcal{N}_r}$$

where  $\hat{\gamma}_{sj}$  is the risk in the  $j$ th stratum of the study population,  $\mathcal{N}_{rj}$  is the number of observations in the  $j$ th stratum of the reference population, and  $\mathcal{N}_r = \sum_k \mathcal{N}_{rk}$  is the total number of observations in the reference population.

That is, the directly standardized rate and risk of a study population are weighted averages of the stratum-specific rates and risks, respectively, where the weights are the corresponding strata population sizes in the reference population. The direct standardization can be used when the study population is large enough to provide stable stratum-specific rates or risks. When the same reference population is used for multiple study populations, directly standardized rates and risks provide valid comparisons between study populations.

The variances of the directly standardized rate and risk are

$$V(\hat{\lambda}_{ds}) = V\left(\frac{\sum_j \mathcal{T}_{rj} \hat{\lambda}_{sj}}{\mathcal{T}_r}\right) = \frac{\sum_j \mathcal{T}_{rj}^2 V(\hat{\lambda}_{sj})}{\mathcal{T}_r^2}$$

$$V(\hat{\gamma}_{ds}) = V\left(\frac{\sum_j \mathcal{N}_{rj} \hat{\gamma}_{sj}}{\mathcal{N}_r}\right) = \frac{\sum_j \mathcal{N}_{rj}^2 V(\hat{\gamma}_{sj})}{\mathcal{N}_r^2}$$

By using the method of statistical differentials (Elandt-Johnson and Johnson 1980, pp. 70–71), the variance of the logarithm of directly standardized rate and risk can be estimated by

$$V(\log(\hat{\lambda}_{ds})) = \frac{1}{\hat{\lambda}_{ds}^2} V(\hat{\lambda}_{ds})$$

$$V(\log(\hat{\gamma}_{ds})) = \frac{1}{\hat{\gamma}_{ds}^2} V(\hat{\gamma}_{ds})$$

The confidence intervals for  $\hat{\lambda}_{ds}$  and  $\hat{\gamma}_{ds}$  can be constructed based on normal and lognormal distributions. A gamma distribution confidence interval can also be constructed for  $\hat{\lambda}_{ds}$ .

In the next four subsections,  $\beta = \lambda$  denotes the rate statistic and  $\beta = \gamma$  denotes the risk statistic.

### Normal Distribution Confidence Intervals for Standardized Rate and Risk

A  $(1 - \alpha)$  confidence interval for  $\hat{\beta}_{ds}$  based on a normal distribution is then given by

$$\left( \hat{\beta}_{ds} - z \sqrt{V(\hat{\beta}_{ds})}, \hat{\beta}_{ds} + z \sqrt{V(\hat{\beta}_{ds})} \right)$$

where  $z = \Phi^{-1}(1 - \alpha/2)$  is the  $(1 - \alpha/2)$  quantile of the standard normal distribution.

### Lognormal Distribution Confidence Intervals for Standardized Rate and Risk

A  $(1 - \alpha)$  confidence interval for  $\log(\hat{\beta}_{ds})$  based on a normal distribution is given by

$$\left( \log(\hat{\beta}_{ds}) - z \sqrt{V(\log(\hat{\beta}_{ds}))}, \log(\hat{\beta}_{ds}) + z \sqrt{V(\log(\hat{\beta}_{ds}))} \right)$$

where  $z = \Phi^{-1}(1 - \alpha/2)$  is the  $(1 - \alpha/2)$  quantile of the standard normal distribution.

Thus, a  $(1 - \alpha)$  confidence interval for  $\hat{\beta}_{ds}$  based on a lognormal distribution is given by

$$\left( \hat{\beta}_{ds} e^{-z \sqrt{V(\log(\hat{\beta}_{ds}))}}, \hat{\beta}_{ds} e^{z \sqrt{V(\log(\hat{\beta}_{ds}))}} \right)$$

### Gamma Distribution Confidence Interval for Standardized Rate

Fay and Feuer (1997) use the relationship between the Poisson and gamma distributions to derive approximate confidence intervals for the standardized rate  $\hat{\lambda}_{ds}$  based on the gamma distribution. As in the construction of the asymptotic normal confidence intervals, it is assumed that the number of events has a Poisson distribution, and the standardized rate is a weighted sum of independent Poisson random variables. A confidence interval for  $\hat{\lambda}_{ds}$  is then given by

$$\left( \frac{v}{2\hat{\lambda}_{ds}} (\chi^2)^{-1}_{\frac{2\hat{\lambda}_{ds}^2}{v}} \left( \frac{\alpha}{2} \right), \frac{v + w_x^2}{2(\hat{\lambda}_{ds} + w_x)} (\chi^2)^{-1}_{\frac{2(\hat{\lambda}_{ds} + w_x)^2}{v + w_x^2}} \left( 1 - \frac{\alpha}{2} \right) \right)$$

where

$$v = \sum_j w_j^2 \frac{\hat{\lambda}_{sj}}{\mathcal{T}_{sj}}$$

$$w_j = \frac{\mathcal{T}_{rj}}{\mathcal{T}_r} \frac{1}{\mathcal{T}_{sj}}$$

and  $w_x$  is the maximum  $w_j$ .

Tiwari, Clegg, and Zou (2006) propose a less conservative confidence interval for  $\hat{\lambda}_{ds}$  with a different upper confidence limit,

$$\left( \frac{v}{2\hat{\lambda}_{ds}} (\chi^2)^{-1}_{\frac{2\hat{\lambda}_{ds}^2}{v}} \left( \frac{\alpha}{2} \right), \frac{v + w_{2m}}{2(\hat{\lambda}_{ds} + w_m)} (\chi^2)^{-1}_{\frac{2(\hat{\lambda}_{ds} + w_m)^2}{v + w_{2m}}} \left( 1 - \frac{\alpha}{2} \right) \right)$$

where  $w_m$  is the average  $w_j$  and  $w_{2m}$  is the average  $w_j^2$ .

## Comparing Standardized Rates and Comparing Standardized Risks

By using the same reference population, two directly standardized rates or risks from different populations can be compared. Both the difference and ratio statistics can be used in the comparison. Assume that  $\hat{\beta}_1$  and  $\hat{\beta}_2$  are directly standardized rates or risks for two populations with variances  $V(\hat{\beta}_1)$  and  $V(\hat{\beta}_2)$ , respectively. The difference test assumes that the difference statistic

$$\hat{\beta}_1 - \hat{\beta}_2$$

has a normal distribution with mean 0 under the null hypothesis  $H_0 : \beta_1 = \beta_2$ . The variance is given by

$$V(\hat{\beta}_1 - \hat{\beta}_2) = V(\hat{\beta}_1) + V(\hat{\beta}_2)$$

The ratio test assumes that the log ratio statistic,

$$\log \left( \frac{\hat{\beta}_1}{\hat{\beta}_2} \right)$$

has a normal distribution with mean 0 under the null hypothesis  $H_0 : \beta_1 = \beta_2$ , or equivalently,  $\log(\beta_1/\beta_2) = 0$ . An estimated variance is given by

$$V \left( \log \left( \frac{\hat{\beta}_1}{\hat{\beta}_2} \right) \right) = V(\log(\hat{\beta}_1)) + V(\log(\hat{\beta}_2)) = \frac{1}{\hat{\beta}_1^2} V(\hat{\beta}_1) + \frac{1}{\hat{\beta}_2^2} V(\hat{\beta}_2)$$

---

## Mantel-Haenszel Effect Estimation

In direct standardization, the derived standardized rates and risks in a study population are the weighted average of the stratum-specific rates and risks in the population, respectively, where the weights are given by the population-time for standardized rate and the number of observations for standardized risk in a reference population.

Assuming that an effect, such as rate difference, rate ratio, risk difference, and risk ratio between two populations, is homogeneous across strata, the Mantel-Haenszel estimates of this effect can be constructed from directly standardized rates or risks in the two populations, where the weights are constructed from the stratum-specific population-times for rate and number of observations for risk of the two populations.

That is, for population  $k$ ,  $k=1$  and  $2$ , the standardized rate and risk are

$$\hat{\lambda}_k = \frac{\sum_j w_j \hat{\lambda}_{kj}}{\sum_j w_j} \quad \text{and} \quad \hat{\gamma}_k = \frac{\sum_j w_j \hat{\gamma}_{kj}}{\sum_j w_j}$$

where the weights are

$$w_j = \frac{\mathcal{T}_{1j} \mathcal{T}_{2j}}{\mathcal{T}_{1j} + \mathcal{T}_{2j}}$$

for standardized rate, and

$$w_j = \frac{\mathcal{N}_{1j} \mathcal{N}_{2j}}{\mathcal{N}_{1j} + \mathcal{N}_{2j}}$$

for standardized risk.



### Rate and Risk Difference Statistics

Denote  $\beta = \lambda$  for rate and  $\beta = \gamma$  for risk. The variance is

$$V(\hat{\beta}_k) = V\left(\frac{\sum_j w_j \hat{\beta}_{kj}}{\sum_j w_j}\right) = \frac{1}{(\sum_j w_j)^2} \sum_j w_j^2 V(\hat{\beta}_{kj})$$

The Mantel-Haenszel difference statistic is

$$\hat{\beta}_1 - \hat{\beta}_2$$

with variance

$$V(\hat{\beta}_1 - \hat{\beta}_2) = V(\hat{\beta}_1) + V(\hat{\beta}_2)$$

Under the null hypothesis  $H_0 : \beta_1 = \beta_2$ , the difference statistic  $\hat{\beta}_1 - \hat{\beta}_2$  has a normal distribution with mean 0.

### Rate Ratio Statistic

The Mantel-Haenszel rate ratio statistic is  $\hat{\lambda}_1/\hat{\lambda}_2$ , and the log ratio statistic is

$$\log\left(\frac{\hat{\lambda}_1}{\hat{\lambda}_2}\right)$$

Under the null hypothesis  $H_0 : \lambda_1 = \lambda_2$  (or equivalently,  $\log(\lambda_1/\lambda_2) = 0$ ), the log ratio statistic has a normal distribution with mean 0 and variance

$$V\left(\log\left(\frac{\hat{\lambda}_1}{\hat{\lambda}_2}\right)\right) = \frac{\sum_j w_j \hat{\lambda}_{pj}}{(\sum_j w_j \hat{\lambda}_{1j}) (\sum_j w_j \hat{\lambda}_{2j})}$$

where

$$\hat{\lambda}_{pj} = \frac{d_{1j} + d_{2j}}{\mathcal{T}_{1j} + \mathcal{T}_{2j}}$$

is the combined rate estimate in stratum  $j$  under the null hypothesis of equal rates (Greenland and Robins 1985; Greenland and Rothman 2008, p. 273).

### Risk Ratio Statistic

The Mantel-Haenszel risk ratio statistic is  $\hat{\gamma}_1/\hat{\gamma}_2$ , and the log ratio statistic is

$$\log\left(\frac{\hat{\gamma}_1}{\hat{\gamma}_2}\right)$$

Under the null hypothesis  $H_0 : \gamma_1 = \gamma_2$  (or equivalently,  $\log(\gamma_1/\gamma_2) = 0$ ), the log ratio statistic has a normal distribution with mean 0 and variance

$$V\left(\log\left(\frac{\hat{\gamma}_1}{\hat{\gamma}_2}\right)\right) = \frac{\sum_j w_j (\hat{\gamma}_{pj} - \hat{\gamma}_{1j}\hat{\gamma}_{2j})}{(\sum_j w_j \hat{\gamma}_{1j}) (\sum_j w_j \hat{\gamma}_{2j})}$$

where

$$\hat{\gamma}_{pj} = \frac{d_{1j} + d_{2j}}{\mathcal{N}_{1j} + \mathcal{N}_{2j}}$$

is the combined risk estimate in stratum  $j$  under the null hypothesis of equal risks (Greenland and Robins 1985; Greenland and Rothman 2008, p. 275).

---

## Indirect Standardization and Standardized Morbidity/Mortality Ratio

Indirect standardization compares the rates of the study and reference populations by applying the stratum-specific rates in the reference population to the study population, where the stratum-specific rates might not be reliable.

The expected number of events in the study population is

$$\mathcal{E} = \sum_j \mathcal{T}_{sj} \hat{\lambda}_{rj}$$

where  $\mathcal{T}_{sj}$  is the population-time in the  $j$ th stratum of the study population and  $\hat{\lambda}_{rj}$  is the rate in the  $j$ th stratum of the reference population.

With the expected number of events,  $\mathcal{E}$ , the standardized morbidity ratio or standardized mortality ratio can be expressed as

$$\mathcal{R}_{sm} = \frac{\mathcal{D}}{\mathcal{E}}$$

where  $\mathcal{D}$  is the observed number of events (Breslow and Day 1987, p. 65).

The ratio  $\mathcal{R}_{sm} > 1$  indicates that the mortality rate or risk in the study population is larger than the estimate in the reference population, and  $\mathcal{R}_{sm} < 1$  indicates that the mortality rate or risk in the study population is smaller than the estimate in the reference population.

With the ratio  $\mathcal{R}_{sm}$ , an indirectly standardized rate for the study population is computed as

$$\hat{\lambda}_{is} = \mathcal{R}_{sm} \hat{\lambda}_r$$

where  $\hat{\lambda}_r$  is the overall crude rate in the reference population.

Similarly, to compare the risks of the study and reference populations, the stratum-specific risks in the reference population are used to compute the expected number of events in the study population

$$\mathcal{E} = \sum_j \mathcal{N}_{sj} \hat{\gamma}_{rj}$$

where  $\mathcal{N}_{sj}$  is the number of observations in the  $j$ th stratum of the study population and  $\hat{\gamma}_{rj}$  is the risk in the  $j$ th stratum of the reference population.

Also, with the standardized morbidity ratio  $\mathcal{R}_{sm} = \mathcal{D}/\mathcal{E}$ , an indirectly standardized risk for the study population is computed as

$$\hat{\gamma}_{is} = \mathcal{R}_{sm} \hat{\gamma}_r$$

where  $\hat{\gamma}_r$  is the overall crude risk in the reference population.

The observed number of events in the study population is  $\mathcal{D} = \sum_j d_{sj}$ , where  $d_{sj}$  is the number of events in the  $j$ th stratum of the population. For the rate estimate, if  $d_{sj}$  has a Poisson distribution, then the variance of the standardized mortality ratio  $\mathcal{R}_{sm} = \mathcal{D} / \mathcal{E}$  is

$$V(\mathcal{R}_{sm}) = \frac{1}{\mathcal{E}^2} \sum_j V(d_{sj}) = \frac{1}{\mathcal{E}^2} \sum_j d_{sj} = \frac{\mathcal{D}}{\mathcal{E}^2} = \frac{\mathcal{R}_{sm}}{\mathcal{E}}$$

For the risk estimate, if  $d_{sj}$  has a binomial distribution, then the variance of  $\mathcal{R}_{sm} = \mathcal{D} / \mathcal{E}$  is

$$V(\mathcal{R}_{sm}) = V\left(\frac{1}{\mathcal{E}} \sum_j d_{sj}\right) = \frac{1}{\mathcal{E}^2} \sum_j V(d_{sj}) = \frac{1}{\mathcal{E}^2} \sum_j \mathcal{N}_{sj}^2 V(\hat{\gamma}_{sj})$$

where

$$V(\hat{\gamma}_{sj}) = \frac{\hat{\gamma}_{sj}(1 - \hat{\gamma}_{sj})}{\mathcal{N}_{sj}}$$

By using the method of statistical differentials (Elandt-Johnson and Johnson 1980, pp. 70–71), the variance of the logarithm of  $\mathcal{R}_{sm}$  can be estimated by

$$V(\log(\mathcal{R}_{sm})) = \frac{1}{\mathcal{R}_{sm}^2} V(\mathcal{R}_{sm})$$

For the rate estimate,

$$V(\log(\mathcal{R}_{sm})) = \frac{1}{\mathcal{R}_{sm}^2} V(\mathcal{R}_{sm}) = \frac{1}{\mathcal{R}_{sm}^2} \frac{\mathcal{R}_{sm}}{\mathcal{E}} = \frac{1}{\mathcal{R}_{sm}} \frac{1}{\mathcal{E}} = \frac{1}{\mathcal{D}}$$

The confidence intervals for  $\mathcal{R}_{sm}$  can be constructed based on normal, lognormal, and Poisson distributions.

### Normal Distribution Confidence Interval for SMR

A  $(1 - \alpha)$  confidence interval for  $\mathcal{R}_{sm}$  based on a normal distribution is given by

$$(\mathcal{R}_l, \mathcal{R}_u) = \left( \mathcal{R}_{sm} - z \sqrt{V(\mathcal{R}_{sm})}, \mathcal{R}_{sm} + z \sqrt{V(\mathcal{R}_{sm})} \right)$$

where  $z = \Phi^{-1}(1 - \alpha/2)$  is the  $(1 - \alpha/2)$  quantile of the standard normal distribution.

A test statistic for the null hypothesis  $H_0 : \text{SMR} = 1$  is then given by

$$\frac{\mathcal{R}_{sm} - 1}{\sqrt{V(\mathcal{R}_{sm})}}$$

The test statistic has an approximate standard normal distribution under  $H_0$ .

### Lognormal Distribution Confidence Interval for SMR

A  $(1 - \alpha)$  confidence interval for  $\log(\mathcal{R}_{sm})$  based on a normal distribution is given by

$$\left( \log(\mathcal{R}_{sm}) - z \sqrt{V(\log(\mathcal{R}_{sm}))}, \log(\mathcal{R}_{sm}) + z \sqrt{V(\log(\mathcal{R}_{sm}))} \right)$$

where  $z = \Phi^{-1}(1 - \alpha/2)$  is the  $(1 - \alpha/2)$  quantile of the standard normal distribution.

Thus, a  $(1 - \alpha)$  confidence interval for  $\mathcal{R}_{sm}$  based on a lognormal distribution is given by

$$\left( \mathcal{R}_{sm} e^{-z \sqrt{V(\log(\mathcal{R}_{sm}))}}, \mathcal{R}_{sm} e^{z \sqrt{V(\log(\mathcal{R}_{sm}))}} \right)$$

A test statistic for the null hypothesis  $H_0 : \text{SMR} = 1$  is then given by

$$\frac{\log(\mathcal{R}_{sm})}{\sqrt{V(\log(\mathcal{R}_{sm}))}}$$

The test statistic has an approximate standard normal distribution under  $H_0$ .

### Poisson Distribution Confidence Interval for Rate SMR

Denote the  $(\alpha/2)$  quantile for the  $\chi^2$  distribution with  $2\mathcal{D}$  degrees of freedom by

$$q_l = (\chi_{2\mathcal{D}}^2)^{-1}(\alpha/2)$$

Denote the  $(1 - \alpha/2)$  quantiles for the  $\chi^2$  distribution with  $2(\mathcal{D} + 1)$  degrees of freedom by

$$q_u = (\chi_{2(\mathcal{D}+1)}^2)^{-1}(1 - \alpha/2)$$

Then a  $(1 - \alpha)$  confidence interval for  $\mathcal{R}_{sm}$  based on the  $\chi^2$  distribution is given by

$$(\mathcal{R}_l, \mathcal{R}_u) = \left( \frac{q_l}{2\mathcal{E}}, \frac{q_u}{2\mathcal{E}} \right)$$

A  $p$ -value for the test of the null hypothesis  $H_0 : \text{SMR} = 1$  is given by

$$2 \min \left( \sum_{k=0}^{\mathcal{D}} \frac{e^{-\mathcal{E}} \mathcal{E}^k}{k!}, \sum_{k=\mathcal{D}}^{\infty} \frac{e^{-\mathcal{E}} \mathcal{E}^k}{k!} \right)$$

### Indirectly Standardized Rate and Its Confidence Interval

With a rate-standardized mortality ratio  $\mathcal{R}_{sm}$ , an indirectly standardized rate for the study population is computed as

$$\hat{\lambda}_{is} = \mathcal{R}_{sm} \hat{\lambda}_r$$

where  $\hat{\lambda}_r$  is the overall crude rate in the reference population.

The  $(1 - \alpha/2)$  confidence intervals for  $\hat{\lambda}_{is}$  can be constructed as

$$(\mathcal{R}_l \hat{\lambda}_r, \mathcal{R}_u \hat{\lambda}_r)$$

where  $(\mathcal{R}_l, \mathcal{R}_u)$  is the confidence interval for  $\mathcal{R}_{sm}$ .

### Indirectly Standardized Risk and Its Confidence Interval

With a risk-standardized mortality ratio  $\mathcal{R}_{sm}$ , an indirectly standardized risk for the study population is computed as

$$\hat{\gamma}_{is} = \mathcal{R}_{sm} \hat{\gamma}_r$$

where  $\hat{\gamma}_r$  is the overall crude risk in the reference population.

The  $(1 - \alpha/2)$  confidence intervals for  $\hat{\gamma}_{is}$  can be constructed as

$$(\mathcal{R}_l \hat{\gamma}_r, \mathcal{R}_u \hat{\gamma}_r)$$

where  $(\mathcal{R}_l, \mathcal{R}_u)$  is the confidence interval for  $\mathcal{R}_{sm}$ .

---

### Attributable Fraction and Population Attributable Fraction

The attributable fraction measures the excess event rate or risk fraction in the exposed population that is attributable to the exposure. That is, it is the proportion of event rate or risk in the exposed population that would be reduced if the exposure were not present. In contrast, the population attributable fraction measures the excess event rate or risk fraction in the total population that is attributable to the exposure.

In the STDRATE procedure, you can compute the attributable fraction by using either indirect standardization or Mantel-Haenszel estimation.

### Indirect Standardization

With indirect standardization, you specify a study population that consists of subjects who are exposed to a factor, such as smoking, and a reference population that consists of subjects who are not exposed to the factor. Denote the numbers of events in the study and reference populations by  $\mathcal{D}_s$  and  $\mathcal{D}_r$ , respectively.

For the rate estimate, denote the population-times in the study and reference populations by  $\mathcal{T}_s$  and  $\mathcal{T}_r$ , respectively. Then the event rates in the two populations can be expressed as the following equations, respectively:

$$\hat{\lambda}_s = \frac{\mathcal{D}_s}{\mathcal{T}_s} \quad \text{and} \quad \hat{\lambda}_r = \frac{\mathcal{D}_r}{\mathcal{T}_r}$$

Similarly, for the risk estimate, denote the numbers of observations in the study and reference populations by  $\mathcal{N}_s$  and  $\mathcal{N}_r$ , respectively. Then the event risks in the two populations can be expressed as the following equations, respectively:

$$\hat{\gamma}_s = \frac{\mathcal{D}_s}{\mathcal{N}_s} \quad \text{and} \quad \hat{\gamma}_r = \frac{\mathcal{D}_r}{\mathcal{N}_r}$$

In the next two subsections,  $\beta = \lambda$  denotes the rate statistic and  $\beta = \gamma$  denotes the risk statistic.

### Attributable Fraction with Indirect Standardization

The attributable fraction is the fraction of event rate or risk in the exposed population that is attributable to exposure:

$$\mathcal{R}_a = \frac{\hat{\beta}_s - \hat{\beta}_r}{\hat{\beta}_s}$$

With a standardized mortality ratio  $\mathcal{R}_{sm}$ , the attributable fraction is estimated by

$$\mathcal{R}_a = \frac{\mathcal{R}_{sm} - 1}{\mathcal{R}_{sm}}$$

The confidence intervals for the attributable fraction can be computed using the confidence intervals for  $\mathcal{R}_{sm}$ . That is, with a confidence interval  $(\mathcal{R}_l, \mathcal{R}_u)$  for  $\mathcal{R}_{sm}$ , the corresponding  $\mathcal{R}_a$  confidence interval is given by

$$\left( \frac{\mathcal{R}_l - 1}{\mathcal{R}_l}, \frac{\mathcal{R}_u - 1}{\mathcal{R}_u} \right)$$

### Population Attributable Fraction with Indirect Standardization

The population attributable fraction for a population is the fraction of event rate or risk in a given time period that is attributable to exposure. The population attributable fraction is

$$\mathcal{R}_{pa} = \frac{\hat{\beta}_0 - \hat{\beta}_r}{\hat{\beta}_0}$$

where

$$\hat{\beta}_0 = \frac{\mathcal{D}_s + \mathcal{D}_r}{\mathcal{T}_s + \mathcal{T}_r}$$

is the combined rate in the total population for the rate statistic and where

$$\hat{\beta}_0 = \frac{\mathcal{D}_s + \mathcal{D}_r}{\mathcal{N}_s + \mathcal{N}_r}$$

is the combined risk in the total population for the risk statistic.

Denote  $\rho = \mathcal{D}_s / (\mathcal{D}_s + \mathcal{D}_r)$ , the proportion of exposure among events, then  $\mathcal{R}_{pa}$  can also be expressed as

$$\mathcal{R}_{pa} = \rho \frac{\mathcal{R}_{sm} - 1}{\mathcal{R}_{sm}}$$

where  $\mathcal{R}_{sm}$  is the standardized mortality ratio.

An approximate confidence interval for the population attributable rate  $\mathcal{R}_{pa}$  can be derived by using the complementary log transformation (Greenland 2008, p. 296). That is, with

$$\mathcal{H} = \log(1 - \mathcal{R}_{pa})$$

a variance estimator for the estimated  $\mathcal{H}$  is given by

$$\text{Var}(\hat{\mathcal{H}}) = \frac{\mathcal{R}_{pa}^2}{(1 - \mathcal{R}_{pa})^2} \left( \frac{\hat{V}}{(\mathcal{R}_{sm} - 1)^2} + \frac{2}{\mathcal{D}_s (\mathcal{R}_{sm} - 1)} + \frac{\mathcal{D}_r}{\mathcal{D}_s (\mathcal{D}_s + \mathcal{D}_r)} \right)$$

where  $\hat{V}$  is a variance estimate for  $\log(\mathcal{R}_{sm})$ .

### Mantel-Haenszel Estimation

With Mantel-Haenszel estimation, you specify one study population that consists of subjects who are exposed to a factor and another study population that consists of subjects who are not exposed to the factor. Denote the numbers of events in the exposed and nonexposed study populations by  $\mathcal{D}_1$  and  $\mathcal{D}_2$ , respectively.

For the rate estimate, denote the population-times in the two populations by  $\mathcal{T}_1$  and  $\mathcal{T}_2$ , respectively. Then the event rates in the two populations can be expressed as the following equations, respectively:

$$\hat{\lambda}_1 = \frac{\mathcal{D}_1}{\mathcal{T}_1} \quad \text{and} \quad \hat{\lambda}_2 = \frac{\mathcal{D}_2}{\mathcal{T}_2}$$

Similarly, for the risk estimate, denote the numbers of observations in the two populations by  $\mathcal{N}_1$  and  $\mathcal{N}_2$ , respectively. Then the event risks in the two populations can be expressed as the following equations, respectively:

$$\hat{\gamma}_1 = \frac{\mathcal{D}_1}{\mathcal{N}_1} \quad \text{and} \quad \hat{\gamma}_2 = \frac{\mathcal{D}_2}{\mathcal{N}_2}$$

In the next two subsections,  $\beta = \lambda$  denotes the rate statistic and  $\beta = \gamma$  denotes the risk statistic.

### Attributable Fraction with Mantel-Haenszel Estimation

The attributable fraction is the fraction of event rate or risk in the exposed population that is attributable to exposure:

$$\mathcal{R}_a = \frac{\hat{\beta}_1 - \hat{\beta}_2}{\hat{\beta}_1}$$

Denote the rate or risk ratio by  $\mathcal{R} = \hat{\beta}_1 / \hat{\beta}_2$ . The attributable fraction is given by

$$\mathcal{R}_a = \frac{\mathcal{R} - 1}{\mathcal{R}}$$

The confidence intervals for the attributable fraction can be computed using the confidence intervals for the rate or risk ratio  $\mathcal{R}$ . That is, with a confidence interval  $(\mathcal{R}_l, \mathcal{R}_u)$  for  $\mathcal{R}$ , the corresponding  $\mathcal{R}_a$  confidence interval is given by

$$\left( \frac{\mathcal{R}_l - 1}{\mathcal{R}_l}, \frac{\mathcal{R}_u - 1}{\mathcal{R}_u} \right)$$

For Mantel-Haenszel estimation, you can use the Mantel-Haenszel rate or risk ratio to estimate  $\mathcal{R}$ .

### Population Attributable Fraction with Mantel-Haenszel Estimation

The population attributable fraction for a population is the fraction of event rate or risk in a given time period that is attributable to exposure. The population attributable fraction is

$$\mathcal{R}_{pa} = \frac{\hat{\beta}_0 - \hat{\beta}_2}{\hat{\beta}_0}$$

where

$$\hat{\beta}_0 = \frac{\mathcal{D}_1 + \mathcal{D}_2}{\mathcal{T}_1 + \mathcal{T}_2}$$

is the combined rate in the total population for the rate statistic and where

$$\hat{\beta}_0 = \frac{\mathcal{D}_1 + \mathcal{D}_2}{\mathcal{N}_1 + \mathcal{N}_2}$$

is the combined risk in the total population for the risk statistic.

Denote the proportion of exposure among events as  $\rho = \mathcal{D}_1 / (\mathcal{D}_1 + \mathcal{D}_2)$ . Then  $\mathcal{R}_{pa}$  can also be expressed as

$$\mathcal{R}_{pa} = \rho \frac{\mathcal{R} - 1}{\mathcal{R}}$$

where  $\mathcal{R} = \hat{\beta}_1 / \hat{\beta}_2$  is the rate or risk ratio.

An approximate confidence interval for the population attributable rate  $\mathcal{R}_{pa}$  can be derived by using the complementary log transformation (Greenland 2008, p. 296). That is, with

$$\mathcal{H} = \log(1 - \mathcal{R}_{pa})$$

a variance estimator for the estimated  $\mathcal{H}$  is given by

$$\text{Var}(\hat{\mathcal{H}}) = \frac{\mathcal{R}_{pa}^2}{(1 - \mathcal{R}_{pa})^2} \left( \frac{\hat{V}}{(\mathcal{R} - 1)^2} + \frac{2}{\mathcal{D}_1 (\mathcal{R} - 1)} + \frac{\mathcal{D}_2}{\mathcal{D}_1 (\mathcal{D}_1 + \mathcal{D}_2)} \right)$$

where  $\hat{V}$  is a variance estimate for  $\log(\mathcal{R})$ .

For Mantel-Haenszel estimation, you can use the Mantel-Haenszel rate or risk ratio to estimate  $\mathcal{R}$ .

## Applicable Data Sets and Required Variables for Method Specifications

The METHOD= and DATA= options are required in the STD RATE procedure. The METHOD= option specifies the standardization method, and the DATA= and REFDATA= options specify the study populations and reference population, respectively. You can use the GROUP= option in the POPULATION statement to identify various study populations. Table 88.2 lists applicable data sets for each method.

**Table 88.2** Applicable Data Sets for Method Specifications

METHOD=	Number of Populations	
	in DATA= Data Set	REFDATA= Data Set
DIRECT	1	X
	2	X
MH	2	
INDIRECT	1	X



Table 88.3 lists the required variables for each method.

**Table 88.3** Required Variables for Method Specifications

METHOD=	STAT=	DATA= Data Set			REFDATA= Data Set		
		RATE	RISK	TOTAL	RATE	RISK	TOTAL
DIRECT	RATE	X					X
	RISK		X				X
MH	RATE	X		X			
	RISK		X	X			
INDIRECT	RATE			X	X		
	RISK			X		X	

The symbol “X” indicates that the variable is either explicitly specified or implicitly available from other variables. For example, when STAT=RATE, the variable RATE is available if the corresponding variables EVENT and TOTAL are specified.

## Applicable Confidence Limits for Rate and Risk Statistics

In the STD RATE procedure, the METHOD= option specifies the standardization method, and the STAT= option specifies either either rate or risk for standardization. Table 88.4 lists applicable confidence limits for different methods with standardized rate, rate SMR, standardized risk, and risk SMR.

**Table 88.4** Applicable Confidence Limits for Standardized Rate and Risk Statistics

Statistic	METHOD=	Confidence Limits			
		Normal	Lognormal	Gamma	Poisson
Rate	DIRECT	X	X	X	
	MH	X	X	X	
	INDIRECT	X	X		X
Rate SMR	INDIRECT	X	X		X
Risk	DIRECT	X	X		
	MH	X	X		
	INDIRECT	X	X		X
Risk SMR	INDIRECT	X	X		X

Table 88.5 lists applicable confidence limits for stratum-specific rate, rate SMR, risk, and risk SMR.

**Table 88.5** Applicable Confidence Limits for Strata Rate and Risk Statistics

Statistic	Confidence Limits		
	Normal	Lognormal	Poisson
Rate	X	X	X
Rate SMR	X	X	X
Risk	X	X	
Risk SMR	X	X	

## Table Output

The STD RATE procedure displays the “Standardization Information” table by default. In addition, the procedure also displays the “Standardized Rate Estimates” table (with the default STAT=RATE option in the PROC STD RATE statement) and the “Standardized Risk Estimates” table (with the STAT=RISK option) by default. The rest of this section describes the output tables in alphabetical order.

## Attributable Fraction Estimates

The “Attributable Fraction Estimates” table displays the following information:

- Parameter: attributable rate and population attributable rate for the rate statistic, and attributable risk and population attributable risk for the risk statistic
- Estimate: estimate of the parameter
- Method: method to construct confidence limits
- Lower and Upper: lower and upper confidence limits

## Effect Estimates

The “Effect Estimates” table displays the following information:

- Standardized Rate: directly standardized rates for study populations
- Standardized Risk: directly standardized risks for study populations

When EFFECT=RATIO, the table displays the following:

- Estimate: the rate or risk ratio estimate
- Log Ratio: the logarithm of rate ratio or risk ratio estimate

- Standard Error: standard error of the logarithm of the ratio estimate
- Z: the standard Z statistic
- $\Pr > |Z|$ : the  $p$ -value for the test

When EFFECT=DIFF, the table displays the following:

- Estimate: the rate or risk difference estimate
- Standard Error: standard error of the difference estimate
- Z: the standard Z statistic
- $\Pr > |Z|$ : the  $p$ -value for the test

### Standardization Information

The “Standardization Information” table displays the input data sets, type of statistic to be standardized, standardization method, and number of strata. The table also displays the variance divisor for the risk estimate, and the rate multiplier for the rate estimate. With a rate multiplier  $c$ , the rates per  $c$  population-time units are displayed in the output tables.

### Standardized Morbidity/Mortality Ratio

The “Standardized Morbidity/Mortality Ratio” table displays the following information:

- SMR: standardized morbidity/mortality ratio
- Standard Error: standard error for SMR
- Lower and Upper: lower and upper confidence limits for SMR
- Test Statistic:  $\text{SMR}-1$ , for the test of  $\text{SMR}=1$
- Estimate: value of test statistic
- Standard Error: standard error of the estimate
- Z: the standard Z statistic
- $\Pr > |Z|$ : the  $p$ -value for the test

## Standardized Rate Estimates

The “Standardized Rate Estimates” table displays the following information:

- Population: study populations, and reference population for indirect standardization
- Number of Events: number of events in population
- Population-Time: total contributed time in population, for the rate statistic
- Crude Rate: event rate in the population
- Expected Number of Events
- SMR: standardized morbidity/mortality ratio, for indirect standardization
- Standardized Rate: for the rate statistic
- Standard Error: standard error of the standardized estimate of rate
- Confidence Limits: lower and upper confidence limits for standardized estimate

## Standardized Risk Estimates

The “Standardized Risk Estimates” table displays the following information:

- Population: study populations, and reference population for indirect standardization
- Number of Events: number of events in population
- Number of Observations: number of observations in population
- Crude Risk: event risk in the population
- Expected Number of Events
- SMR: standardized morbidity/mortality ratio, for indirect standardization
- Standardized Risk
- Standard Error: standard error of the standardized estimate of risk
- Confidence Limits: lower and upper confidence limits for standardized estimate

## Strata Effect Estimates

The “Strata Effect Estimates” table displays the following information for each stratum:

- Stratum Index: a sequential stratum identification number
- STRATA variables: the levels of STRATA variables
- Rate: rates for the study populations, for the rate statistic
- Risk: risks for the study populations, for the risk statistic

When EFFECT=DIFF, the table displays the following information for each stratum:

- Estimate: rate or risk difference estimate of the study populations
- Standard Error: the standard error of the difference estimate
- Confidence Limits: confidence limits for the difference estimate

When EFFECT=RATIO, the table displays the following information for each stratum:

- Estimate: rate or risk ratio estimate of the study populations
- Confidence Limits: confidence limits for the ratio estimate

## Strata Statistics

For each POPULATION statement, the “Strata Information” table displays the following information for each stratum:

- Stratum Index: a sequential stratum identification number
- STRATA variables: the levels of STRATA variables

If the REFERENCE statement is specified, the table displays the following information for each stratum in the reference population:

- Population-Time Value: population-time for the rate statistic for direct standardization
- Population-Time Proportion: proportion for the population-time
- Number of Observations Value: number of observations for the risk statistic for direct standardization
- Number of Observations Proportion: proportion for the number of observations
- Rate: for the rate statistic for indirect standardization
- Risk: for the risk statistic for indirect standardization

For the rate statistic, the table displays the following information for each stratum in the specified study data set:

- Number of Events
- Population-Time Value
- Population-Time Proportion
- Rate Estimate
- Standard Error: standard error for the rate estimate if the CL=NORMAL suboption is specified in the STATS option in the STRATA statement
- Confidence Limits: confidence limits for the risk estimate if the CL suboption is specified in the STATS option in the STRATA statement
- Expected Number of Events: expected number of events that use the reference population population-time for direct standardization, Mantel-Haenszel weight for Mantel-Haenszel estimation, or reference population rate for indirect standardization

For the risk statistic, the table displays the following information for each stratum in the specified study data set:

- Number of Events
- Number of Observations Value
- Number of Observations Proportion
- Risk
- Standard Error: standard error for the risk estimate, if the CL=NORMAL suboption is specified in the STATS option in the STRATA statement
- Confidence Limits: confidence limits for the risk estimate, if the CL suboption is specified in the STATS option in the STRATA statement
- Expected Number of Events: expected number of events that uses the reference population number of observations for direct standardization, Mantel-Haenszel weight for Mantel-Haenszel estimation, or reference population risk for indirect standardization

### Strata SMR Estimates

The “Strata SMR Estimates” table displays the following information for each stratum:

- Stratum Index: a sequential stratum identification number
- STRATA variables: the levels of STRATA variables
- Number of Events

- Expected Number of Observations
- SMR Estimate
- Standard Error: standard error for the SMR estimate, if the CL=NORMAL suboption is specified in the SMR option in the STRATA statement
- Confidence Limits: confidence limits for the SMR estimate, if the CL suboption is specified in the SMR option in the STRATA statement

For the rate statistic, the table also displays the following information for each stratum:

- Population-Time
- Reference Rate

For the risk statistic, the table also displays the following information for each stratum:

- Number of Observations
- Reference Risk

## ODS Table Names

PROC STD RATE assigns a name to each table it creates. You must use these names to refer to tables when you use the Output Delivery System (ODS). These names are listed in [Table 88.6](#). For more information about ODS, see Chapter 20, “Using the Output Delivery System.”

**Table 88.6** ODS Tables Produced by PROC STD RATE

ODS Table Name	Description	Statement	Option
AttrFraction	Attributable fraction	PROC STD RATE	METHOD=INDIRECT(AF)
		PROC STD RATE	METHOD=MH(AF)
Effect	Effect estimates	PROC STD RATE	EFFECT
SMR	Standardized morbidity/mortality ratio	PROC STD RATE	METHOD=INDIRECT
StdInfo	Standardization information	PROC STD RATE	
StdRate	Standardized rate estimates	PROC STD RATE	STAT=RATE
StdRisk	Standardized risk estimates	PROC STD RATE	STAT=RISK
StrataEffect	Strata effect estimates	STRATA	EFFECT
StrataStats	Strata statistics	STRATA	STATS
StrataSMR	Strata SMR estimates	STRATA	SMR

---

## Graphics Output

This section describes the use of ODS for creating graphics with the STD RATE procedure. To request these graphs, ODS Graphics must be enabled and you must specify the associated graphics options in the PROC STD RATE statement. For more information about ODS Graphics, see Chapter 21, “Statistical Graphics Using ODS.”

### Strata Distribution Plot

The PLOTS=DIST option displays the proportion of exposed time or sample size for each stratum in the populations.

### Strata Effect Plot

The PLOTS=EFFECT option displays the stratum-specific effect measure of rate difference, rate ratio, risk difference, or risk ratio. In addition, the crude effect measure and confidence limits of these stratum-specific effect estimates are also displayed.

### Strata Rate Plot

The PLOTS=RATE option displays the stratum-specific rate estimates and their confidence limits of populations. In addition, the overall crude rates of populations are also displayed.

### Strata Risk Plot

The PLOTS=RISK option displays the stratum-specific risk estimates and their confidence limits of populations. In addition, the overall crude risks of populations are also displayed.

### Strata SMR Plot

The PLOTS=SMR option displays the SMR for each stratum in the populations.

---

## ODS Graphics

Statistical procedures use ODS Graphics to create graphs as part of their output. ODS Graphics is described in detail in Chapter 21, “Statistical Graphics Using ODS.”

Before you create graphs, ODS Graphics must be enabled (for example, by specifying the ODS GRAPHICS ON statement). For more information about enabling and disabling ODS Graphics, see the section “Enabling and Disabling ODS Graphics” on page 600 in Chapter 21, “Statistical Graphics Using ODS.”

The overall appearance of graphs is controlled by ODS styles. Styles and other aspects of using ODS Graphics are discussed in the section “A Primer on ODS Statistical Graphics” on page 599 in Chapter 21, “Statistical Graphics Using ODS.”

PROC STD RATE assigns a name to each graph it creates. You can use these names to refer to the graphs when you use ODS. To request the graph, ODS Graphics must be enabled and you must use the PLOTS option in the PROC STD RATE statement to specify the *plot-request* indicated in Table 88.7.



**Table 88.7** Graphs Produced by PROC STD RATE

ODS Graph Name	Plot Description	<i>plot-request</i> in PLOTS Option
StrataDistPlot	Strata proportion of exposed time or sample size	DIST
StrataEffectPlot	Strata effect measure of rate difference, rate ratio, risk difference, or risk ratio	EFFECT
StrataRatePlot	Strata rate estimates	RATE
StrataRiskPlot	Strata risk estimates	RISK
StrataSMRPlot	Strata SMR of rates or risks	SMR

## Examples: STD RATE Procedure

### Example 88.1: Comparing Directly Standardized Rates

This example computes directly standardized mortality rates for populations in the states of Alaska and Florida, and then compares these two standardized rates with a rate ratio statistic.

The Alaska data set contains the stratum-specific mortality information in a given period of time for the state of Alaska (Alaska Bureau of Vital Statistics 2000a, b). Variables Sex and Age are the grouping variables that form the strata in the standardization, and variables Death and PYear indicate the number of events and person-years, respectively. The COMMA9. format is specified in the DATA step to input numerical values that contain commas in PYear.

```
data Alaska;
  State='Alaska';
  input Sex $ Age $ Death PYear comma9.;
  datalines;
Male    00-14    37    81,205
Male    15-34    68    93,662
Male    35-54   206   108,615
Male    55-74   369    35,139
Male    75+     556     5,491
Female  00-14    78    77,203
Female  15-34   181    85,412
Female  35-54   395   100,386
Female  55-74   555    32,118
Female  75+     479     7,701
;
```

The Florida data set contains the corresponding stratum-specific mortality information for the state of Florida (Florida Department of Health 2000, 2012). Variables Sex and Age are the grouping variables that form the strata in the standardization, and variables Death and PYear indicate the number of events and person-years, respectively.

```

data Florida;
  State='Florida';
  input Sex $ Age $ Death comma8. PYear comma11.;
  datalines;
Male    00-14    1,189  1,505,889
Male    15-34    2,962  1,972,157
Male    35-54   10,279  2,197,912
Male    55-74   26,354  1,383,533
Male    75+     42,443   554,632
Female  00-14     906  1,445,831
Female  15-34    1,234  1,870,430
Female  35-54    5,630  2,246,737
Female  55-74   18,309  1,612,270
Female  75+     53,489   868,838
;

```

The TwoStates data set contains the data sets Alaska and Florida:

```

data TwoStates;
  length State $ 7.;
  set Alaska Florida;
run;

```

The US data set contains the corresponding stratum-specific person-years information for the United States (U.S. Bureau of Census 2011). Variables Sex and Age are the grouping variables that form the strata in the standardization, and the variable PYear indicates the person-years.

```

data US;
  input Sex $ Age $ PYear comma12.;
  datalines;
Male    00-14   30,854,207
Male    15-34   40,199,647
Male    35-54   40,945,028
Male    55-74   19,948,630
Male    75+     6,106,351
Female  00-14   29,399,168
Female  15-34   38,876,268
Female  35-54   41,881,451
Female  55-74   22,717,040
Female  75+    10,494,416
;

```

The following statements invoke the STD RATE procedure and compute the direct standardized rates for the states of Florida and Alaska by using the United States as the reference population. The DATA= option names the data set for the study populations, and the REFDATA= option names the data set for the reference population.

```

ods graphics on;
proc stdrate data=TwoStates
  refdata=US
  method=direct
  stat=rate(mult=1000)
  effect

```

```

      plots(only)=(dist effect)
    ;
    population group=State event=Death total=PYear;
    reference total=PYear;
    strata Sex Age / effect;
run;
ods graphics off;

```

The METHOD=DIRECT option requests direct standardization, and the STAT=RATE option specifies the rate statistic for standardization. With the EFFECT option, the procedure computes the rate effect between the study populations with the default rate ratio statistics.

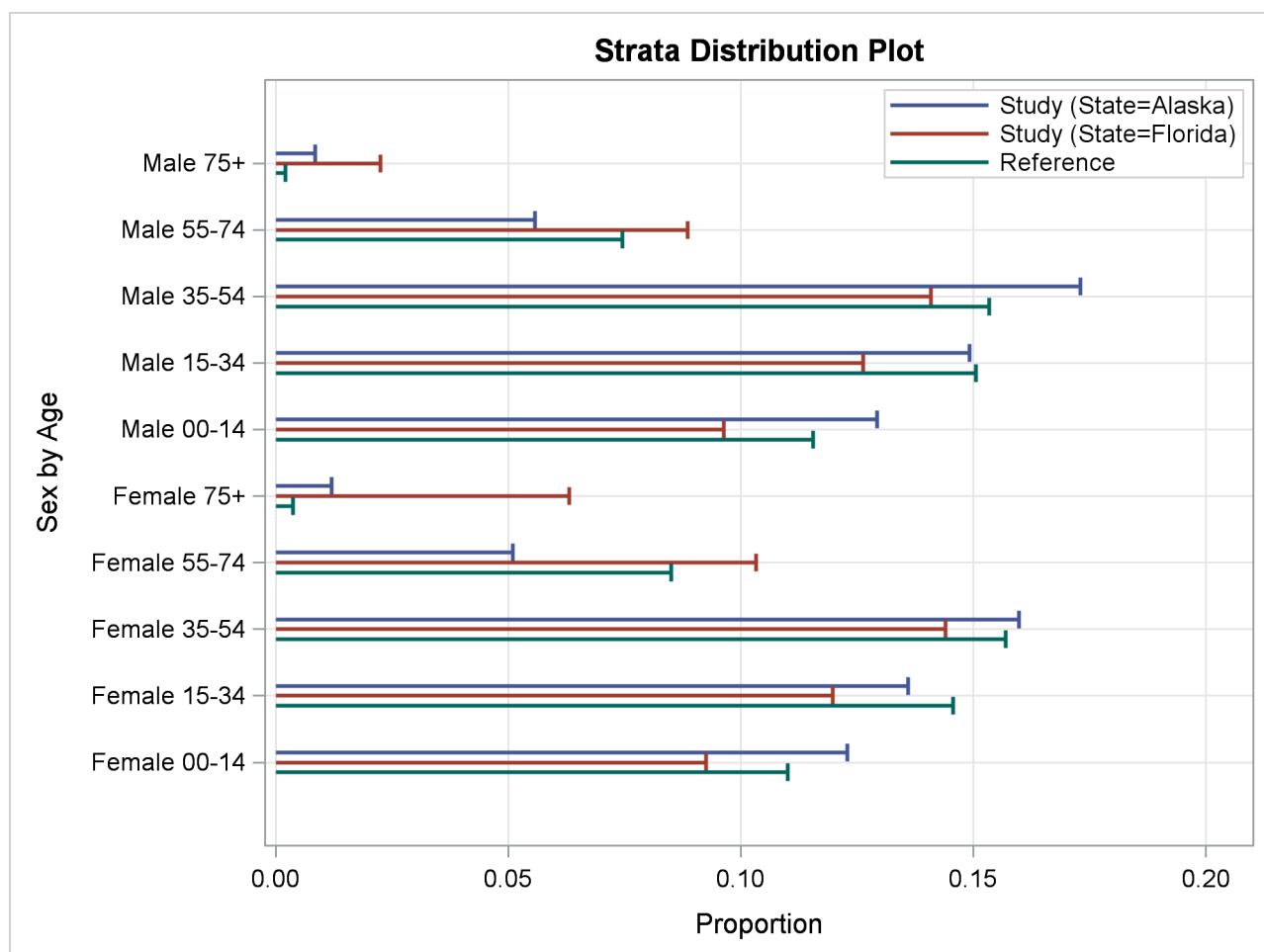
The “Standardization Information” table in [Output 88.1.1](#) displays the standardization information.

**Output 88.1.1** Standardization Information

The STD RATE Procedure	
Standardization Information	
Data Set	WORK.TWOSTATES
Group Variable	State
Reference Data Set	WORK.US
Method	Direct Standardization
Statistic	Rate
Number of Strata	10
Rate Multiplier	1000

With ODS Graphics enabled, the PLOTS(ONLY)=(DIST EFFECT) option displays the strata distribution plot and the strata effect plot, but does not display the default strata rate plot.

The strata distribution plot displays proportions for stratum-specific person-years in the study populations and reference population, as shown in [Output 88.1.2](#).

**Output 88.1.2** Strata Distribution Plot

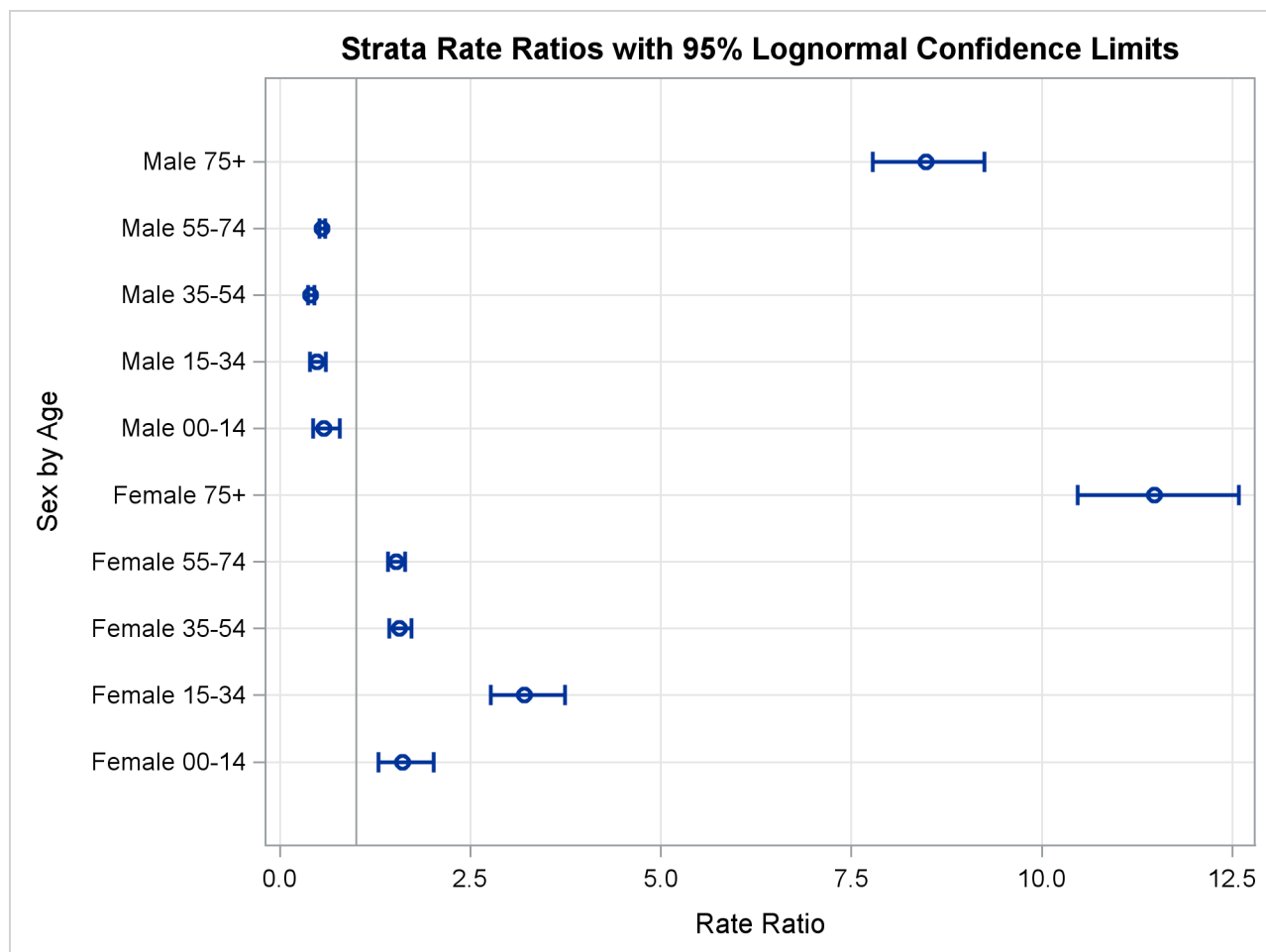
The EFFECT option in the STRATA statement and the STAT=RATE option request that the “Strata Rate Effect Estimates” table in [Output 88.1.3](#) display the stratum-specific rate effect statistics between the two study populations. The default EFFECT=RATIO in the PROC STD RATE statement requests that the stratum-specific rate ratio statistics be displayed.

**Output 88.1.3** Strata Effect Estimates

Strata Rate Effect Estimates (Rate Multiplier = 1000)							
Stratum Index	Sex	Age	-----State----- Alaska Florida		Rate Ratio	95% Lognormal Confidence Limits	
1	Female	00-14	1.010	0.6266	1.6123	1.2794	2.0319
2	Female	15-34	2.119	0.6597	3.2121	2.7481	3.7544
3	Female	35-54	3.935	2.5059	1.5702	1.4180	1.7389
4	Female	55-74	17.280	11.3560	1.5217	1.3984	1.6557
5	Female	75+	62.200	5.4191	11.4779	10.4536	12.6026
6	Male	00-14	0.456	0.7896	0.5771	0.4160	0.8004
7	Male	15-34	0.726	1.5019	0.4834	0.3801	0.6148
8	Male	35-54	1.897	4.6767	0.4055	0.3533	0.4655
9	Male	55-74	10.501	19.0483	0.5513	0.4975	0.6109
10	Male	75+	101.257	11.9394	8.4809	7.7634	9.2647

The “Strata Rate Effect Estimates” table shows that except for the age group 75+, Alaska has lower mortality rates for male groups and higher mortality rates for female groups than Florida. For age group 75+, Alaska has much higher mortality rates than Florida for both male and female groups.

With ODS Graphics enabled and two study populations, the PLOTS=EFFECT option displays the stratum-specific effect measures and their associated confidence limits, as shown in [Output 88.1.4](#). The STAT=RATE option and the default EFFECT=RATIO option request that the strata rate ratios be displayed. By default, confidence limits are generated with 95% confidence level. This plot displays the stratum-specific rate ratios in the “Strata Rate Effect Estimates” table in [Output 88.1.3](#).

**Output 88.1.4** Strata Effect Measure Plot

The “Directly Standardized Rate Estimates” table in [Output 88.1.5](#) displays directly standardized rates and related statistics.

**Output 88.1.5** Directly Standardized Rate Estimates

Directly Standardized Rate Estimates					
Rate Multiplier = 1000					
	-----Study Population-----			-Reference Population-	
State	Observed Events	Population- Time	Crude Rate	Expected Events	Population- Time
Alaska	2924	626932	4.6640	1126924	266481515
Florida	76455	15577105	4.9082	1076187	266481515
Directly Standardized Rate Estimates					
Rate Multiplier = 1000					
	-----Standardized Rate-----				
		Standard	95% Normal		
State	Estimate	Error	Confidence	Limits	
Alaska	4.2289	0.0901	4.0522	4.4056	
Florida	4.0385	0.0156	4.0079	4.0691	

The MULT=1000 suboption in the STAT=RATE option requests that rates per 1,000 person-years be displayed. The table shows that although the crude rate in the Florida population (4.908) is higher than the crude rate in the Alaska population (4.664), the resulting standardized rate in the Florida population (4.0385) is lower than the crude rate in the Alaska population (4.2289).

The EFFECT option requests that the “Rate Effect Estimates” table in [Output 88.1.6](#) display the log rate ratio statistics of the two directly standardized rates by default.

**Output 88.1.6** Effect Estimates

Rate Effect Estimates (Rate Multiplier = 1000)						
-----State-----		Rate Ratio	Log Rate Ratio	Standard Error	Z	Pr >  Z
Alaska	Florida					
4.2289	4.0385	1.0471	0.0461	0.0217	2.13	0.0335

The table shows that with a log rate ratio statistic 1.047, the resulting *p*-value is 0.0335, which indicates that the death rate is significantly higher in Alaska than in Florida at the 5% significance level.

**Example 88.2: Computing Mantel-Haenszel Risk Estimation**

This example uses Mantel-Haenszel method to estimate the effect of household smoking on respiratory symptoms of school children, after adjusting for the effects of the student’s grade and household pets.

Suppose that the School data set contains the stratum-specific numbers of cases of respiratory symptoms in a given school year for a school district. Variables Pet and Grade are the grouping variables that form

the strata in the standardization, and the variable Smoking identifies students who have smokers in their households. The variables Case and Student indicate the number of cases with respiratory symptoms and the total number of students, respectively.

```
data School;
  input Smoking $ Pet $ Grade $ Case Student;
  datalines;
Yes   Yes   K-1   109   807
Yes   Yes   2-3   106   791
Yes   Yes   4-5   112   868
Yes   No    K-1   168   1329
Yes   No    2-3   162   1337
Yes   No    4-5   183   1594
No    Yes   K-1   284   2403
No    Yes   2-3   266   2237
No    Yes   4-5   273   2279
No    No    K-1   414   3398
No    No    2-3   372   3251
No    No    4-5   382   3270
;
```

The following statements invoke the STD RATE procedure and compute the Mantel-Haenszel rate difference statistic between students with household smokers and students without household smokers:

```
ods graphics on;
proc stdrate data=School
  method=mh
  stat=risk
  effect=diff
  plots=all
;
  population group=Smoking event=Case total=Student;
  strata Pet Grade / order=data stats(cl=none) effect;
run;
ods graphics off;
```

The ORDER=DATA option in the STRATA statement sorts the strata by the order of appearance in the input data set.

The “Standardization Information” table in [Output 88.2.1](#) displays the standardization information.



**Output 88.2.1** Standardization Information

The STDRATE Procedure	
Standardization Information	
Data Set	WORK.SCHOOL
Group Variable	Smoking
Method	Mantel-Haenszel
Statistic	Risk
Number of Strata	6

The STATS option in the STRATA statement requests that the STDRATE procedure display a “Mantel-Haenszel Standardized Strata Statistics” table for study populations, as shown in [Output 88.2.2](#). The table displays the strata information and the expected number of events in each stratum. The Expected Events column shows the expected number of events when the Mantel-Haenszel weights are applied to the corresponding stratum-specific risks in the study populations. The CL=NONE suboption requests that confidence limits for strata risks not be displayed.

**Output 88.2.2** Mantel-Haenszel Standardized Strata Statistics

The STDRATE Procedure				
Mantel-Haenszel Standardized Strata Statistics				
Smoking	Stratum Index	Pet	Grade	
No	1	Yes	K-1	
No	2	Yes	2-3	
No	3	Yes	4-5	
No	4	No	K-1	
No	5	No	2-3	
No	6	No	4-5	
Yes	1	Yes	K-1	
Yes	2	Yes	2-3	
Yes	3	Yes	4-5	
Yes	4	No	K-1	
Yes	5	No	2-3	
Yes	6	No	4-5	

Mantel-Haenszel Standardized Strata Statistics					
Smoking	Stratum Index	-----Study Population-----			Crude Risk
		Observed Events	-Number of Observations- Value Proportion		
No	1	284	2403	0.1427	0.118186
No	2	266	2237	0.1329	0.118909
No	3	273	2279	0.1353	0.119789
No	4	414	3398	0.2018	0.121836
No	5	372	3251	0.1931	0.114426
No	6	382	3270	0.1942	0.116820
Yes	1	109	807	0.1200	0.135068
Yes	2	106	791	0.1176	0.134008
Yes	3	112	868	0.1291	0.129032
Yes	4	168	1329	0.1976	0.126411
Yes	5	162	1337	0.1988	0.121167
Yes	6	183	1594	0.2370	0.114806

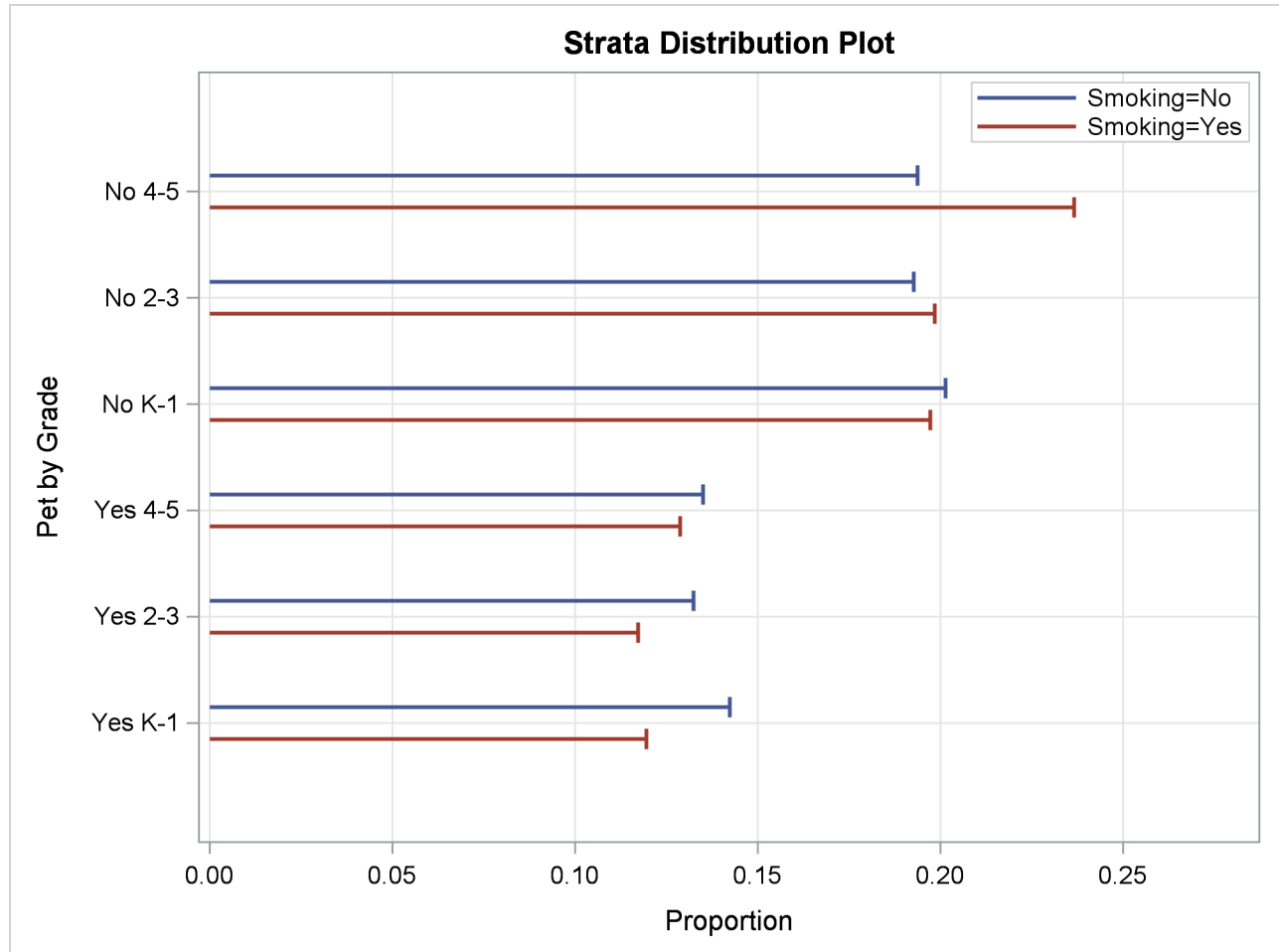
  

Mantel-Haenszel Standardized Strata Statistics				
Smoking	Stratum Index	--Mantel-Haenszel--		Expected Events
		Weight		
No	1	604.12	71.398	
No	2	584.37	69.487	
No	3	628.59	75.298	
No	4	955.35	116.396	
No	5	947.38	108.405	
No	6	1071.62	125.187	
Yes	1	604.12	81.597	
Yes	2	584.37	78.310	
Yes	3	628.59	81.108	
Yes	4	955.35	120.767	
Yes	5	947.38	114.791	
Yes	6	1071.62	123.028	

With ODS Graphics enabled, the PLOTS=ALL option displays all appropriate plots. With the METHOD=MH and STAT=RISK options, these plots include the strata distribution plot, strata risk plot, and strata effect plot.

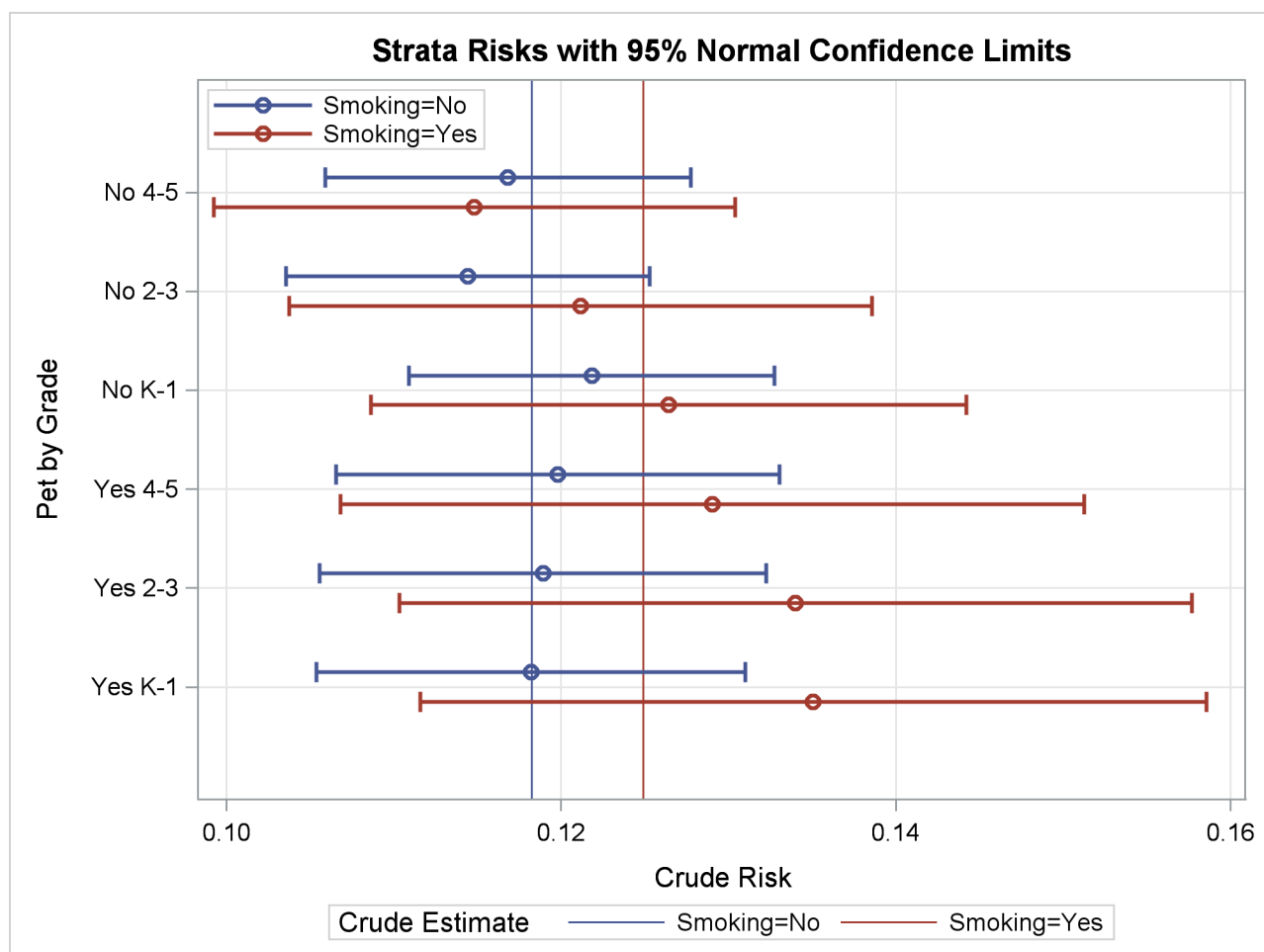
The strata distribution plot displays proportions for stratum-specific numbers of students in the study populations, as shown in [Output 88.2.3](#).

**Output 88.2.3** Strata Distribution Plot



The strata risk plot displays stratum-specific risk estimates with confidence limits in the study populations, as shown in [Output 88.2.4](#). This plot displays stratum-specific risk estimates in the “Mantel-Haenszel Standardized Strata Statistics” table in [Output 88.2.2](#). In addition, the overall crude risks for the two study populations are also displayed. By default, strata levels are displayed on the vertical axis.

Output 88.2.4 Strata Risk Plot



The EFFECT option in the STRATA statement requests that the “Strata Risk Effect Estimates” table be displayed, as shown in [Output 88.2.5](#). The EFFECT=DIFF option in the PROC STDRA statement requests that strata risk differences be displayed.

**Output 88.2.5** Strata Effect Estimates

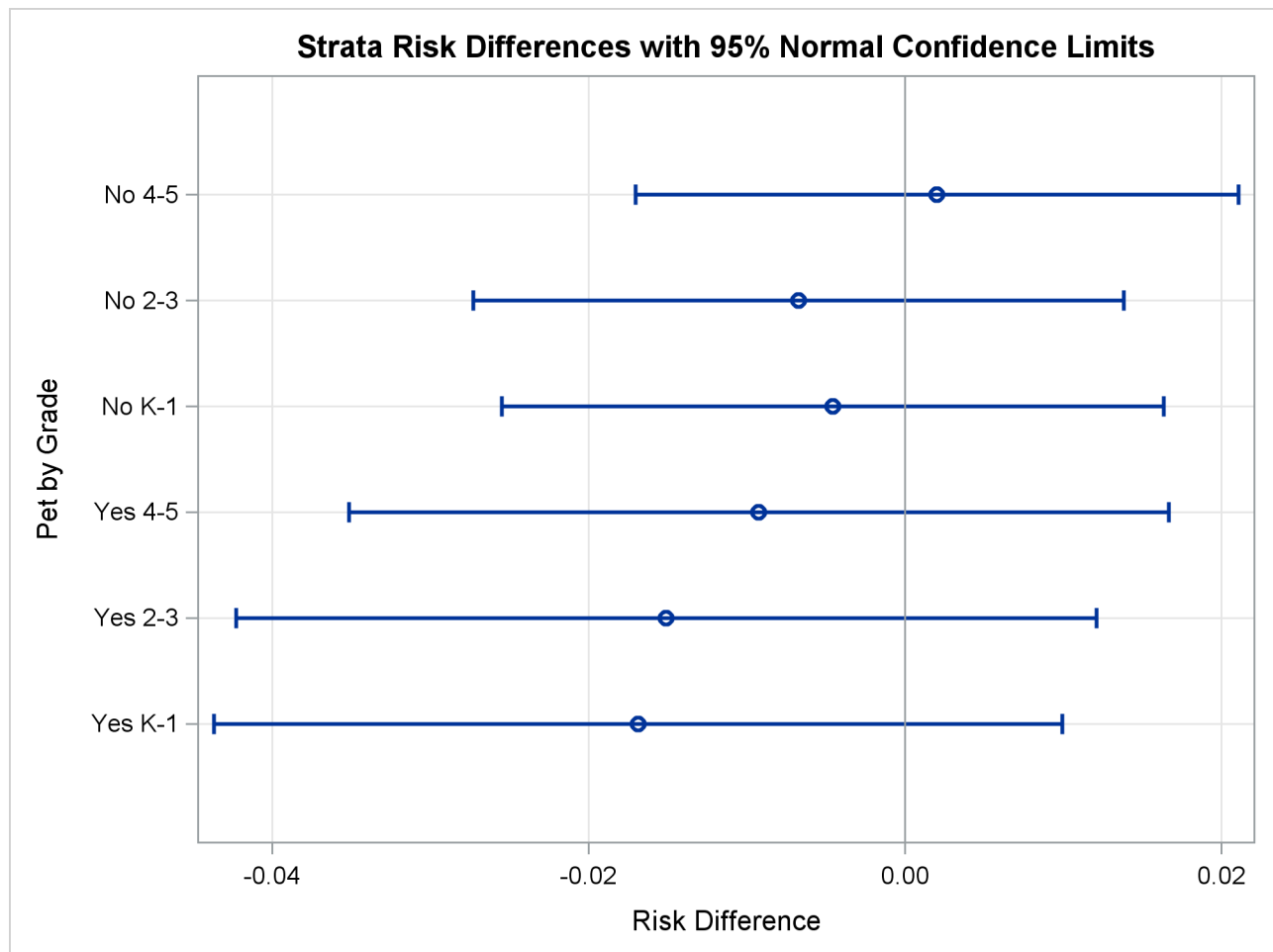
Strata Risk Effect Estimates						
Stratum Index	Pet	Grade	-----Smoking-----		Risk Difference	Standard Error
			No	Yes		
1	Yes	K-1	0.11819	0.13507	-.016883	0.013716
2	Yes	2-3	0.11891	0.13401	-.015098	0.013912
3	Yes	4-5	0.11979	0.12903	-.009243	0.013257
4	No	K-1	0.12184	0.12641	-.004574	0.010704
5	No	2-3	0.11443	0.12117	-.006740	0.010527
6	No	4-5	0.11682	0.11481	0.002014	0.009762

Strata Risk Effect Estimates		
Stratum Index	95% Normal Confidence Limits	
1	-.043766	0.010001
2	-.042366	0.012169
3	-.035225	0.016740
4	-.025554	0.016405
5	-.027373	0.013892
6	-.017120	0.021148

The “Strata Risk Effect Estimates” table shows that for the stratum of students without household pets in Grade 4–5, the risk is higher for students without household smokers than for students with household smokers. For all other strata, the risk is lower for students without household smokers than for students with household smokers. The difference is not significant in each stratum because the null value 0 is between the lower and upper confidence limits.

With ODS Graphics enabled, the PLOTS=EFFECT option displays the plot with the stratum-specific risk effect measures and their associated confidence limits, as shown in [Output 88.2.6](#). The EFFECT=DIFF option requests that the risk difference be displayed. By default, confidence limits are generated with 95% confidence level. This plot displays the stratum-specific risk differences in the “Strata Risk Effect Estimates” table in [Output 88.2.5](#).

**Output 88.2.6** Strata Risk Plot

The “Mantel-Haenszel Standardized Risk Estimates” table in [Output 88.2.7](#) displays the Mantel-Haenszel standardized risks and related statistics.

**Output 88.2.7** Standardized Risk Estimates (Mantel-Haenszel Estimation)

Mantel-Haenszel Standardized Risk Estimates					
Smoking	-----Study Population-----			--Mantel-Haenszel--	
	Observed Events	Number of Observations	Crude Risk	Expected Events	Weight
No	1991	16838	0.1182	566.172	4791.43
Yes	840	6726	0.1249	599.602	4791.43
Mantel-Haenszel Standardized Risk Estimates					
Smoking	-----Standardized Risk-----				
	Estimate	Standard Error	95% Normal Confidence Limits		
No	0.1182	0.00250	0.1133	0.1231	
Yes	0.1251	0.00404	0.1172	0.1331	

The EFFECT=DIFF option requests that the “Risk Effect Estimates” table display the risk difference statistic for the two directly standardized risks, as shown in [Output 88.2.8](#).

**Output 88.2.8** Mantel-Haenszel Effect Estimates

Risk Effect Estimates					
-----Smoking-----		Risk Difference	Standard Error	Z	Pr >  Z
No	Yes				
0.1182	0.1251	-0.00698	0.00475	-1.47	0.1418

The table shows that although the standardized risk for students without household smokes is lower than the standardized risk for students with household smokes, the difference (−0.00698) is not significant at the 5% significance level, ( $p$ -value 0.1418).

---

## Example 88.3: Computing Attributable Fraction Estimates

This example computes the excess event risk fraction that is attributable to a specific chemical exposure for workers in a factory.

Suppose that the Factory data set contains the stratum-specific event information for exposure to a specific chemical agent. The variable Age is the grouping variable that forms the strata. The variables Event\_E and Count\_E indicate the number of events and number of workers for workers with the specific chemical exposure, respectively. The variables Event\_NE and Count\_NE indicate the number of events and number of workers for workers without the specific chemical exposure, respectively.

```

data Factory;
  input Age $ Event_E Count_E Event_NE Count_NE;
  datalines;
20-29    31   352   143   2626
30-39    57   486   392   4124
40-49    62   538   459   4662
50-59    50   455   337   3622
60-69    38   322   199   2155
70+      9    68    35    414
;

```

The following statements invoke the STD RATE procedure and compute the attributable risk and population attributable risk for the chemical exposure:

```

ods graphics on;
proc stdrate data=Factory
  reldata=Factory
  method=indirect (af)
  stat=risk
  plots (stratum=horizontal)
;
  population event=Event_E total=Count_E;
  reference event=Event_NE total=Count_NE;
  strata Age / stats;
run;
ods graphics off;

```

The “Standardization Information” table in [Output 88.3.1](#) displays the standardization information.

**Output 88.3.1** Standardization Information

The STD RATE Procedure	
Standardization Information	
Data Set	WORK.FACTORY
Reference Data Set	WORK.FACTORY
Method	Indirect Standardization
Statistic	Risk
Number of Strata	6

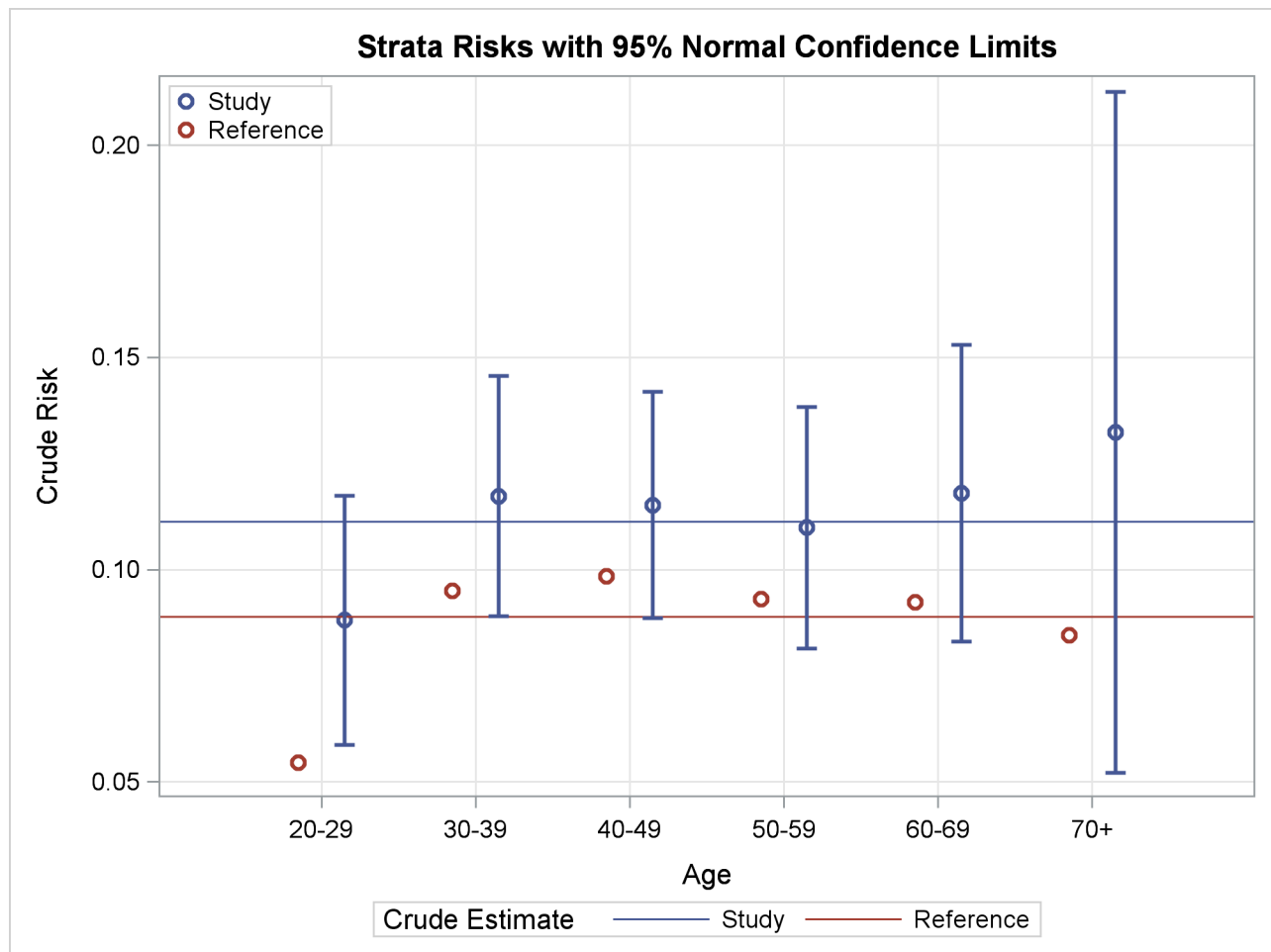
The STATS option in the STRATA statement requests that the “Indirectly Standardized Strata Statistics” table in [Output 88.3.2](#) display the strata information and the expected number of events at each stratum. The Expected Events column shows the expected numbers of events when the stratum-specific risks in the reference data set are applied to the corresponding numbers of workers in the study data set.



**Output 88.3.2** Strata Information (Indirect Standardization)

The STD RATE Procedure						
Indirectly Standardized Strata Statistics						
Stratum Index	Age	-----Study Population-----				
		Observed Events	-Number of Observations- Value	Proportion	Crude Risk	Standard Error
1	20-29	31	352	0.1585	0.088068	0.015105
2	30-39	57	486	0.2188	0.117284	0.014595
3	40-49	62	538	0.2422	0.115242	0.013767
4	50-59	50	455	0.2049	0.109890	0.014662
5	60-69	38	322	0.1450	0.118012	0.017979
6	70+	9	68	0.0306	0.132353	0.041095
Indirectly Standardized Strata Statistics						
Stratum Index	-Study Population- 95% Normal		-----Reference Population-----		Crude Risk	Expected Events
	Confidence Limits		-Number of Observations- Value	Proportion		
1	0.058463	0.117673	2626	0.1492	0.05446	19.1683
2	0.088678	0.145890	4124	0.2343	0.09505	46.1959
3	0.088260	0.142224	4662	0.2648	0.09846	52.9691
4	0.081153	0.138627	3622	0.2058	0.09304	42.3343
5	0.082774	0.153251	2155	0.1224	0.09234	29.7346
6	0.051809	0.212897	414	0.0235	0.08454	5.7488

With ODS Graphics enabled and the specified STAT=RISK option, the default PLOTS=RISK option displays the stratum-specific risk estimates in the study and reference populations, as shown in [Output 88.3.3](#). The STRATUM=HORIZONTAL global option in the PLOTS option displays the strata information on the horizontal axis. The plot displays the stratum-specific risk estimates in the “Indirect Standardized Strata Statistics” table in [Output 88.3.2](#). In addition, confidence limits for the risk estimates in the study population and the overall crude risks for the two populations are also displayed

**Output 88.3.3** Strata Risk Plot

The METHOD=INDIRECT option requests that the “Standardized Morbidity/Mortality Ratio” table in [Output 88.3.4](#) display the SMR, its 95% confidence limits, and the test for the null hypothesis  $H_0 : \text{SMR} = 1$ .

**Output 88.3.4** Standardized Morbidity/Mortality Ratio

Standardized Morbidity/Mortality Ratio							
Observed Events	Expected Events	SMR	Standard Error	95% Normal Confidence Limits		Z	Pr >  Z
247	196.151	1.2592	0.0755	1.1113	1.4072	3.43	0.0006

The “Standardized Morbidity/Mortality Ratio” table shows that  $\text{SMR}=1.259$ , the 95% confidence limits do not contain the null value  $\text{SMR}=1$ , and the null hypothesis of  $\text{SMR}=1$  is rejected at  $\alpha = 0.05$  level from the normal test.

The “Indirectly Standardized Risk Estimates” table in [Output 88.3.5](#) displays the standardized risks and related statistics.

**Output 88.3.5** Standardized Risks (Indirect Standardization)

Indirectly Standardized Risk Estimates					
-----Study Population-----			Reference		
Observed Events	Number of Observations	Crude Risk	Crude Risk	Expected Events	SMR
247	2221	0.1112	0.0889	196.151	1.2592
Indirectly Standardized Risk Estimates					
-----Standardized Risk-----					
Estimate	Standard Error	95% Normal Confidence Limits			
0.1120	0.00671	0.0988	0.1251		

The AF suboption in the METHOD=INDIRECT option requests that the “Attributable Fraction Estimates” table display the attributable risk and population attributable risk, as shown in [Output 88.3.6](#)

**Output 88.3.6** Attributable Fraction Estimates

Attributable Fraction Estimates			
Parameter	Estimate	-95% Confidence Limits-	
Attributable Risk	0.20587	0.10013	0.28937
Population Attributable Risk	0.02806	0.01159	0.04426

The attributable risk fraction 0.206 indicates that 20.6% of all events in the chemical exposure group are attributed to the chemical exposure, and the population attributable risk fraction 0.028 indicates that about 2.8% of all events in the total population are attributed to the chemical exposure.

The Attributable fraction can also be computed by using Mantel-Haenszel method.

Suppose that the Factory1 data set contains the stratum-specific event information for exposure to a specific chemical agent. The variable Age is the grouping variable that forms the strata, and the variable Exposure identifies workers with chemical exposure. The variables Event and Count indicate the number of events and number of workers, respectively.

```
data Factory1;
  input Exposure $ Age $ Event Count;
  datalines;
Yes 20-29 31 352
Yes 30-39 57 486
Yes 40-49 62 538
Yes 50-59 50 455
Yes 60-69 38 322
Yes 70+ 9 68
No 20-29 143 2626
```

```

No    30-39    392    4124
No    40-49    459    4662
No    50-59    337    3622
No    60-69    199    2155
No    70+      35     414
;

```

The following statements invoke the STD RATE procedure and compute the attributable risk and population attributable risk for the chemical exposure:

```

proc stdrate data=Factory1
    method=mh (af)
    stat=risk
    effect
;
    population group(order=data)=Exposure event=Event total=Count;
    strata Age;
run;

```

The “Standardization Information” table in [Output 88.3.7](#) displays the standardization information.

**Output 88.3.7** Standardization Information

The STD RATE Procedure	
Standardization Information	
Data Set	WORK.FACTORY1
Group Variable	Exposure
Method	Mantel-Haenszel
Statistic	Risk
Number of Strata	6

The “Mantel-Haenszel Standardized Risk Estimates” table in [Output 88.3.8](#) displays the Mantel-Haenszel standardized risks and related statistics.

**Output 88.3.8** Standardized Risk Estimates (Mantel-Haenszel Estimation)

Mantel-Haenszel Standardized Risk Estimates					
Exposure	-----Study Population-----			--Mantel-Haenszel--	
	Observed Events	Number of Observations	Crude Risk	Expected Events	Weight
Yes	247	2221	0.1112	219.122	1970.26
No	1565	17603	0.0889	174.134	1970.26

Mantel-Haenszel Standardized Risk Estimates				
Exposure	-----Standardized Risk-----			
	Estimate	Standard Error	95% Normal Confidence Limits	
Yes	0.1112	0.00667	0.0981	0.1243
No	0.0884	0.00214	0.0842	0.0926

The EFFECT option requests that the “Risk Effect Estimates” table display the risk ratio statistic for the two directly standardized risks, as shown in [Output 88.3.9](#).

**Output 88.3.9** Mantel-Haenszel Effect Estimates

Risk Effect Estimates						
-----Exposure-----		Risk Ratio	Log Risk Ratio	Standard Error	Z	Pr >  Z
Yes	No					
0.1112	0.0884	1.2584	0.2298	0.0647	3.55	0.0004

The AF suboption in the METHOD=MH option requests that the “Attributable Fraction Estimates” table display the attributable risk and population attributable risk, as shown in [Output 88.3.10](#)

**Output 88.3.10** Attributable Fraction Estimates

Attributable Fraction Estimates			
Parameter	Estimate	-95% Confidence Limits-	
Attributable Risk	0.20531	0.09789	0.29994
Population Attributable Risk	0.02799	0.01070	0.04497

Similar to the results of using the SMR estimates, the attributable risk fraction (0.205) indicates that 20.5% of all events in the chemical exposure group are attributed to the chemical exposure, and the population attributable risk fraction (0.028) indicates that about 2.8% of all events in the total population are attributed to the chemical exposure.

### Example 88.4: Displaying SMR Results from BY Groups

This example illustrates the use of ODS OUTPUT statement to save standardized mortality ratios for different causes and to display these statistics together in a table and in a plot.

The Florida\_Cs data set contains the stratum-specific mortality information for stomach cancer and skin cancer in year 2000 for the state of Florida (Florida Department of Health 2000, 2012). The variable Age is the grouping variable that forms the strata in the standardization. The variables Event\_C16, Event\_C43, and PYear identify the number of events for stomach cancer, the number of events for skin cancer, and the person-years, respectively. The COMMA12. format is specified in the DATA step to input numerical values that contain commas in PYear.

```
data Florida_Cs;
  input Age $1-5 Event_C16 Event_C43 PYear comma12.;
  datalines;
00-04    0    0    953,785
05-14    0    0  1,997,935
15-24    0    4  1,885,014
25-34    1   14  1,957,573
35-44   19   43  2,356,649
45-54   64   72  2,088,000
55-64  114   70  1,548,371
65-74  201  126  1,447,432
75-84  294  136  1,087,524
85+    136   73   335,944
;
```

The following statements construct and list the mortality information by cancer cause:

```
data Florida_Cs;
  set Florida_Cs;
  Cause='Stomach'; Event=Event_C16; output;
  Cause='Skin'; Event=Event_C43; output;
  drop Event_C16 Event_C43;
run;

proc sort data=Florida_Cs;
  by Cause;
run;

proc print data=Florida_Cs;
  var Cause Age Event PYear;
run;
```

**Output 88.4.1** Florida Data

Obs	Cause	Age	Event	PYear
1	Skin	00-04	0	953785
2	Skin	05-14	0	1997935
3	Skin	15-24	4	1885014
4	Skin	25-34	14	1957573
5	Skin	35-44	43	2356649
6	Skin	45-54	72	2088000
7	Skin	55-64	70	1548371
8	Skin	65-74	126	1447432
9	Skin	75-84	136	1087524
10	Skin	85+	73	335944
11	Stomach	00-04	0	953785
12	Stomach	05-14	0	1997935
13	Stomach	15-24	0	1885014
14	Stomach	25-34	1	1957573
15	Stomach	35-44	19	2356649
16	Stomach	45-54	64	2088000
17	Stomach	55-64	114	1548371
18	Stomach	65-74	201	1447432
19	Stomach	75-84	294	1087524
20	Stomach	85+	136	335944

The US\_Cs data set contains the corresponding stratum-specific mortality information for the United States (Miniño et al. 2002; U.S. Bureau of Census 2011). The variable Age is the grouping variable that forms the strata in the standardization. The variables Event\_C16, Event\_C43, and PYear identify the number of events for stomach cancer, the number of events for skin cancer, and the person-years, respectively.

```
data US_Cs;
  input Age $1-5 Event_C16 Event_C43 PYear comma12.;
  datalines;
00-04      0      0 19,175,798
05-14      1      1 41,077,577
15-24     14     41 39,183,891
25-34    124    186 39,892,024
35-44    484    626 45,148,527
45-54   1097   1199 37,677,952
55-64   1804   1303 24,274,684
65-74   3054   1637 18,390,986
75-84   3833   1624 12,361,180
85+     2234    803  4,239,587
;
```

The following statements construct and list the mortality information by cancer cause:

```
data US_Cs;
  set US_Cs;
  Cause='Stomach'; Event=Event_C16; output;
  Cause='Skin'; Event=Event_C43; output;
  drop Event_C16 Event_C43;
run;
```

```
proc sort data=US_Cs;
  by Cause;
run;

proc print data=US_Cs;
  var Cause Age Event PYear;
run;
```

Output 88.4.2 lists the mortality information by cancer cause.

**Output 88.4.2** Florida Data

Obs	Cause	Age	Event	PYear
1	Skin	00-04	0	19175798
2	Skin	05-14	1	41077577
3	Skin	15-24	41	39183891
4	Skin	25-34	186	39892024
5	Skin	35-44	626	45148527
6	Skin	45-54	1199	37677952
7	Skin	55-64	1303	24274684
8	Skin	65-74	1637	18390986
9	Skin	75-84	1624	12361180
10	Skin	85+	803	4239587
11	Stomach	00-04	0	19175798
12	Stomach	05-14	1	41077577
13	Stomach	15-24	14	39183891
14	Stomach	25-34	124	39892024
15	Stomach	35-44	484	45148527
16	Stomach	45-54	1097	37677952
17	Stomach	55-64	1804	24274684
18	Stomach	65-74	3054	18390986
19	Stomach	75-84	3833	12361180
20	Stomach	85+	2234	4239587

The following statements invoke the STD RATE procedure and request indirect standardization to compute the skin and stomach SMR estimates for the state of Florida. The BY statement requests separate analyses of causes that are defined by the Cause variable.

```
ods graphics on;
ods select StdInfo StrataSmrPlot Smr;
proc stdrate data=Florida_Cs refdata=US_Cs
  stat=rate
  method=indirect
  plots=smr
  ;
  population event=Event total=PYear;
  reference event=Event total=PYear;
  strata Age;
  by Cause;
ods output smr=Smr_Cs;
run;
ods graphics off;
```



Only the tables and plots that are specified in the ODS SELECT statement are displayed.

The STDINFO option in the ODS SELECT statement requests that the “Standardization Information” table display the standardization information for the first BY group, skin cancer, as shown in [Output 88.4.3](#)

**Output 88.4.3** Standardization Information

----- Cause=Skin -----	
The STDRATE Procedure	
Standardization Information	
Data Set	WORK.FLORIDA_CS
Reference Data Set	WORK.US_CS
Method	Indirect Standardization
Statistic	Rate
Number of Strata	10
Rate Multiplier	100000

The STRATASMRPLOT option in the ODS SELECT statement requests that the strata SMR plot display stratum-specific SMR estimates for skin cancer with confidence limits, as shown in [Output 88.4.4](#).

**Output 88.4.4** Strata SMR Plot

The SMR option in the ODS SELECT statement requests that the “Standardized Morbidity/Mortality Ratio” table display the SMR, its confidence limits, and the test for the null hypothesis  $H_0 : \text{SMR} = 1$  for skin cancer, as shown in [Output 88.4.5](#). With the default ALPHA=0.05, 95% confidence limits are constructed.

**Output 88.4.5** Standardized Morbidity/Mortality Ratio

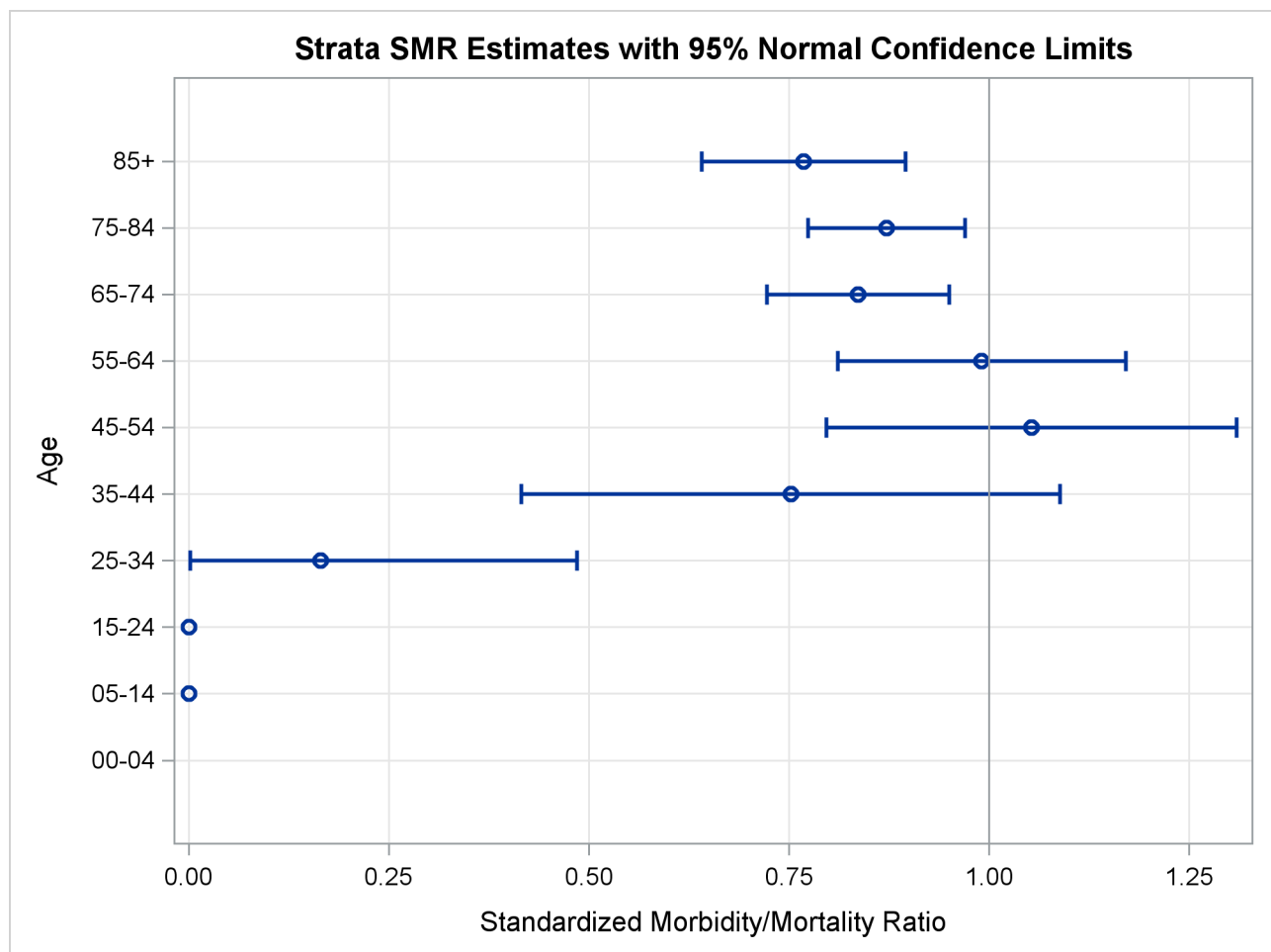
----- Cause=Skin -----							
Standardized Morbidity/Mortality Ratio							
Observed Events	Expected Events	SMR	Standard Error	95% Normal Confidence Limits		Z	Pr >  Z
538	528.726	1.0175	0.0439	0.9316	1.1035	0.40	0.6893

Similarly, the “Standardization Information” table in [Output 88.4.6](#) displays the standardization information for the second BY group, stomach cancer.

**Output 88.4.6** Standardization Information

----- Cause=Stomach -----	
The STDRATE Procedure	
Standardization Information	
Data Set	WORK.FLORIDA_CS
Reference Data Set	WORK.US_CS
Method	Indirect Standardization
Statistic	Rate
Number of Strata	10
Rate Multiplier	100000

The “Strata SMR Plot” displays stratum-specific SMR estimates with confidence limits for stomach cancer, as shown in [Output 88.4.7](#).

**Output 88.4.7** Strata SMR Plot

The “Standardized Morbidity/Mortality Ratio” table displays the SMR, its confidence limits, and the test for the null hypothesis  $H_0 : \text{SMR} = 1$  for stomach cancer, as shown in [Output 88.4.8](#).

**Output 88.4.8** Standardized Morbidity/Mortality Ratio

----- Cause=Stomach -----							
Standardized Morbidity/Mortality Ratio							
Observed Events	Expected Events	SMR	Standard Error	95% Normal Confidence Limits		Z	Pr >  Z
829	962.537	0.8613	0.0299	0.8026	0.9199	-4.64	<.0001

The ODS OUTPUT SMR=SMR\_CS statement requests that the “Standardized Morbidity/Mortality Ratio” tables for the two cancer causes be saved in the data set Smr\_Cs. The following statements display the selected output variables for the data set:

```
proc print data=Smr_Cs;
  var Cause ObservedEvents ExpectedEvents Smr SmrLcl SmrUcl;
run;
```

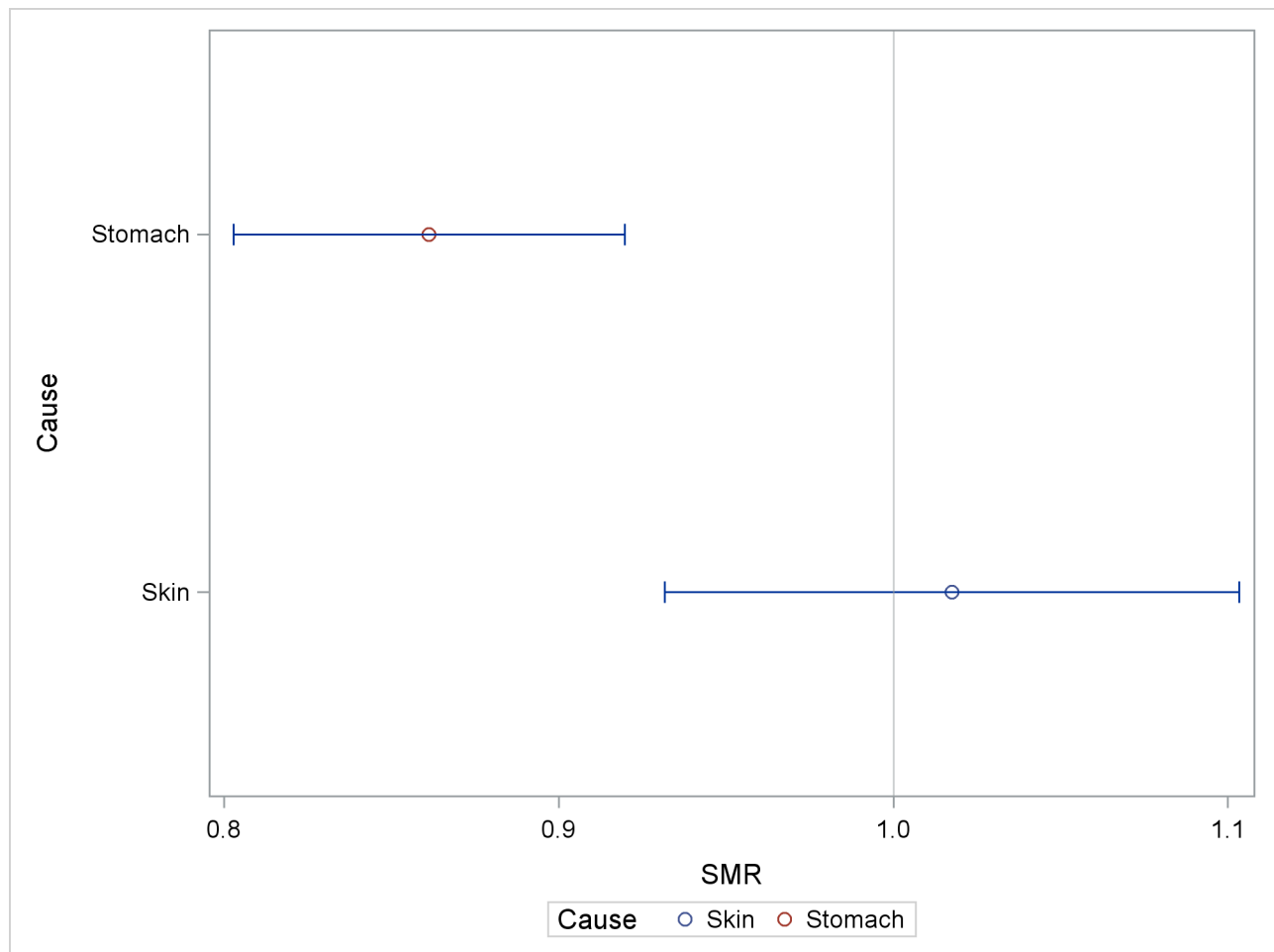
**Output 88.4.9** SMR Results from BY Groups

Obs	Cause	Observed Events	Expected Events	Smr	SmrLcl	SmrUcl
1	Skin	538	528.726	1.0175	0.9316	1.1035
2	Stomach	829	962.537	0.8613	0.8026	0.9199

The table in [Output 88.4.9](#) shows that the study population (state of Florida) has a higher skin cancer rate and a lower stomach cancer rate than the reference population (United States), but only the lower stomach cancer rate is significant because its corresponding SMR upper confidence limit (0.9199) is less than 1.

The following statements display the standardized morbidity/mortality ratios for the two causes in a plot:

```
proc sgplot data=Smr_Cs;
  scatter y=Cause x=Smr / group=Cause;
  highlow y=Cause high=SmrUcl low=SmrLcl / highcap=serif lowcap=serif;
  yaxis type=discrete;
  xaxis label="SMR";
  refline 1 / axis=x transparency=0.5;
run;
```

**Output 88.4.10** SMR Results

Alternatively, you can also use the following statements to obtain separate analyses for the two cancer causes, and then to display these standardized mortality ratios together in a table and in a plot:

```

/*----- Perform Separate Analyses for Different Causes -----*/
ods graphics on;
ods select StdInfo StrataSmrPlot Smr;
proc stdrate data=Florida-Cs refdata=US-Cs
    stat=rate
    method=indirect
    plots=smr
    ;
    population event=Event_C43 total=PYear;
    reference event=Event_C43 total=PYear;
    strata Age;
ods output smr=Smr_c43;
run;

```

```

ods select StdInfo StrataSmrPlot Smr;
proc stdrate data=Florida_Cs refdata=US_Cs
            stat=rate
            method=indirect
            plots=smr
            ;
    population event=Event_C16 total=PYear;
    reference  event=Event_C16 total=PYear;
    strata Age;
ods output smr=Smr_c16;
run;
ods graphics off;

/*----- Combine SMRs -----*/
data Smr_C43;
    set Smr_C43;
    length Cause $ 7.;
    Cause='Skin';
run;

data Smr_C16;
    set Smr_C16;
    length Cause $ 7.;
    Cause='Stomach';
run;

data Smr_Cs;
    set Smr_C43 Smr_C16;
run;

/*----- Display the Cause-Specific SMRs -----*/
proc print data=Smr_Cs;
    var Cause ObservedEvents ExpectedEvents Smr SmrLcl SmrUcl;
run;

proc sgplot data=Smr_Cs;
    scatter y=Cause x=Smr / group=Cause;
    highlow y=Cause high=SmrUcl low=SmrLcl / highcap=serif lowcap=serif;
    yaxis type=discrete;
    xaxis label="SMR";
    refline 1 / axis=x transparency=0.5;
run;

```

---

## References

- Alaska Bureau of Vital Statistics (2000a), “2000 Annual Report: Appendix I: Population Overview,” Accessed February 2012.  
 URL [http://www.hss.state.ak.us/dph/bvs/PDFs/2000/annual\\_report/Appendix\\_I.pdf](http://www.hss.state.ak.us/dph/bvs/PDFs/2000/annual_report/Appendix_I.pdf)

- Alaska Bureau of Vital Statistics (2000b), "2000 Annual Report: Deaths," Accessed February 2012.  
URL [http://www.hss.state.ak.us/dph/bvs/PDFs/2000/annual\\_report/Death\\_chapter.pdf](http://www.hss.state.ak.us/dph/bvs/PDFs/2000/annual_report/Death_chapter.pdf)
- Breslow, N. E. and Day, N. E. (1987), *Statistical Methods in Cancer Research, Volume II: The Design and Analysis of Cohort Studies*, IARC Scientific Publications, No. 82, Lyon, France: International Agency for Research on Cancer.
- Elandt-Johnson, R. C. and Johnson, N. L. (1980), *Survival Models and Data Analysis*, New York: John Wiley & Sons.
- Fay, M. P. and Feuer, E. J. (1997), "Confidence Intervals for Directly Standardized Rates: A Method Based on the Gamma Distribution," *Statistics in Medicine*, 16, 791–801.
- Florida Department of Health (2000), "Florida Vital Statistics Annual Report 2000," Accessed February 2012.  
URL <http://www.flpublichealth.com/VSB00K/pdf/2000/Population.pdf>
- Florida Department of Health (2012), "Florida Death Query System," Accessed February 2012.  
URL <http://www.floridacharts.com/charts/DeathQuery.aspx>
- Friss, R. H. and Sellers, T. A. (2009), *Epidemiology for Public Health Practice*, Fourth Edition, Sudbury, MA: Jones and Bartlett.
- Greenland, S. (2008), "Applications of Stratified Analysis Methods," in K. J. Rothman, S. Greenland, and T. L. Lash, eds., *Modern Epidemiology*, Third Edition, Philadelphia: Lippincott Williams & Wilkins.
- Greenland, S. and Robins, J. M. (1985), "Estimation of a Common Effect Parameter from Sparse Follow-up Data," *Biometrics*, 41, 55–68.
- Greenland, S. and Rothman, K. J. (2008), "Introduction to Stratified Analysis," in K. J. Rothman, S. Greenland, and T. L. Lash, eds., *Modern Epidemiology*, Third Edition, Philadelphia: Lippincott Williams & Wilkins.
- Kleinbaum, D. G., Kupper, L. L., and Morgenstern, H. (1982), *Epidemiologic Research: Principles and Quantitative Methods*, Research Methods Series, New York: Van Nostrand Reinhold.
- Miniño, A. M., Arias, E., Kochanek, K. D., Murphy, S. L., and Smith, B. L. (2002), "Deaths: Final Data for 2000," Accessed February 2012.  
URL [http://www.cdc.gov/nchs/data/nvsr/nvsr50/nvsr50\\_15.pdf](http://www.cdc.gov/nchs/data/nvsr/nvsr50/nvsr50_15.pdf)
- Porta, M. (2008), *A Dictionary of Epidemiology*, Fifth Edition, New York: Oxford University Press.
- Tiwari, R. C., Clegg, L. X., and Zou, Z. (2006), "Efficient Interval Estimation for Age-Adjusted Cancer Rates," *Statistics Methods in Medical Research*, 15, 547–569.
- U.S. Bureau of Census (2011), "Age and Sex Composition: 2010," Accessed February 2012.  
URL <http://www.census.gov/prod/cen2010/briefs/c2010br-03.pdf>





# Subject Index

- AF
  - STD RATE procedure, 7517
- attributable fraction
  - STD RATE procedure, 7517
- attributable fraction estimates
  - STD RATE procedure, 7522
- confidence limits
  - STD RATE procedure, 7498
- directly standardized rate
  - STD RATE procedure, 7510
- distribution plot
  - STD RATE procedure, 7528
- effect estimates
  - STD RATE procedure, 7522
- effect measure plot
  - STD RATE procedure, 7528
- indirect standardization
  - STD RATE procedure, 7514
- Mantel-Haenszel effect estimation
  - STD RATE procedure, 7512
- method specifications
  - STD RATE procedure, 7520
- population attributable fraction
  - STD RATE procedure, 7517
- rate
  - STD RATE procedure, 7505
- risk
  - STD RATE procedure, 7508
- significance level
  - STD RATE procedure, 7498
- SMR
  - STD RATE procedure, 7514
- SMR plot
  - STD RATE procedure, 7528
- standardization information
  - STD RATE procedure, 7523
- standardized morbidity ratio
  - STD RATE procedure, 7523
- standardized mortality ratio
  - STD RATE procedure, 7523
- standardized rate, 7487
- standardized rate estimates
  - STD RATE procedure, 7524
- standardized risk estimates
  - STD RATE procedure, 7524
- STD RATE procedure, 7521
  - AF, 7517
  - attributable fraction, 7517
  - attributable fraction estimates, 7522
  - confidence limits, 7498
  - directly standardized rate, 7510
  - distribution plot, 7528
  - effect estimates, 7522
  - effect measure plot, 7528
  - indirect standardization, 7514
  - introductory example, 7489
  - Mantel-Haenszel effect estimation, 7512
  - method specifications, 7520
  - ODS Graphics names, 7528
  - ODS table names, 7527
  - population attributable fraction, 7517
  - rate, 7505
  - risk, 7508
  - significance level, 7498
  - SMR, 7514
  - SMR plot, 7528
  - standardization information, 7523
  - standardized morbidity ratio, 7523
  - standardized mortality ratio, 7523
  - standardized rate estimates, 7524
  - standardized risk estimates, 7524
  - strata effect estimates, 7525
  - strata information, 7525
  - strata rate plot, 7528
  - strata risk plot, 7528
  - strata SMR estimates, 7526
  - syntax, 7497
  - table output, 7522
- strata effect estimates
  - STD RATE procedure, 7525
- strata information
  - STD RATE procedure, 7525
- strata rate plot
  - STD RATE procedure, 7528
- strata risk plot
  - STD RATE procedure, 7528
- strata SMR estimates
  - STD RATE procedure, 7526



# Syntax Index

ALPHA= option  
PROC STDRA TE statement, 7498

BY statement  
STDRA TE procedure, 7501

CL= option  
PROC STDRA TE statement, 7498

DATA option  
PROC STDRA TE statement, 7499

EFFECT option  
PROC STDRA TE statement, 7499  
STRATA statement (STDRA TE), 7504

EVENT= option  
POPULATION statement (STDRA TE), 7501  
REFERENCE statement (STDRA TE), 7502

GROUP= option  
POPULATION statement (STDRA TE), 7501

METHOD= option  
PROC STDRA TE statement, 7499

MISSING option  
STRATA statement (STDRA TE), 7504

MULT= option  
POPULATION statement (STDRA TE), 7502  
PROC STDRA TE statement, 7501  
REFERENCE statement (STDRA TE), 7502, 7503

ORDER= option  
POPULATION statement (STDRA TE), 7501  
STRATA statement (STDRA TE), 7504

PLOTS option  
PROC STDRA TE statement, 7499

POPEVENT= option  
POPULATION statement (STDRA TE), 7502

POPULATION statement  
STDRA TE procedure, 7501

PROC STDRA TE statement, *see* STDRA TE procedure

RATE= option  
POPULATION statement (STDRA TE), 7502  
REFERENCE statement (STDRA TE), 7502

REFDATA option  
PROC STDRA TE statement, 7500

REFERENCE statement  
STDRA TE procedure, 7502

REFEVENT= option  
REFERENCE statement (STDRA TE), 7503

REFRATE= option  
REFERENCE statement (STDRA TE), 7503

RISK= option  
POPULATION statement (STDRA TE), 7502  
REFERENCE statement (STDRA TE), 7503

SMR option  
STRATA statement (STDRA TE), 7504

STAT= option  
PROC STDRA TE statement, 7501

STATS option  
STRATA statement (STDRA TE), 7504

STDRA TE procedure, BY statement, 7501  
STDRA TE procedure, POPULATION statement, 7501

EVENT= option, 7501

GROUP= option, 7501

MULT= option, 7502

ORDER= option, 7501

POPEVENT= option, 7502

RATE= option, 7502

RISK= option, 7502

TOTAL= option, 7502

STDRA TE procedure, PROC STDRA TE statement, 7498

ALPHA= option, 7498

CL option, 7498

DATA= option, 7499

EFFECT option, 7499

METHOD= option, 7499

MULT= option, 7501

PLOTS option, 7499

REFDATA= option, 7500

STAT= option, 7501

STDRA TE procedure, REFERENCE statement, 7502

EVENT= option, 7502

MULT= option, 7502, 7503

RATE= option, 7502

REFEVENT= option, 7503

REFRATE= option, 7503

REFRISK= option, 7503

REFTOTAL= option, 7503

RISK= option, 7503

TOTAL= option, 7503

STD RATE procedure, STRATA statement, [7503](#)

EFFECT option, [7504](#)

INFO option, [7504](#)

MISSING option, [7504](#)

ORDER= option, [7504](#)

SMR option, [7504](#)

STRATA statement

STD RATE procedure, [7503](#)

TOTAL= option

POPULATION statement (STD RATE), [7502](#)

REFERENCE statement (STD RATE), [7503](#)

## Your Turn

---

We welcome your feedback.

- If you have comments about this book, please send them to **`yourturn@sas.com`**. Include the full title and page numbers (if applicable).
- If you have comments about the software, please send them to **`suggest@sas.com`**.



# SAS® Publishing Delivers!

Whether you are new to the work force or an experienced professional, you need to distinguish yourself in this rapidly changing and competitive job market. SAS® Publishing provides you with a wide range of resources to help you set yourself apart. Visit us online at [support.sas.com/bookstore](http://support.sas.com/bookstore).

## SAS® Press

Need to learn the basics? Struggling with a programming problem? You'll find the expert answers that you need in example-rich books from SAS Press. Written by experienced SAS professionals from around the world, SAS Press books deliver real-world insights on a broad range of topics for all skill levels.

**[support.sas.com/saspress](http://support.sas.com/saspress)**

## SAS® Documentation

To successfully implement applications using SAS software, companies in every industry and on every continent all turn to the one source for accurate, timely, and reliable information: SAS documentation. We currently produce the following types of reference documentation to improve your work experience:

- Online help that is built into the software.
- Tutorials that are integrated into the product.
- Reference documentation delivered in HTML and PDF – **free** on the Web.
- Hard-copy books.

**[support.sas.com/publishing](http://support.sas.com/publishing)**

## SAS® Publishing News

Subscribe to SAS Publishing News to receive up-to-date information about all new SAS titles, author podcasts, and new Web site features via e-mail. Complete instructions on how to subscribe, as well as access to past issues, are available at our Web site.

**[support.sas.com/spn](http://support.sas.com/spn)**



**THE  
POWER  
TO KNOW®**

