Paper 423-2013

Computing Direct and Indirect Standardized Rates and Risks with the STDRATE Procedure

Yang Yuan, SAS Institute Inc.

ABSTRACT

In epidemiological and health care studies, a common goal is to establish relationships between various factors and event outcomes. But outcome measures such as rates or risks can be biased by confounding. You can control for confounding by dividing the population into homogeneous strata and estimating rate or risk based on a weighted average of stratum-specific rate or risk estimates. This paper reviews the concepts of standardized rate and risk and introduces the STDRATE procedure, which is new in SAS/STAT® 12.1. PROC STDRATE computes directly standardized rates and risks by using Mantel-Haenszel estimates, and it computes indirectly standardized rates and risks by using standardized morbidity/mortality ratios (SMR). PROC STDRATE also provides stratum-specific summary statistics, such as rate and risk estimates and confidence limits.

INTRODUCTION

Epidemiology is the study of the occurrence and distribution of health-related states or events in specified populations. It is also the study of causal mechanisms for health phenomena in populations (Friss and Sellers 2009, p. 5). A goal of epidemiology is to establish relationships between various factors (such as exposure to a specific chemical) and event outcomes (such as incidence of disease). Two commonly used event frequency measures are rate and risk, which are defined as follows:

- An event rate in a defined population is a measure of the frequency with which an event occurs in a specified period of time. That is, an event rate is the number of new events divided by population-time (for example, person-years) over the time period (Kleinbaum, Kupper, and Morgenstern 1982, p. 100).
- An event risk in a defined population is the probability that an event occurs in a specified time period. That is, an event risk is the number of events divided by the population size in the time period.

Event rates and risks can be biased by confounding, which occurs when other variables that are associated with exposure influence the outcome. For example, when event rates vary for different age groups of a population, the crude rate for the population (unadjusted for age structure) might not be a meaningful summary statistic. In particular, the crude rate might be misleading when it is used to compare two populations that differ in their age structures.

A common strategy for controlling confounding is stratification. You begin by subdividing the population into several strata that are defined by levels of the confounding variables, such as age. You estimate the effect of exposure on the event outcome within each stratum, and then you combine the resulting stratum-specific effect estimates into an overall estimate.

Standardized overall rate and risk estimates that are based on stratum-specific estimates adjust for the effects of confounding variables. These estimates provide meaningful summary statistics and allow valid comparisons of populations. There are two types of standardization:

- Direct standardization uses the weights from a standard or reference population to compute the weighted average of stratum-specific rate or risk estimates in the study population. When you use the same reference population to compute directly standardized estimates for two populations, you can also compare the resulting estimates.

- Indirect standardization uses the stratum-specific rate or risk estimates in the reference population to compute the expected number of events in the study population. The ratio of the observed number of events to the computed expected number of events in the study population is the standardized morbidity ratio (SMR). SMR is also the standardized mortality ratio if the event is death; you can use it to compare rates or risks between the study and reference populations.

The STDRATE (pronounced "standard rate") procedure provides both directly standardized and indirectly standardized rate and risk estimates. In addition, if an effect (such as the rate difference between two populations) is homogeneous across strata, PROC STDRATE also provides the Mantel-Haenszel method (Greenland and Rothman 2008, p. 271) to compute a pooled estimate of the effect that is based on these stratum-specific effect estimates.

Note: The term standardization has different meanings in other statistical applications. For example, the STANDARD procedure standardizes numeric variables in a SAS data set to a given mean and standard deviation.

The following three sections describe the main features of PROC STDRATE: direct standardization, Mantel-Haenszel estimation, and indirect standardization and SMR. Each section includes an example. These sections are followed by a summary section that summarizes the main features of PROC STDRATE.

DIRECT STANDARDIZATION

Direct standardization uses the weights from a standard or reference population to compute the weighted average of stratum-specific estimates in the study population. The directly 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 jth stratum of the study population, $\mathcal{T}_{rj}$ is the population-time in the jth stratum of the reference population, and $\mathcal{T}_r = \sum_k \mathcal{T}_{rk}$ is the total population-time in the reference population.

The standardized risk can also be computed similarly.

The direct standardization is applicable when the study population is large enough to provide stable stratum-specific estimates. The directly standardized estimate is the overall crude estimate in the study population if it has the same strata distribution as the reference population.

When you use the same reference population to derive standardized estimates for different populations, you can also use the estimated difference and estimated ratio statistics to compare the resulting estimates.

EXAMPLE: 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 following 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):

```
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 variables **Sex** and **Age** are the grouping variables that form the strata in the standardization, and the 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**.

The following Florida data set contains the corresponding stratum-specific mortality information for the state of Florida (Florida Department of Health 2000, 2012):

```
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 crude rate for Alaska (2924/626932 = 0.004664) is less than the crude rate for Florida (76455/15577105 = 0.004908). However, because the age distributions in the two states differ widely, these crude rates might not provide a valid comparison.

To compare standardized rates for the two populations, you can combine the two data sets to form a single data set to be used in the DATA= option. The following TwoStates data set contains the data sets Alaska and Florida, where the variable **State** identifies the two states:

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

The following US data set contains the stratum-specific person-years information for the United States (U.S. Bureau of the Census 2011):

```
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 PROC STDRATE and compute the direct standardized rates for the states of Florida and Alaska by using the United States as the reference population:

```
ods graphics on;
proc stdrate data=TwoStates
             refdata=US
             method=direct
             stat=rate(mult=1000)
             effect=ratio
             plots(only)=effect
             ;
   population group=State event=Death total=PYear;
   reference  total=PYear;
   strata Sex Age / effect;
run;
ods graphics off;
```

The DATA= option names the data set for the study populations, and the REFDATA= option names the data set for the reference population. The METHOD=DIRECT option requests direct standardization. The STAT=RATE option specifies the rate statistic for standardization, and the MULT=1000 suboption requests that rates per 1,000 person-years be displayed. When you specify the EFFECT=RATIO and STAT=RATE options, PROC STDRATE computes the rate ratio effect between the study populations.

The POPULATION and REFERENCE statements specify the options that are related to the study and reference populations, respectively. The EVENT= option specifies the variable for the number of events in the study population, the TOTAL= option specifies the variable for the person-years in the populations, and the GROUP=STATE option specifies the variable that identifies the Alaska and Florida populations in the DATA= data set.

The “Standardization Information” table in Figure 1 displays the standardization information.

Figure 1 Standardization Information

The STDRATE 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

The EFFECT option in the STRATA statement and the STAT=RATE option in the PROC STDRATE statement display the “Strata Rate Effect Estimates” table, as shown in Figure 2. The EFFECT=RATIO option in the PROC STDRATE statement requests that the stratum-specific rate ratio statistics be displayed.

Figure 2 Strata Effect Estimates

The STDRATE Procedure

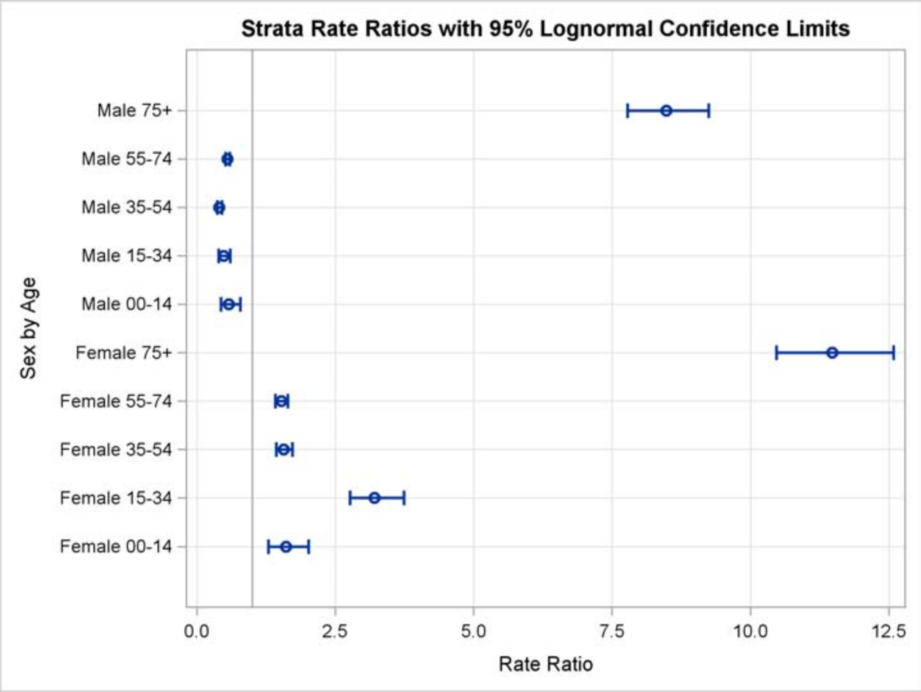
Strata Rate Effect Estimates (Rate Multiplier = 1000)

Stratum Index	Sex	Age	State Alaska	State 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 than Florida for male groups and higher mortality rates for female groups. For the age group 75+, Alaska has much higher mortality rates than Florida for both male and female groups.

With ODS Graphics enabled, the PLOTS(ONLY)=EFFECT option displays only the strata effect plot; the default strata rate plot is not displayed. The strata effect measure plot includes the stratum-specific effect measures and their associated confidence limits, as shown in Figure 3. The STAT=RATE option and the EFFECT=RATIO option request that the strata rate ratios be displayed. By default, confidence limits are generated at a 95% confidence level. This plot displays the stratum-specific rate ratios that are shown in the “Strata Rate Effect Estimates” table in Figure 2.

Figure 3 Strata Effect Measure Plot



The “Directly Standardized Rate Estimates” table in Figure 4 displays directly standardized rates and related statistics.

Figure 4 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			
State	Estimate	Standard Error	95% Normal 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 in Figure 4 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 actually lower than the standardized rate in the Alaska population (4.2289).

The EFFECT=RATIO option requests that the “Rate Effect Estimates” table in Figure 5 display the log rate ratio statistics of the two directly standardized rates.

Figure 5 Effect Estimates

Rate Effect Estimates (Rate Multiplier = 1000)

State						
Alaska	Florida	Rate Ratio	Log Rate Ratio	Standard Error	Z	Pr > |Z|
4.2289	4.0385	1.0471	0.0461	0.0217	2.13	0.0335

The table in Figure 5 shows that when the log rate ratio statistic is 1.047, the resulting p-value is 0.0335, indicating that the death rate is significantly higher in Alaska than in Florida at the 5% significance level.

MANTEL-HAENSZEL ESTIMATION

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. You can derive a pooled estimate of the effect from these stratum-specific effect estimates, and you can use the Mantel-Haenszel method to estimate such an effect. For a homogeneous rate difference effect between two populations, the Mantel-Haenszel estimate is identical to the difference between two directly standardized rates, but it uses weights that are derived from the two populations instead of from an explicitly specified reference population.

That is, for population k, k=1 and k=2, the standardized rates are

$$\hat{\lambda}_k = \frac{\sum_j w_j \, \hat{\lambda}_{kj}}{\sum_j w_j}$$

where $\hat{\lambda}_{kj}$ is the rate in the jth stratum of population k and the weights are derived from the two population-times,

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

where $\mathcal{T}_{kj}$ is the population-time in the jth stratum of population k.

The Mantel-Haenszel difference statistic is then given by

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

You can also apply the Mantel-Haenszel method to other homogeneous effects between populations, such as the rate ratio, risk difference, and risk ratio.

EXAMPLE: COMPUTING MANTEL-HAENSZEL RISK ESTIMATION

This example uses the Mantel-Haenszel method to estimate the effect of household smoking on respiratory symptoms of school children, after adjusting for the effects of the students' grades and household pets.

The following School data set contains hypothetical stratum-specific numbers of cases of respiratory symptoms in a given school year for a school district:

```
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 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 students who have respiratory symptoms and the total number of students, respectively.

The following statements invoke PROC STDRATE and compute the Mantel-Haenszel rate difference statistic between students who have smokers in their household and students who do not:

```
ods graphics on;
proc stdrate data=School
             method=mh
             stat=risk
             effect=diff
             plots=effect
             ;
   population group=Smoking event=Case total=Student;
   strata Pet Grade / order=data effect;
run;
ods graphics off;
```

The METHOD=MH option requests the Mantel-Haenszel estimation, and the STAT=RISK option specifies the risk statistic for standardization. When you specify the EFFECT=DIFF option, PROC STDRATE uses the default risk difference statistics to compute the risk effect between the study populations.

The POPULATION statement specifies the options that are related to the study populations. The EVENT= option specifies the variable for the number of cases, the TOTAL= option specifies the number of students, and the GROUP=SMOKING option specifies the variable **Smoking**, which identifies the smoking groups in the DATA= data set.

The STRATA statement names the variables, **Pet** and **Grade**, that form the strata in the standardization. The ORDER=DATA option sorts the strata by order of their appearance in the input data set, and the EFFECT option displays the strata effects.

The "Standardization Information" table in Figure 6 displays the standardization information.

Figure 6 Standardization Information

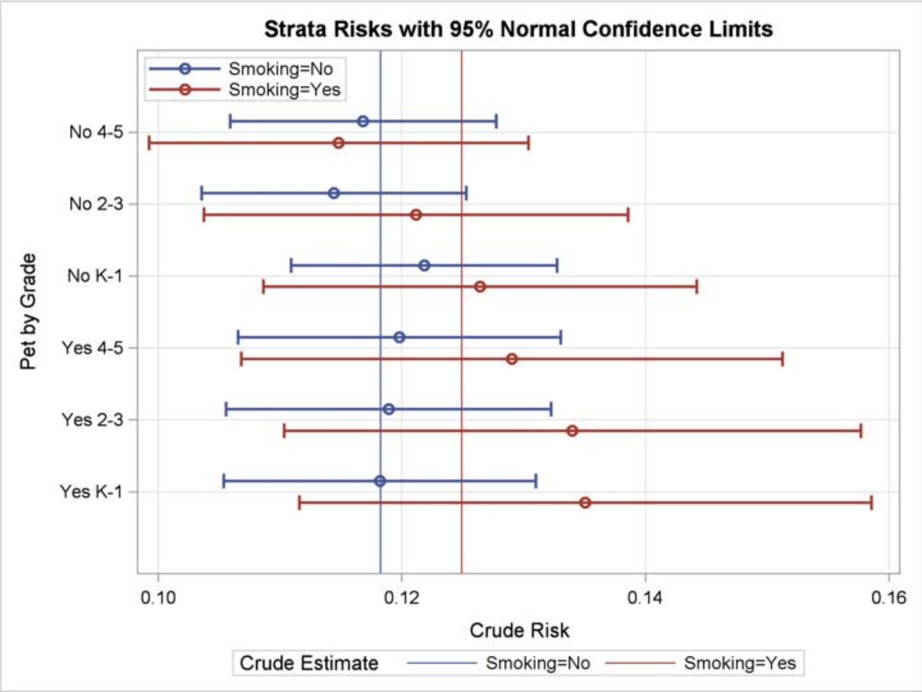
```
                 The STDRATE Procedure

              Standardization Information

        Data Set                        WORK.SCHOOL
        Group Variable                      Smoking
        Method                      Mantel-Haenszel
        Statistic                              Risk
        Number of Strata                          6
```

With ODS Graphics enabled, PROC STDRATE displays the strata risk plot by default. The strata risk plot displays stratum-specific risk estimates and their confidence limits in the study populations, as shown in Figure 7. This plot displays stratum-specific risk estimates and the overall crude risks for the two study populations. By default, strata levels are displayed on the vertical axis.

Figure 7 Strata Risk Plot



When you specify the STAT=RISK option in the PROC STDRATE statement, the EFFECT option in the STRATA statement displays the "Strata Risk Effect Estimates" table, as shown in Figure 8. The EFFECT=DIFF option in the PROC STDRATE statement requests that strata risk differences be displayed.

Figure 8 Strata Risk Effect Estimates

Strata Risk Effect Estimates

Stratum Index	Pet	Grade	Smoking No	Smoking Yes	Risk Difference	Standard Error
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

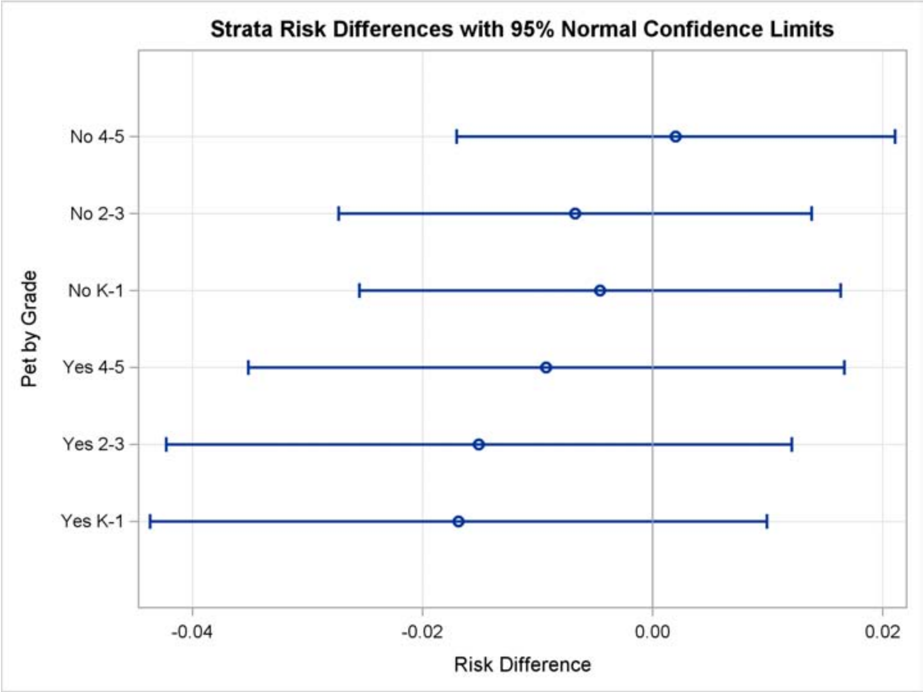
Strata Risk Effect Estimates

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 in Grade 4–5 who have no pets in their households, the risk is higher for students who have no smokers in their households than for students who do have smokers in their households. 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 that includes the stratum-specific risk effect measures and their associated confidence limits, as shown in Figure 9. The EFFECT=DIFF option requests that the risk difference be displayed. By default, confidence limits are generated with a 95% confidence level. This plot displays the stratum-specific risk differences in the "Strata Risk Effect Estimates" table in Figure 8.

Figure 9 Strata Risk Plot



The “Mantel-Haenszel Standardized Risk Estimates” table in Figure 10 displays the Mantel-Haenszel standardized risks and related statistics.

Figure 10 Mantel-Haenszel Standardized Risk Estimates

Mantel-Haenszel Standardized Risk Estimates

	Study Population			Mantel-Haenszel	
Smoking	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

	Standardized Risk			
Smoking	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 Mantel-Haenszel standardized risks, as shown in Figure 11.

Figure 11 Mantel-Haenszel Effect Estimates

Risk Effect Estimates

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

The table in Figure 11 shows that although the standardized risk for students without household smokers is lower than the standardized risk for students with household smokers, the difference (–0.00698) is not significant at the 5% significance level (p-value = 0.1418).

INDIRECT STANDARDIZATION AND SMR

Indirect standardization begins with the computation of SMR (the ratio of the observed number of events to the expected number of events in the study population). For rate statistics, you compute the expected number of events by applying the stratum-specific rate estimates in the reference population to the corresponding population-time in the study population. That is,

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

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

With the expected number of events, $\mathcal{E}$, SMR is

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

where $\mathcal{D}$ is the observed number of events.

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

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

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

You can also compute SMR for the risk statistic similarly.

SMR compares rates or risks in the study and reference populations, and it is applicable even when the study population is so small that the resulting stratum-specific rates are not stable.

EXAMPLE: COMPUTING SMR AND INDIRECTLY STANDARDIZED RATE

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

The following 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):

```
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
;
```

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**.

The following US_C43 data set contains the corresponding stratum-specific mortality information for the United States in 2000 (Miniño et al. 2002; U.S. Bureau of the Census 2011):

```
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 PROC STDRATE and request indirect standardization to compare death rates between 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 plots that are appropriate for indirect standardization.

The POPULATION and REFERENCE statements specify the options that are related to the study and reference populations, respectively. The EVENT= and TOTAL= options specify variables for the number of events and person-years in the populations, 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 crude rates, and the SMR option requests a strata SMR estimates table.

The "Standardization Information" table in Figure 12 displays the standardization information.

Figure 12 Standardization Information

The STDRATE 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 13 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 13 Strata Information (Indirect Standardization)

The STDRATE Procedure

Indirectly Standardized Strata Statistics
Rate Multiplier = 100000

Stratum Index	Age	Study Population: Observed Events	Study Population: Population-Time Value	Study Population: Population-Time Proportion	Study Population: Crude Rate	Study Population: Standard Error
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

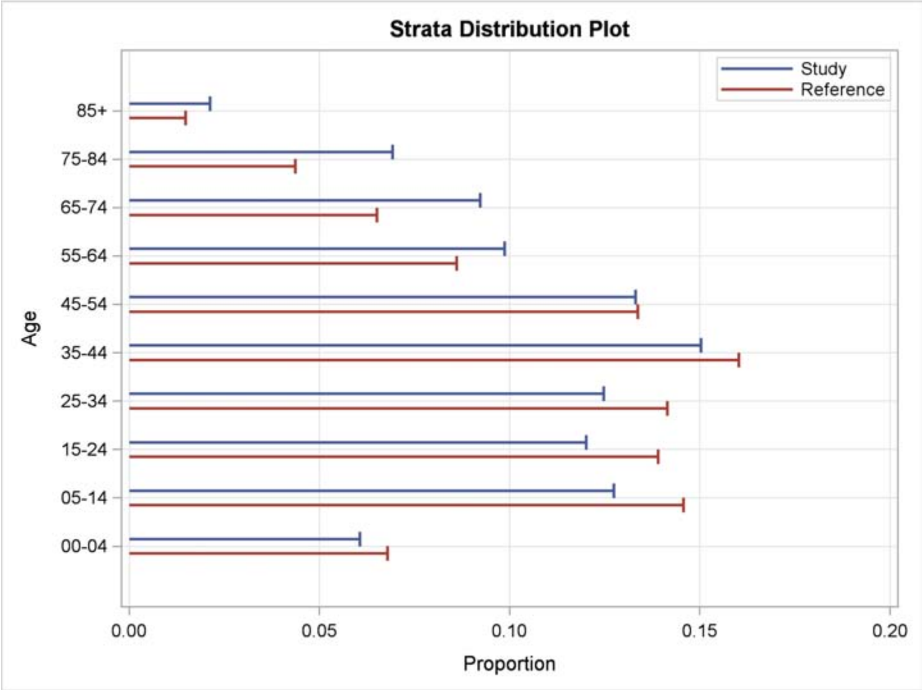
Indirectly Standardized Strata Statistics
Rate Multiplier = 100000

Stratum Index	Study Population: 95% Normal Confidence Limits		Reference Population: Population-Time Value	Reference Population: Population-Time Proportion	Reference Population: Crude Rate	Expected Events
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. When you request 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 14.

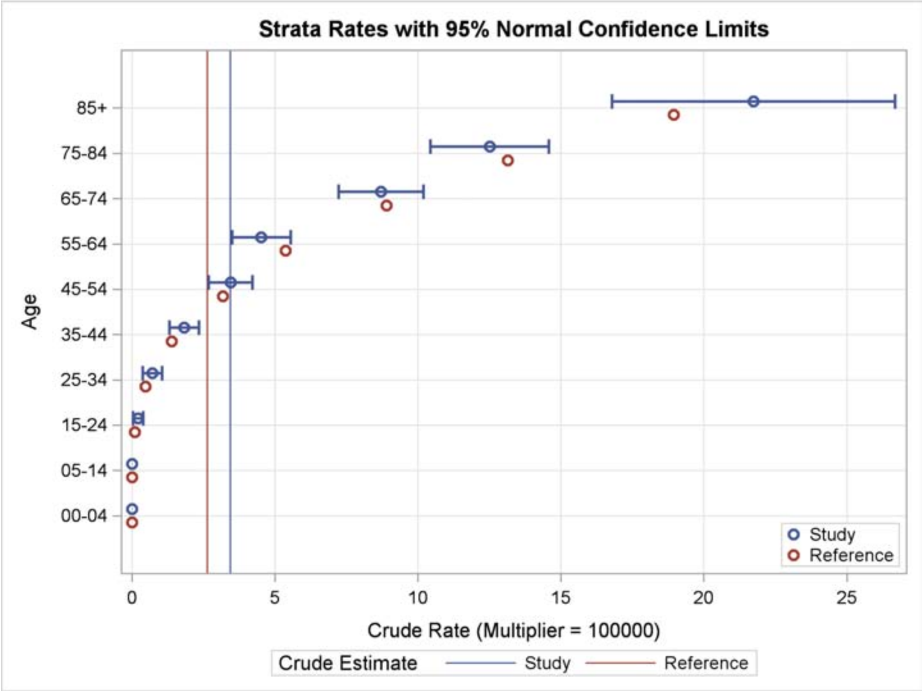
Figure 14 Strata Distribution Plot



The strata distribution plot displays the proportions in the "Indirectly Standardized Strata Statistics" table in Figure 13. In the plot in Figure 14, 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 of skin cancer deaths 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 15. This plot displays the rate estimates in the "Indirectly Standardized Strata Statistics" table in Figure 13. 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 15 Strata Rate Plot



The SMR option in the STRATA statement requests that the “Strata SMR Estimates” table display the strata SMR at each stratum. (See Figure 16.) The MULT=100000 suboption in the STAT=RATE option requests that the reference rates per 100,000 person-years be displayed. The table shows that SMR is less than 1 at three age strata (55–64, 65–74, and 75–84).

Figure 16 Strata SMR Information

Strata SMR Estimates
Rate Multiplier = 100000

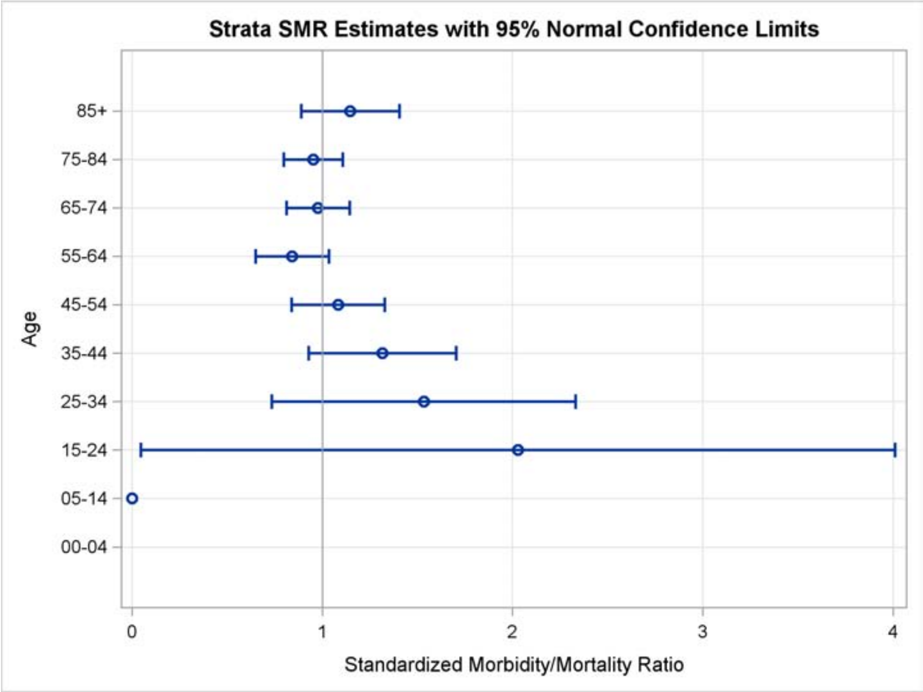
		---Study Population--		Reference			
Stratum Index	Age	Observed Events	Population-Time	Crude Rate	Expected Events	SMR	Standard Error
1	00-04	0	953785	0.0000	0.000	.	.
2	05-14	0	1997935	0.0024	0.049	0.0000	.
3	15-24	4	1885014	0.1046	1.972	2.0280	1.0140
4	25-34	14	1957573	0.4663	9.127	1.5339	0.4099
5	35-44	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

	95% Normal	
Stratum Index	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 plot displays stratum-specific SMR estimates and their confidence limits, as shown in Figure 17. The plot displays the SMR estimates in the “Strata SMR Estimates” table in Figure 16.

Figure 17 Strata SMR Plot



The METHOD=INDIRECT option requests that the "Standardized Morbidity/Mortality Ratio" table be displayed. (See Figure 18.) The table displays the SMR, its confidence limits, and the test for the null hypothesis H_0 : SMR $= 1$. The default ALPHA=0.05 option requests that 95% confidence limits be constructed.

Figure 18 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 SMR=1, and the hypothesis of SMR=1 is not rejected at the α=0.05 level from the normal test.

The “Indirectly Standardized Rate Estimates” table in Figure 19 displays the indirectly standardized rate and related statistics.

Figure 19 Standardized Rate Estimates (Indirect Standardization)

Indirectly Standardized Rate Estimates
Rate Multiplier = 100000

Study Population			Reference		
Observed Events	Population-Time	Crude Rate	Crude Rate	Expected Events	SMR
538	15658227	3.4359	2.6366	528.726	1.0175

Indirectly Standardized Rate Estimates
Rate Multiplier = 100000

Standardized Rate			
Estimate	Standard Error	95% Normal 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 in Figure 19 shows that although the crude rate in the state of Florida (3.4359) is much higher than the crude rate in the United States (2.6366), the resulting standardized rate (2.6829) is close to the crude rate in the United States.

SUMMARY

In comparing the outcome measure of rate or risk between two populations, the use of the overall crude rate or risk might not be appropriate because of confounding. You can derive directly standardized and indirectly standardized rate or risk estimates based on stratum-specific estimates by removing the effects of confounding variables. These estimates provide useful summary statistics and allow valid comparison of the populations.

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

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

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

Florida Department of Health (2000), “Florida Vital Statistics Annual Report 2000,” Accessed February 2012.
URL http://www.flpublichealth.com/VSBOOK/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*, 4th Edition, Sudbury, MA: Jones & Bartlett.

Greenland, S. and Rothman, K. J. (2008), "Introduction to Stratified Analysis," in K. J. Rothman, S. Greenland, and T. L. Lash, eds., *Modern Epidemiology*, 3rd 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

U.S. Bureau of the Census (2011), "Age and Sex Composition: 2010," Accessed February 2012.
URL http://www.census.gov/prod/cen2010/briefs/c2010br-03.pdf

ACKNOWLEDGMENT

The author is grateful to Bob Rodriguez of the SAS Advanced Analytics Division for his valuable assistance in the preparation of this paper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author:

Yang Yuan
SAS Institute Inc.
111 Rockville Pike, Suite 1000
Rockville, MD 20850
301-838-7030
310-838-7410 (Fax)
Yang.Yuan@sas.com