Standardization of Rates and Ratios - Assignment solutions

1. a & b

Population and Deaths in 1980 in Rateboro Adults by Age and Sex and U.S. Total (hypothetical data)

	Rateboro						United States		
	Males			Females			Both Sexes		
Age	Pop.	Deaths	Rate	Pop.	Deaths	Rate	Pop*	Deaths*	Rate
18-34	900	6	.0067	800	1	.0013	60,000	90	.0015
35-59	800	3	.0038	800	5	.0063	45,000	270	.0060
60-74	300	15	.0500	500	10	.0200	20,000	600	.0300
75 +	200	22	.1100	500	38	.0760	15,000	1500	.1000
Total	2200	46	.0209	2600	54	.0208	140,000	2460	.0176

(*In thousands. Population and deaths for Rateboro are actual figures.)

Calculations:

c. Directly standardized death rates for Rateboro males and females (separately) using the U.S. population (both sexes) as a standard.

$$\text{Directly standardized rate} \quad = \quad \frac{\Sigma(r_t N_t)}{N_t}$$

$$\text{Male rate} = \frac{[(.0067 \times 60{,}000) + (.0038 \times 45{,}000) + (.05 \times 20{,}000) + (.11 \times 15{,}000)]}{140{,}000}$$

= 0.0230, or 23 deaths per thousand

$$\text{Female rate} = \frac{[(.0013 \times 60{,}000) + (.0063 \times 45{,}000) + (.02 \times 20{,}000) + (.076 \times 15{,}000)]}{140{,}000}$$

= 0.0136, or 13.6 deaths per thousand

d.

Indirectly standardized rates: $= \dfrac{d_t}{\Sigma(R_i n_i)} R_t$

$$\text{Male rate} = \frac{46}{[(0.0015 \times 900) + (0.006 \times 800) + (0.03 \times 300) + (0.1 \times 200)]} (.0176)$$

= 0.0230, or 23 deaths per thousand

[the similarity to the directly-standardized rate is coincidental.]

$$\text{Female rate} = \frac{54}{[(0.0015 \times 800) + (0.006 \times 800) + (0.03 \times 500) + (0.1 \times 500)]} (.0176)$$

= 0.0134, or 13.4 deaths per thousand

[the similarity to the directly-standardized rate is coincidental.]

2.

a. Females have a more favorable mortality experience. Although the crude death rates for males and females are very close (20.9/1000 vs. 20.8/1000), when age-standardized (direct or indirect) rates are compared, the lower death rates for women are clear.

i. direct: 23 deaths/1000 (men) vs. 13.6 deaths/1000 (women)

ii. indirect: 23 death/1000 (men) vs. 13.4 deaths/1000 (women)

b. The similarity in the crude death rates is a function of the respective age distributions of males and females in Rateboro. A greater proportion of women are found in the older age groups, where the morality rates are higher. The crude death rate gives more weight to these larger strata.

c. i. Reasons for rate adjustment are:

- adjustment procedures attempt to permit valid comparisons by minimizing the effect of extraneous variables (e.g., age) that are differientially distributed across the populations of interest;
- summary indices from two or more populations are more easily compared than multiple strata with specific rates; and
- small numbers in some strata may lead to unstable rates.

ii. Disadvantages of adjustment are:

- information is lost when summary measures are used (opposing trends in subgroups may be masked);
- the absolute and relative magnitudes of the standardized rates will depend on the standard used (i.e., the age groups weighted most heavily); and
- standardized rates are fictitious – they do not estimate any "true" parameter.

ii. Direct vs. indirect methods: indirect methods of adjustment are used when the numbers of deaths in the individual strata are too small to yield meaningful rates. The major disadvantage of indirectly standardized rates is that they can properly be compared only to the crude rate in the standard population (that is, it is technically incorrect to compare the indirectly standardized rates for males to the indirectly standardized rates for females as was shown in 2.a.2 above). Conversely, the major advantage of using direct adjustment is that the standardized rates are comparable to one another if they were based on the same standard weights. However, in several of the strata the numbers of observed deaths are small (e.g., 1,3 , 5 and 6), so the estimates of the rates for those strata are imprecise (likely to be heavily influenced by random error) and therefore weighting them in a direct adjustment procedure is hazardous.

d. Agree with the first part (consistency of Rateboro experience and U.S.) but question the second part (Rateboro environment more suitable for males age 35-59) since the rates cited are based on only 3 male and 5 female deaths and are therefore too imprecise to warrant such a conclusion.

3.

a. Indirect adjustment was used, as age-calendar-year-specific rates from a standard population (Connecticut) were applied to the age-calendar-year distribution (of women-years) in the study population. Here is a detailed explanation:

For the indirect standardization or adjustment procedure, "standard rates" were obtained from the Connecticut population. These rates were both age-specific and calendar-year specific, to control for changes in incidence over time. Thus, a table of standard rates like the following would have been used:

Breast cancer incidence (per 100,000 Connecticut women per year) (hypothetical data)

	Period					
Age	1935-39	1940-44	1945-49	1950-54	1955-59	etc.
30-34	20	22	26	28	30	
35-39	30	33	35	38	40	
40-44	50	54	57	59	62	
45-49	70	72	75	78	81	

Source: Connecticut Cancer Registry (1950-1969)

The second ingredient for an standardized rate is the weight. The weight could be population or population-time (person-years, or in this case, women-years). Boice and Monson tell us that they computed women-years within 5-year age groups and 5-year calendar time intervals (quinquennia) (which is why the above table is constructed as it is). Boice and Monson also divided the follow-up period for each woman into 5- (their lucky number!?) year intervals since the start of observation (sanitarium admission or fluoroscopy exposure) for the women. Dividing up the follow-up period is not part of the adjustment procedure, but enables the investigators to analyze the results for different lengths of follow-up after exposure. Thus the investigators can allow for latency periods in cancer development.

Suppose that the distribution of women-years for all women followed between 11 and 15 years after admission or exposure was:

Distribution of Women-Years (WY) among exposed subjects
between 11 and 15 years (inclusive) following admission or exposure
(hypothetical data)

	Period					
Age	1935-39	1940-44	1945-49	1950-54	1955-59	etc.
30-34	1900	1800	--	--	--	
35-39	1800	1700	1600	--	--	
40-44	1700	1600	1500	1400	--	
45-49	1600	1500	1400	1300	1200	

Source: U.S. Census

With the rates and the weights, the next step is: "Multiplication of the age-calendar year specific WY [women-years] at risk by the corresponding Connecticut incidence rates determined the number of expected breast cancers."

So the expected number of breast cancer cases would be:

$$
\begin{array}{llll}
0.00020 & \times & 1900 & + \\
0.00022 & \times & 1800 & + \\
0.00030 & \times & 1800 & + \\
0.00033 & \times & 1700 & + \\
0.00035 & \times & 1600 & + \\
0.00050 & \times & 1700 & + \\
0.00054 & \times & 1600 & + \\
0.00057 & \times & 1500 & + \\
0.00059 & \times & 1400 & + \\
\text{etc.} & & &
\end{array}
$$

This expected number of breast cancer cases (expected if the women in the study had the same age- and calendar-year-specific breast cancer incidence as women in Connecticut) would be compared to the number of breast cancer cases actually observed.

b. It is not possible to calculate by the method used by Boice and Monson, since their method requires age-calendar-year specific incidence rates whereas the rates given in the question are not specific for calendar year.

c. The advantage of this more complex adjustment procedure is that it controls for secular changes in breast cancer incidence.

4. a. Race-sex-specific and overall TB rates for the three counties:

Incidence of tuberculosis, per 100,000, in three N.C. counties during January 1, 1986 - December 31, 1990

County	White males	White females	Nonwhite males	Nonwhite females	Overall
Johnston	7.0	4.7	124.5	32.2	18.6
Orange	2.9	1.6	8.0	9.1	3.4
Wilson	6.0	9.0	95.4	42.2	28.7

E.g., mean annual TB incidence for nonwhite females in Johnston county =
13 / (8,078 × 5) = 32.2 per 100,000. The 5 in the denominator is needed to obtain the annual incidence, since the numerator contains cases accumulated during 5 years.

Overall annual TB incidence in Johnston county =

75 / (80,664 × 5) = 93 / 5 = 18.6 per 100,000

b. SMR's:

SMRs for tuberculosis in three N.C. counties during January 1, 1986 - December 31, 1990

County	Expected	Observed / Expected	SMR
Johnston	11.7 + 6.1 + 13.5 + 8 = 39.3	75 / 39.3	1.8
Orange	12.7 + 6.8 + 14.8 + 8.7 = 43	15 / 43	0.35
Wilson	7.34 + 4 + 21 + 12.7 = 45	94 / 45	2.1

E.g., overall SMR for Johnston County:

Expected (over 5 years) based on national rates =

Group	US rate /100,000		County pop.		5 years		Expected cases in 5 yrs	
WM	0.000074	×	31,721	×	5	=	11.74	+
WF	0.000036	×	33,955	×	5	=	6.11	+
NM	0.000392	×	6,910	×	5	=	13.54	+
NF	0.000198	×	8,078	×	5	=	8.00	+
							39.39	

SMR = Observed/Expected = 75 / 39.39 ≈ 1.9

Interpretation: Both Johnston and Wilson Counties have higher TB incidence than the U.S. average. The greater TB incidence in these counties is apparently due to the higher rates in nonwhites of both sexes than in the U.S. as a whole. In Johnston County there are 56 cases in nonwhites vs. 21.5 expected; in Wilson County there are 78 cases in nonwhites vs. 33.7 expected. There is also a slight increased incidence in whites in Wilson County: 16 white cases observed vs. 11 expected. Note that the incidence of TB in Johnston County is nearly 18 times as great in nonwhite males compared to white males.

In this case comparison of the SMR's between Johnston and Orange counties is not problematic, since the race-sex population distributions (i.e., the "weights") are similar for the two counties. The population distribution in Wilson County is different, however, so comparing its SMR to the others is indeed problematic.

5. a. Intuitively, we know this assertion to be true, since:

   i. a directly standardized rate is a weighted average of stratum-specific rates;

   ii. the crude rate is a weighted average of stratum-specific rates, weighted in this case by the stratum sizes of the study population;

   iii. a weighted average of identical rates will be equal to the value of those rates, no matter what weights are used.

Using the notation from the Standardization chapter of the *Evolving Text*, with subscript "a" or "b" referring to group A or B, respectively, we have the directly-standardized rate for group A (from the formula under "Standardization of rates by the direct method" and using the information in the problem):

$$\text{Directly standardized rate for A} \quad = \quad \frac{\sum(r_{ai}\ N_i)}{N_t} \quad = \quad \frac{\sum(r_a\ N_i)}{N_t}$$

$$= \quad \frac{r_a \sum N_I}{N_t} \quad = \quad \frac{r_a\ N_t}{N_t} \quad = \quad r_a$$

$$\text{Crude rate for A} = \frac{\sum r_{ai}\, n_{ai}}{n_t} = \frac{r_a \sum n_{ai}}{n_t} = r_a$$

So the directly standardized rate equals the crude rate (equals the stratum-specific rates). The same can be shown, in a identical manner, for B. Therefore the ratio of directly-standardized rates equals the ratio of crude rates.

Moral: if there is no variation in your stratum-specific rates, you don't need to adjust--the crude is fine.

c. This question asks about the situation in which there is a constant rate ratio between groups A and B within each age stratum. Since the SMR is calculated using the rates in the standard population (in this case, r_{bi}) for the denominator (the "expected" deaths), that denominator will be 1/K times the observed deaths, since the rates from the standard population are 1/K times the rates observed in the study population.

Using the formulas on pages 4 and 8:

$$\text{SMR} = \frac{\text{Observed deaths}}{\text{Expected deaths}} = \frac{\sum (r_{ai} n_{ai})}{\sum (r_{bi} n_{ai})} = \frac{\sum (r_{ai} n_{ai})}{\sum \left(\frac{r_{ai}}{K} \right) n_{ai}}$$

$$= \frac{\sum (r_{ai} n_{ai})}{\frac{1}{K} \sum (r_{ai} n_{ai})} = K$$

This exercise illustrates the underlying rationale for the SMR, i.e., in a situation in which there are too few data to make meaningful judgments about specific rates, we assume that each is a constant multiple of the specific rates in a standard population and then estimate that constant multiple with the SMR. The assumption of a constant multiple may not hold in reality, but it may be reasonably correct with study group we are examining. In any case it is the best we can do given the limited amount of data.

d. Intuitively, if two populations are alike in terms of a particular variable, then that variable cannot be responsible for observed differences between them.

<u>Directly standardized rates</u> are comparable, regardless of age distributions, because the specific rates in each population are weighted by the same external standard.

Crude rates are comparable because the crude rate for each group may be thought of as a weighted average of the group's specific rates, with weighting by the proportional size of the strata:

$$r_a = \frac{\text{deaths}}{n_{at}} = \frac{\sum (r_{ai}n_{ai})}{n_{at}} = \sum\left(r_{ai} \frac{n_{ai}}{n_{at}} \right)$$

$$r_b = \frac{\text{deaths}}{n_{bt}} = \frac{\sum (r_b n_b)}{n_{bt}} = \sum\left(r_{bi} \frac{n_{bi}}{n_{bt}} \right)$$

To say that both groups have the same proportional age distribution is to say that for any age stratum (i.e., stratum "i"),

$$r_b = \frac{n_{ai}}{n_a} = \frac{n_i}{n_{bt}} = p_i$$

So $r_a=\Sigma[r_{ai}p_i]$, $r_b=\Sigma[r_{bi}p_i]$, and the two sets of specific rates are averaged using the same weights, p_i.

Indirectly standardized rates:

From the formula at the top of page 4,

$$\text{Indirectly standardized rate} = r_t \times \frac{R_t}{\sum (R_i n_i)/n_t} = \frac{r_t R_t}{\sum\left(R_i \frac{n_I}{n_t}\right)}$$

$$= r_t \frac{R_t}{\sum (R_i p_i)}$$

Since R_t and R_i come from the standard population and p_i is the same for groups A and B (though it may vary from stratum to stratum) by the conditions of the problem, the indirectly standardized rates for A and B are each equal to their crude rates times a the same constant. So a comparison of indirectly standardized rates in this case is the same as a comparison of their crude rates, which was shown above to be valid.

6. An Excel® spreadsheet for this problem can be found on the EPID 168 web site at www.sph.unc.edu\courses\epid168\public\Standardization.xls.