PH 250B Week 3, Tab 4 Practice Problems - ANSWERS
Topic: Measures of Disease

Problem 1. (Fall 2017 250B problem set)

You have received data from two Factories, A and B. You know the number of workers in each age group and the total number of observed deaths at each factory. You do not know who died, so you don't know the age distribution of the deaths. You also have mortality rates from Factory C. Use the mortality rates from Factory C as the standard to do the following calculations.

a. Calculate the expected number of deaths for factories A and B.

Factory A:

Age (yrs)	Population size	Standard (Factory C) Mortality Rate /1,000	Expected deaths	Observed deaths
20-39	3,000	3	3/1000*3000=9	
40-49	6,000	9	54	
50-59	8,000	11	88	
Total	17,000		151	165

Factory B:

Age (yrs)	Population size	Standard (Factory C) Mortality Rate /1,000	Expected deaths	Observed deaths
20-39	6,000	3	3/1000*6000=18	
40-49	6,000	9	54	
50-59	3,000	11	33	
Total	15,000		105	113

Expected deaths for A = 151
Expected deaths for B = 105

b. Calculate the standardized mortality ratio (SMR) for each factory.
SMR for A = Observed deaths / Expected deaths = 165/151 = 1.09 (or 109 or 109%)
SMR for B = 113/105 = 1.08 (or 108 or 108%)

c. Give your interpretation of each SMR. Can you compare the SMRs directly?

Factory A has a 9% higher rate of mortality than Factory C.
Factory B has an 8% higher rate of mortality than Factory C.
There appears to be excess mortality in both Factory A and Factory B compared to Factory C. However, we cannot compare these two SMRs directly because each is adjusted to a different standard (i.e., the populations in factories A and B). Let's clarify this with an example:
Let's take our standard mortality rate from a different factory, Factory D.

FACTORY A

Age (yrs)	Population size	Factory D Standard Mortality Rate /1,000	Expected deaths	Observed deaths
20-39	3,000	1	3	
40-49	6,000	5	30	
50-59	8,000	24	192	
Total	17,000		225	165

FACTORY B

Age (yrs)	Population size	Factory D Standard Mortality Rate /1,000	Expected deaths	Observed deaths
20-39	6,000	1	6	
40-49	6,000	5	30	
50-59	3,000	24	72	
Total	15,000		108	113

Expected Deaths at Factory A = 225
Expected Deaths at Factory B = 108

SMR Factory A = 165/225 = 0.73
SMR Factory B = 113/108 = 1.05
Now, the results differ. Just by using a different standard population's rates, we now see that Factory A has lower than expected numbers of deaths, while Factory B has higher than expected numbers of deaths.

SMRs almost never can be compared directly - the main reason is that the age distribution in each study population (i.e., each factory) is different, and that fundamental problem is not eliminated by indirect standardization. If the two study populations (factories) had the same age distribution, the SMRs would be directly comparable.

Problem 2. (Fall 2017 250B Practice Exam problem set)

You are told to more closely examine 2 mines in eastern China which have recently reported especially high rates of CLD. You are instructed to compare their CLD incidence to that of the general Chinese population.

Age group	Chinese pop. CLD cases/ 10,000 person-years	Population of Coal Mine 1	Observed cases of CLD in Coal Mine 1	Expected cases of CLD	Population of Coal Mine 2	Observed cases of CLD	Expected cases of CLD
16-25	15	100	N/A	0.150	55	N/A	0.0825
26-30	24	220	N/A	0.528	135	N/A	0.324
31-35	32	315	N/A	1.01	130	N/A	0.416
36-40	19	110	N/A	0.209	60	N/A	0.114
Total			10	1.90		12	0.937

a. Calculate the relevant standardization measure for Coal Mines 1 and 2.

SMR1 = Observed/Expected = (10/1.9) * 100 = 526 %
SMR2 = Observed/Expected = (12/0.937) * 100 = 1,280 %

b. Interpret these standardization measures.

Coal mine 1 workers have a 426% higher rate of CLD than the general Chinese population.
Coal mine 2 workers have a 1,180% higher rate of CLD than the general Chinese population.

c. One of your coworkers points out that Chinese coal mine workers are, on average, significantly younger than the general population, and that the current generation of young Chinese adults is far more likely than older Chinese adults to be heavy smokers. He suggests that this may account for your findings related to the CLD rates among the miners compared to the general population. If he's right, would this be an age, period, or cohort effect?

   Cohort effect

d. You therefore decide to calculate the SMRs comparing CLD in your town mines to the overall coal-miner population in China (as opposed to the general Chinese population). The SMRs you derive are 110 (for Coal Mine 1) and 132 (for Coal Mine 2). Why are these SMRs so different from the ones you derived in B?

   Because you are using rates from a different standard population (i.e., the coal mining population rather than the general Chinese population).

Problem 3. (Fall 2017 250B Practice Exam problem set)
In general, why do epidemiologists bother using the SMR? [Circle all that apply]

a. Sometimes we do not have the necessary data to calculate directly standardized rates
b. SMRs are often used in occupational settings where numbers of cases are small and the calculated rates (and their variances) would be unstable.
c. SMRs approximate a counterfactual idea
d. After standardization, SMRs allow two study populations with different age structures to be compared.

   All but the last bullet should be checked.

Problem 4. (Fall 2017 250B Practice Exam problem set)
The investigators of a cohort study of women in the United States were curious about how the age-adjusted in-situ breast cancer incidence rates compare for white women and black women in their study. Below you will find the total study population stratified by age, as well as the incidence per 100,000 person-years for the study population, stratified by age and race.

Age at diagnosis	Standard Population	Observed Population	White Women Incidence/ 100,000 person-years	White Women Expected # of cases	Observed Population	Black Women Incidence/ 100,000 person-years	Black Women Expected # of cases
40-49	18,567	13,027	46.058	8.551	5,540	36.101	6.702
50-59	12,673	7,930	63.052	7.991	4,743	63.251	8.016
60+	18,760	12,461	96.3	18.066	6,099	65.58	12.303
Totals	**50,000**	**33,618**		**34.608**	**16,382**		**27.021**

a) The investigators are interested in comparing incidence of in-situ breast cancer for white women and for black women in their study. Calculate the age-adjusted incidence rate (per 100,000 person-years) for black women and for white women using the overall study population as the standard population.

Age-adjusted annual incidence rate for white women in the study:
(34.608/50,000) x 100,000 = 69.2 (per 100,000 person-years)
Age-adjusted annual incidence for black women in the study:
(27.021/50,000) x 100,000 = 54.0 (per 100,000 person-years)

b) What type of standardization did you just conduct (direct or indirect standardization)?

Direct Standardization

c) What is the counterfactual question you were asking by calculating these standardized measures?
What would the overall incidence of in-situ breast cancer be for black women and for white women had each race had the same age structure as the overall study population?