

Simpson's Paradox

NAME:

Does smoking improve survival?

(source: [www.whfreeman.com/scc/content/eesee](http://www.whfreeman.com/scc/content/eesee))

A survey concerned with thyroid and heart disease was conducted in 1972-74 in a district near Newcastle, UK by Tunbridge et al (1977). A follow-up study of the same subjects was conducted twenty years later by Vanderpump et al (1996). Here we explore data from the survey on the smoking habits of 1314 women who were classified as being a current smoker or as never having smoked at the time of the original survey. Of interest is whether or not they survived until the second survey.

**Table 1: Relationship between smoking habits and 20-year survival**

Survival status	Smoking status	
	Yes	No
Dead	139	230
Alive	443	502
Total	582	732

**Table 2: Twenty-year survival status categorized by age and smoking habits at the time of the original survey**

Survival status	Age group (years)					
	18 - 44		44 - 64		above 64	
	Smoker	Non-smoker	Smoker	Non-smoker	Smoker	Non-smoker
Dead	19	13	78	52	42	165
Alive	269	327	167	147	7	28
Total	288	340	245	199	49	193

1. Using Table 1, find the percentage of smokers who survived to the second study.

2. Using Table 1, find the percentage of non-smokers who survived to the second study.

Notice the percentage of smokers who survived is higher than the percentage of non-smokers who survived. This seems to indicate smoking actually improves the chances of survival. Let's look closer.

3. Using Table 2, find the percentage of smokers who survived to the second study for each individual age group. Do the same for the non-smokers. Record your percentages below in the table provided.

<b>Age group</b>	<b>18 - 44</b>	<b>44 - 64</b>	<b>above 64</b>
<b>Percent of smokers who survived</b>			
<b>Percent of non-smokers who survived</b>			

4. For each age group, compare and discuss the rates of survival for non-smokers and smokers. Why do you think the conclusion (stated at the top of this page) reversed itself? Why would age make a difference?

5. Find the percentage of the smokers that were above 64 as well as the percentage of the non-smokers that were above 64. Do these numbers help to explain the reversal?

Simpson's paradox exemplifies how lurking variables can distort, even reverse, a statistical conclusion. We must always think about lurking variables when we read or perform studies.