(a) Is it possible to use these data to see whether preschool instruction can *cause* a lower rate of criminal arrest for some populations? (b) Is this a retrospective or a prospective study?

5. Violence Begets Violence. It is often argued that victims of violence exhibit more violent behavior toward others. To study this hypothesis a researcher searched court records to find 908 individuals who had been victims of abuse as children (11 years or younger). She then found 667 individuals, with similar demographic characteristics, who had not been abused as children. Based on a search through subsequent years of court records, she was able to determine how many in each of these groups became involved in violent crimes, as shown in the following table. (Data from C. S. Widom, "The Cycle of Violence," *Science* 244 (1989): 160). (a) Is this a randomized experiment or an observational study? (b) Is this a prospective or a retrospective study? (c) Consider the populations of abused victims and controls from which these samples were selected. Let Y be a binary response variable taking the value 1 if an individual became involved in violent crime and 0 if not, in each of these populations. If π_1 and π_2 are the means of Y in the two populations, what is the variance of Y in each population? (d) What is the variance of the sampling distribution of $\hat{\pi}_1 - \hat{\pi}_2$? (e) Let Z be a binary response variable that takes on the value 1 if an individual did not become involved in violent crime and 0 if the individual did (so Z is $1 - Y$). What are the means of Z in the two populations, in terms of π_1 and π_2 (the means of Y)? (f) What are the variances of Z in the two populations, in terms of π_1 and π_2? (g) How does the analysis of the data differ if Z is used as a response variable instead of Y (in terms of estimating a difference in population proportions and estimating a ratio of odds)?

	Involved in a violent crime?	
	Yes	No
Abuse victim	102	806
Control	53	614

6. In Exercise 5, there are many ways of wording an odds ratio statement. For example:

(i) The odds of Yes among victims relative to controls.
(ii) The odds of Yes among controls relative to victims.
(iii) The odds of No among victims relative to controls.
(iv) The odds of No among controls relative to victims.
(v) The odds of victims among Yes relative to No.
(vi) The odds of victims among No relative to Yes.
(vii) The odds of controls among Yes relative to No.
(viii) The odds of controls among No relative to Yes.
(ix) The odds of victims among Yes relative to controls.

(a) Which, if any, of these are the same? (b) Do all of them make sense?

7. Alcohol and Breast Cancer. The following are partial results from a *case–control* study involving a sample of *cases* (women with breast cancer) and a sample of *controls* (demographically similar women without breast cancer). (Data from L. Rosenberg, J. R. Palmer, D. R. Miller, E. A. Clarke, and S. Shapiro, "A Case–Control Study of Alcoholic Beverage Consumption and Breast Cancer," *American Journal of Epidemiology* 131 (1990): 6–14). The women were asked about their drinking habits, from which the following table was constructed. (a) For assessing the risk of drinking on breast cancer, which factor is the response? (b) Is this a retrospective or a prospective study? (c) Is this a randomized experiment or an observational study?

	Breast cancer: Cases	Controls
Fewer than 4 drinks per week	330	658
4 or more drinks per week	204	386

8. Salk Polio Vaccine. The Salk polio vaccine trials of 1954 included a double-blind experiment in which elementary school children of consenting parents were assigned at random to injection with the Salk vaccine or with a placebo. Both treatment and control groups were set at 200,000 because the target disease, infantile paralysis, was uncommon (but greatly feared). (Data from J. M. Tanur *et al.*, *Statistics: A Guide to the Unknown*, San Francisco: Holden-Day, 1972.) (a) Is this a randomized experiment or an observational study? (b) Is this a retrospective or a prospective study? (c) Describe the population to which inferences can be made. (d) Would it make more sense to look at the difference in disease probabilities or the ratio of disease odds here?

	Infantile paralysis victim?: Yes	No
Placebo	142	199,858
Salk polio vaccine	56	199,944

Computational Exercises

9. Heart Disease and Obesity in American Samoan Men. The study described in Section 18.1.1 also included men. The data for men are shown in the accompanying table. (a) Compute (i) the sample proportions of CVD deaths for the obese and nonobese groups; (ii) the standard error for the difference in sample proportions; (iii) a 95% confidence interval for the difference in population proportions. (b) Find a one-sided p-value for the test of equal population proportions (using the standard error already computed). (c) Compute (i) the sample odds of CVD death for the obese and nonobese groups; (ii) the estimated odds ratio; (iii) the standard error of the estimated log odds ratio; (iv) a 95% confidence interval for the odds ratio. (d) Write a concluding sentence that incorporates the confidence interval from part (c).

	CVD death: Yes	No
Obese	22	1,179
Not obese	22	1,409

10. Salk Polio Vaccine. (a) For the data in Exercise 8 find a one-sided p-value for testing the hypothesis that the odds of infantile paralysis are the same for the placebo and for the Salk vaccine treatments. (b) An alternative method for testing equal proportions is to look only at the 198 children who contracted polio. Each of them had a 50% chance of being assigned to the Salk group or to the placebo group, unless group membership altered their chances. Under the null hypothesis of no difference in disease probabilities, the proportion of the 198 who are in either treatment group has a mean of 1/2. Is the observed proportion in the control group, 142/198, significantly larger than