McNemar's Test for Matched Pairs

The methods in Part 1 of this section are based on independent data. For 2×2 tables consisting of frequency counts that result from matched pairs, the frequency counts within each matched pair are not independent and, for such cases, we can use McNemar's test of the null hypothesis that the frequencies from the discordant (different) categories occur in the same proportion.

Table 11-7 shows a general format for summarizing results from data consisting of frequency counts from matched pairs. Table 11-7 refers to two different treatments (such as two different eye drop solutions) applied to two different parts of each subject (such as left eye and right eye). It's a bit difficult to correctly read a table such as Table 11-7. The total number of subjects is $a + b + c + d$, and each of those subjects yields results from each of two parts of a matched pair. If $a = 100$, then 100 subjects were cured with both treatments. If $b = 50$ in Table 11-7, then each of 50 subjects had no cure with treatment X but they were each cured with treatment Y. Remember, the entries in Table 11-7 are frequency counts of subjects, not the total number of individual components in the matched pairs. If 500 people have each eye treated with two different ointments, the value of $a + b + c + d$ is 500 (the number of subjects), not 1000 (the number of treated eyes).

Table 11-7 2×2 Table with Frequency Counts from Matched Pairs

		Treatment X	
		Cured	**Not Cured**
Treatment Y	**Cured**	a	b
	Not Cured	c	d

McNemar's test requires that for a table such as Table 11-7, the frequencies are such that $b + c \geq 10$. The test is a right-tailed chi-square test with the following test statistic:

$$\chi^2 = \frac{(|b - c| - 1)^2}{b + c}$$

P-values are typically provided by software, and critical values can be found in Table A-4 using 1 degree of freedom. *Caution:* When applying McNemar's test, be careful to use only the two frequency counts from *discordant* (different) pairs, such as the frequency b in Table 11-7 (with different pairs of cured/not cured) and frequency c in Table 11-7 (with different pairs of not cured/cured).

Example 6 Are Hip Protectors Effective?

A randomized controlled trial was designed to test the effectiveness of hip protectors in preventing hip fractures in the elderly. Nursing home residents each wore protection on one hip, but not the other. Results are summarized in Table 11-8 (based on data from "Efficacy of Hip Protector to Prevent Hip Fracture in Nursing Home Residents," by Kiel et al., *Journal of the American Medical Association*, Vol. 298, No. 4). McNemar's test can be used to test the null hypothesis that the following two proportions are the same:

- The proportion of subjects with no hip fracture on the protected hip and a hip fracture on the unprotected hip.