larly acute at higher durations, because the confidence interval remains the same after the last observed event, even though fewer and fewer women actually survive to the longest durations because of censoring. We therefore employ the Peto method to produce conservative estimates of 95% confidence intervals.\[1\]

In the analyses of contraceptive failure, method-related discontinuation and resumption of use following discontinuation, we estimated a Cox proportional hazards model separately for each potential correlate to assess whether risks were statistically different across the categories of each factor. The result of each model is an estimate of relative risks—the risk for a particular category relative to the risk for the reference category. For example, in the analyses of method-related discontinuation by age, we estimated risks for age categories relative to the risk at age 20–24.

It is possible that variations in risk across categories of a particular correlate are not causally related to that factor but are observed only because of the confounding effects of other factors. For example, race or ethnicity might appear to have an effect on method-related discontinuation when that factor is examined alone but might not have a significant impact once the effects of income are controlled. It is not feasible, however, to estimate separate life tables for all 23,040 possible combinations of categories for all the factors.

To assess simultaneously the effects of several factors on the risk of contraceptive failure, method-related discontinuation and resumption of use, we used Stata to estimate Cox proportional hazards models. Our goal was to find the simplest models that captured the observed variation in the propensity to experience those outcomes. We started by estimating an initial model with all factors. We next estimated a model that included only the factors with at least one category having a relative risk significantly different from 1.0 at the 5% level. Finally, we combined categories with similar relative risks to produce the simplest model. At each stage, we performed a likelihood ratio test to ensure that the restricted model fit the data as well as the prior unrestricted model.\[1\]

Observations in these analyses were unweighted, for two reasons: We were examining relative risk factors, not estimating absolute levels of risk; and we wanted to use standard model selection procedures based on likelihood ratio tests.\[2\] We employed the same procedure to estimate a final Cox model for resumption of contraceptive use.

Finally, we estimated age-specific contraceptive failure rates to produce a total lifetime contraceptive failure rate—the number of contraceptive failures that the typical woman would experience in a lifetime if she used reversible methods of contraception continuously (except for the time spent pregnant after a contraceptive failure) from exact age 15 to exact age 45.

### Results

**Contraceptive Failure**

Table 1 displays probabilities of contraceptive failure for all reversible methods combined and for 11 separate methods: the implant, the injectable, the IUD, the pill, the diaphragm, the male condom, spermicides, the sponge, withdrawal, periodic abstinence\[3\] and all other methods combined. Overall, 9% of women experi-

<table>
<thead>
<tr>
<th>Method</th>
<th>N</th>
<th>Duration of use</th>
<th>6 months</th>
<th>12 months</th>
<th>18 months</th>
<th>24 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>6,867</td>
<td>5.5 (4.9–6.3)</td>
<td>9.4 (8.3–10.5)</td>
<td>13.4 (11.8–15.1)</td>
<td>16.7 (14.5–19.2)</td>
<td>&lt;br&gt;Implant</td>
</tr>
</tbody>
</table>

### Appendix

1 N women are observed at duration i, then at least N=N/S women must have initiated use, where S is the life-table probability of surviving to duration i. If exactly N/S women did initiate use, then binomial theory yields the standard error of S as sqrt(S)(1–S)/N]. The standard error of Q=1–S is therefore (1–Q)sqrt(Q/N). This estimate will be conservative if, because of censoring, more than N=N/S women initiated use. To produce 95% confidence intervals for Q, we first used the delta method to find the standard error of logit(Q) and then constructed 95% confidence intervals for logit(Q); the antilogits of the upper and lower bounds of the confidence interval for logit(Q) are the upper and lower bounds of the confidence interval for Q.

2 Performing a test after looking at the results is invalid. We used the tests informally simply to achieve a parsimonious description of the data.

3 The argument for using weights is that they will correct for compositional effects. If all factors that govern the weights are included in the model, there will be no compositional bias. In the NSFG, weights partially reflect the oversampling of blacks and Hispanics. We included race and ethnicity in all models and dropped this variable in the final step only if it did not have a significant effect. The disadvantages of using weights are that estimation is less efficient and that standard model selection strategies based on likelihood ratio tests cannot be employed. The estimates in our final models when weights were used were similar to those when they were not used.

4 Of the 250 intervals of use of periodic abstinence, only 33 (13%) were intervals of natural family planning. Reliable separate estimates for that method could not be computed.