Appendix C: Random Sample Selection Methodology

Two random samples were constructed for the 1996 National Youth Gang Survey. The first was a random sample of cities and towns with populations between 2,500 and 25,000 (small cities), and the second was of rural counties.

On previous surveys, the response rate for cities with populations over 150,000 has been as high as 91.3 percent. The response rate for cities with populations between 25,000 and 150,000 was estimated using a random sample of municipalities in that size range that were surveyed as part of the 1994 U.S. Department of Justice, Office of Justice Programs, National Institute of Justice survey conducted by Curry et al. (1996). The response rate of those cities to the 1995 National Youth Gang Survey has been estimated at 74.5 percent. All of the above computations for the proposed sample size assumed a response rate of 75 percent.

The estimated required sample size *n* was derived using the formula:

$$n = \frac{t^2 NPQ}{(d2(N\text{-}1) + t^2 PQ)}$$

Where:

t is the abscissa of the normal curve that cuts off an area of α at the tails.
N is the true population size.
P is the true proportion of the population with a specific characteristic.
Q is the true proportion of the population without a specific characteristic or (*1–P*).
d is an acceptable error of size that can be incurred at probability α.

This computing formula is derived from the formula provided by Cochran's *Sampling Techniques* (1977) for sample size *n* required for producing an error of size *d* at a specific probability α. Cochran uses *t*, the abscissa of the normal curve that cuts off an area of α at the tails to produce the formula:

$$n = \frac{\frac{t^2 PQ}{d^2}}{1 + \frac{1}{N}\left(\frac{t^2 PQ}{d^2} - 1\right)}$$

All the terms in the computing formula are presented in a form equivalent to those in Cochran's formula.

An error rate *d* was computed as 5 percent. The probability α of an estimated error being greater than *d* used in the computations above is .05. All computations are based on an estimated true population of $P = .5$ and $Q = .5$ because this results in the most conservative and largest estimates for required samples for each stratum.