A sample size calculation for the US Population data for the age range of 18-24

On page 446 of the text is some data from the 2000 Census. We will use this data as a population and aim to take a sample to estimate the total population of the US in 2000 in the age range of 18-24. The applicable formula is:

$$n = \frac{N\sigma^2}{(N-1)(B^2/4N^2) + \sigma^2}.$$

In this case we know that $N = 50$, so we need to provide an estimate of σ^2 and a desired bound B for the final estimate. If we do not have any data on σ^2 from prior studies or the literature, we can use the conservative approximation that:

$$\sigma \approx \frac{\text{range of the observation}}{4}.$$

Suppose that we specify that $B = 3,000,000$ and use the range of population 18-24 in US states $= 1,900,000$ (from $100,000$ to $2,000,000$ - is this accurate?). Then we have:

$$\sigma \approx \frac{1,900,000}{4} = 475,000, \quad B^2/4N^2 = \frac{3,000,000^2}{4(50^2)} = 900,000,000$$

and

$$n = \frac{50(475,000^2)}{(50-1)(900,000,000 + (475,000^2))} = 41.8.$$

With a population of $N = 50$, this sampling fraction is too large, so what went wrong? First, we have underestimated the range of populations in the 18-24 age range, as California has over 3 million people in this range. More importantly, our bound B is too small for this total population in the 18-24 age range. You can check that if we change the bound and/or range to the following values, we get these new sample sizes:

Bound	Range	Sample size
6,000,000	1,900,000	28.1
6,000,000	3,550,000	40.9
8,000,000	1,900,000	20.9
8,000,000	3,550,000	35.8
10,000,000	1,900,000	15.8
10,000,000	3,550,000	30.8