MA 386 - Statistical Programming
Portfolio: Simulations

Background:
In 2016, the Centers for Disease Control and Prevention (CDC) advised pregnant women to avoid traveling to areas where the Zika virus was spreading. The Zika virus, spread through fluid transmissions as well as mosquitoes, can lead to serious birth defects. Most notably, there is an increased risk of microcephaly. One problem with Zika is that patients infected with the virus may not present with any symptoms. This makes the virus more difficult to detect.

Blood banks were notably concerned with detecting Zika in received donations in order to prevent the spread of the disease through blood transfusions. In the summer of 2016, Roche developed a means of testing blood samples that detects the Zika virus genetic matter (see this link for more information). The diagnostic was tested in Puerto Rico. According to recent data, it is believed that 1.1% of donations are contaminated with the Zika virus. The new diagnostic is to be tested on $N = 12,800$ total donations.

The testing can be done in two ways:

(A) Every donation can be tested separately (N individual tests performed).

(B) The blood donations of k people are pooled to be analyzed. We will assume that $N = nk$ with n an integer. If the test is negative, each of the donations tested must be negative for the Zika virus (thus requring only a single test on this group). If the test result is positive, however, each of the individual k donations contributing to the group must be tested separately (resulting in a total of $k + 1$ tests for that particular group).

Suppose a member of the research team at Roche is debating the two methods. She believes the second method will reduce costs by reducing the number of tests performed. But, she is unsure of the cost savings.

Assignment:
Write a statement to the research team addressing the following questions:

- What value of k will minimize the expected number of tests required under strategy (B)?

- Using the k chosen, on average, how many tests does strategy (B) save in comparison with (A)?
- Is there a risk-cost trade-off with changing values of k? That is, while on average the k chosen above may minimize the expected number of tests, is the number of tests that will be required extremely variable or fairly certain (and how does this depend on the choice of k)?