Randomized response in sampling

Often it is difficult to obtain honest answers from sample subjects to questions like "Have you ever used heroin" or "Have you ever cheated on an exam". To reduce bias the method or *randomized response* is used. The sample subject is given one of the two statements below at random:

(1) "I have property A.'
(2) "I do not have property A.'

The subject responds YES or NO to the given question. The pollster does not know to which of the two statements the subject is responding. We assume:

- The subjects are a simple random sample of size n from a larger population of size N.
- The statements are assigned to the chosen subjects independently.
- The assignment of statements is independent of the sampling procedure.
- The subjects respond honestly to the statements they are given.

Let

- p be the probability the a subject will be assigned the statement (1). This probability is known and is part of the design.
- q be the proportion of subjects in the population with property A.
- r be the probability that a randomly selected subject responds YES to the statement assigned.
- R be the proportion of subjects in the sample who respond YES.

a. Justify that the probability that a randomly selected subject in the population responds YES to the statement assigned is equal for all subjects. Express this probability with p and q. Show that R is an unbiased estimate of r. Take into account that the assignment of statements is independent of the selection procedure.

b. Suggest an unbiased estimator of q. When is this possible? Express the variance of the estimator with $\text{var}(R)$.

c. Let N_A be the random number of sample subjects with property A, and let N_Y be the random number of sample subjects who respond YES. Compute $E(N_Y|N_A = n_A)$ and $\text{var}(N_Y|N_A = n_A)$.

d. Compute $\text{var}(R)$. Give the standard error for the unbiased estimate of q.

e. Simulate sampling with randomized response from a hypothetical population. Check empirically that the sampling distribution is approximately normal. Compute the empirical standard error and compare it to the theoretical result.