Below are questions and topics covered in reading and lectures up to this point. If you have difficulty answering any of the questions, you should review your notes and readings, ask questions in class or see us at office hours. There will be additional study review sheets in the future.

1. How is incidence time defined? What is the difference between the induction and latent periods. Do they both pertain to the exposure-disease association? Why are these concepts important to study design?

2. What distinguishes open, closed, and stationary populations?
   a. What is a study design associated with an open population?
   b. What is a study design associated with a closed population?
   c. What is a study design associated with a stationary population?

3. How is person-time defined for people of different disease status in the population?

4. Five different measures of disease frequency are incidence rate, incidence proportion, incidence odds, survival proportion, and prevalence. Explain:
   a. Which of these are likely to be used in a study with a definable exposure period?
   b. Which of these may be used when there are a number of variables to be taken into account when assessing the effect of exposure.
   c. Which of these may be used when the onset time for the disease is unknown or difficult to define?
   d. Which of these may be used when dealing with mortality as an outcome?

5. What are the circumstances that would lead an investigator to use a prevalence rather than an incidence measure of disease? When may the risks associated with a
prevalence not be associated with the incidence?

6. What is a counterfactual? Relate the counterfactual to the statistical process of “controlling” or partialling the effect of covariates.

7. R & G distinguish between an “effect” and an “association”. What is the difference? In epidemiology, when may a study relationship be considered an estimate of an effect?

8. Which measures of effect are unit-free?

9. What is the interpretation of the following effect estimates: risk ratio (RR), risk difference (RD), rate ratio (IR), rate difference (ID), odds ratio (OR)?

10. Under what conditions will the OR approximate the RR?

11. Under what conditions will the IR approximate the RR?

12. When exposure affects average risk, under ordinary conditions, how would the RR, IR, and OR be ordered in relation to the null value?

13. What are attributable fractions?

14. What is the distinction between excess fraction, etiologic fraction, and rate fraction?

15. Why is the etiologic fraction not generally estimable from epidemiologic data?

16. What are the two basic types of error in estimation?

17. What are the major ways of maximizing precision?

18. How do different epidemiological designs relate to likely internal validity?
19. Define the three major sources of bias in epidemiological studies: selection bias, confounding, and information bias. Give an example of each.

19. What is meant by nondifferential and differential misclassification? What is the most common direction of bias when there is nondifferential misclassification?

20. Does differential misclassification affect absolute and relative measures of effect equally in a given study?

21. Are problems of sensitivity or problems of specificity in measures of exposure likely to have greater effects on bias of relative effect measures?

22. How does categorization of measures relate to measurement error? If case and control groups differ on a stratifying or risk measure will categorization affect the estimation of effect? If so, how?

23. How does nondifferential categorization of risk measures affect the precision of the estimated effect? How does it affect the magnitude of the estimated effect?

24. R & G state that “the essence of scientific generalization is the formulation of abstract concepts relating the study factors” (p 118). How does this view differ from a view of external validity as reflecting the generalizability of study findings to a larger population than the source population for the study?

25. What is statistical estimation and why is it preferable to statistical significance testing?

26. What are point and interval estimates?

27. What are common circumstances in which a research group may use a test of statistical significance to make a decision?

28. What are the hallmarks of a confounder?

29. How does the R & G identification of a confounder relate to the counterfactual
understanding of the scientific problem?

30. There are three major ways of minimizing the influence of a confounder: by sample selection, by stratification, and by statistical analysis. Define and illustrate each.

31. Is there any reason to control a variable that is related to the disease outcome but not to the exposure being assessed in a study?

32. Is the effect measure generated by multiple linear regression an absolute or a relative measure?

33. R & G state strongly that correlations and other standardized coefficients are not effect measures. Is there any role for such coefficients in epidemiological studies?

34. Describe verbally what it means to “statistically control” one variable when estimating the effect of another variable on the outcome (dependent variable).

35. “Specification error” may be defined as error in the model that you are estimating. Describe how early examination of the data may bear on the identification of specification error. How may stratification be used to identify potential specification error?