PubHealthStats Midterm, S03

Part A

1. It would have been much more useful to describe the distribution of survival times or at least some measure of survival time variability. *Or*, one could address this question by addressing how survival time were measured and what factors might bias these measurements (e.g., stage of disease at diagnosis).
2. Correct answer is "b." When cases are classified as non-cases or non-cases are classified as cases, the *information* on which the study is based is biased.
3. Correct answer is "a." Small samples result in imprecision. Imprecision is synonymous with "lots of random error," and expresses itself in the form of a wide confidence interval.
4. Correct answer is "d." Confounding is the type of bias caused by the effects of extraneous factors.
5. Randomization encourages group comparability, and thus eliminates (or at least tends to reduce) the type of bias known as confounding. (Some students confused "randomization" with "random selection." These are two distinct concepts. Randomization has to do with assigned an intervention to two or more groups through chance mechanisms. Random selection addresses the selection of subjects from a population through chance mechanisms.
6. Blinding prevents data collectors from imposing their "viewpoint" (conscious or unconscious) on the information they are collecting. (Comment: I try to grade your papers blindly.)
7. The key distinction is that longitudinal studies reconstruct sequence of events in *individuals* over time, regardless of when data are actually collected.
8. Prevalence studies are cross-sectional.
9. Processing error occurs anytime *after* data have been collected.
10. Apply range checks, pre-codes, or consistency checks during data entry. You may also conduct a double entry and validation.
11. Back it up!
12. The main purpose of the .qes is file is to define the data set, which means it is used to create variable names and variable types.
13. .rec files
14. The .sps file contains *instructions* ("syntax") that allow SPSS to *import* the flat data in the flat (.txt) file into SPSS.
15. A *RR* of 1.5 suggests a 50% increase in risk with exposure.
16. A *RR* of 1.5 suggests the exposed group has a risk that is 1.5 times that of the non-exposed group.
17. False. Broad confidence intervals tell you nothing about bias (systematic error). They do, however, quantify the imprecision (random error) of the estimate.
18. True. It is unethical to withhold an effective treatment when experimenting in humans.
19. True. The key term here is "generally accepted benchmarks."
20. False. The flexible method of significance testing avoids arbitrary cut points such as $\alpha = .05$. Instead, the *p* value is used as a continuous inferential statistic. Why should $p = .07$ be discounted when $p = .05$ be so significant? (This is rhetorical, of course.)
21. True. Prevalence ratios are often loosely called "relative risks." We hope that the users of such statistics are cognizant of their assumptions when using these statistics.

Part B

22. 1 pt for the correct shaded region located beyond 4.71, 1 pt for locating the statistic, 1 pt for locating the distributional landmark of 3.84, and one point for the correct p value ($.025 < p < .05$)
23. <A>
24. <mm/dd/yyyy>
25. Because this question was poorly worded, no deductions were applied.
26. A code book or "file structure" documentation file.
27. disease variable
28. $\text{phat}_1 = 9 / 62 = .1452$
29. phat = 10 / 166 = .0602
30. RRhat = .1452 / .0602 = 2.41196
31. One point each: $\ln RR\text{^} = 0.8804$; $se_{\ln RR\text{^}} = 0.4347$; LCL for the $\ln RR = .0284$; UCL for the $\ln RR = 1.7324$
32. The question was graded on a 4 point scale with A = 4 pts, B = 3 pts, and so on.
33. One-half point for each correct expected value

		+	-
	+	5.17	56.83
	-	13.83	152.17

One-half point for each part of:
$\chi^2_Y = 2.14 + 0.20 + 0.80 + 0.07$