S10 Ex1

HS 267 Exam—March 11, 2010
Coverage: L1, L2, L3, L4AB

TWO-SIDED COPY

Name: KEY

Allowed: SPSS, WinPEPI, calculator, Procedure Notebook (*de novo* contents only; NOT an open notes test), Formulas and Tables pamphlet.

Instructions: *Write neatly*. Be accurate and concise. In narrative response, please be brief.

Time limit: ~~1.25 hours~~ 2 HOURS

There are 4 questions with multiple parts. Each question is worth 10 points; 40 points total.

The remainder of the page is left intentionally blank. It may be used as scrap paper.

HS 267 Exam—March 11, 2010
Coverage: L1, L2, L3, L4AB

1. Risk factor X. A study finds 12 of 269 individuals self-report the occurrence of risk factor X.

(a) Estimate the **prevalence** of the risk factor in the population. Include a **95% confidence interval** for the prevalence. *Show your calculations.*

$\hat{p} = \frac{12}{269} = .0446$

$\tilde{p} = .05128 \quad \tilde{q} = .94872$

$SE_{\tilde{p}} = \sqrt{\frac{.05128 \cdot .94872}{273}} = .01335$

95% CI FOR p = $.05128 \pm 1.96 \cdot .01335$

$= .05128 \pm .02616$

$= (.0251, .0774)$

(b) What is the **margin of error** in the above 95% confidence interval? (Additional calculations not required.)

.026

(c) Use WinPEPI to determine the sample size that would be required to bring the **margin of error down to ±.01**. Use the sample proportion calculated in Part (a) as a reasonable estimate for the expected proportion.

1637 IF $p = .0446$

(d) This study was based on volunteers. Of the three main types of systematic errors in public health research, what type of bias is most typically associated with the use of volunteers? (☑ check)

☐ information bias ☑ selection bias ☐ confounding ☐ random error

(e) This study was based on self-reported data. Of the three main types of systematic errors, which is typically associated with self-reporting?

☑ information bias ☐ selection bias ☐ confounding ☐ random error

(f) Research from a previous time period established a population prevalence of 8%. *Use the confidence interval calculated in part* (a) to state **whether the prevalence has changed significantly**. **Explain your reasoning.** (Do *not* perform a formal hypothesis test; do *not* report a *P* value.)

YES. PREVALENCE DECLINED SIGNIFICANTLY

BASED ON UPPER CONF. LIMIT OF 7.7%

/10

2. Naturalistic sample. A study published by Jolsen et al. forms the basis of this analysis. Data are from a naturalistic sample of patients undergoing chemotherapy. Download the data set http://www.sjsu.edu/faculty/gerstman/datasets/toxic.sav from the course calendar (right-click > Save As). Open the data file in SPSS.

(a) Let us address the exposure of generic drug use (variable GENERIC: 1 = generic drug; 2 = innovator drug) on the occurrence of toxic reactions (variable TOX: 1 = yes, 2 = no). **Cross-tabulate** the data in SPSS and place the counts in this table:

	Tox+	Tox-	Total
Generic+	11 ✓	14	25 ✓
Generic–	3 ✓	31	34 ✓
Total	14	45	59 ✓

(b) Calculate the **incidence of toxicity in the Generic+ group**. Show work.

✓ $\hat{p}_1 = \frac{11}{25} = .4400$

(c) Calculate the **incidence of toxicity in the Generic– group**. Show work.

✓ $\hat{p}_2 = \frac{3}{34} = .08824$

(d) Calculate the **relative risk** associated with generic drug use. Show work.

✓ $\widehat{RR} = \frac{.4400}{.08824} = 4.987$

(e) **Interpret** the relative risk.

✓ GENERIC GROUP HAD 5 TIMES THE RISK

(f) Use SPSS of WinPEPI to calculate the **95% CI for the RR**.

✓ 95% CI FOR RR = (1.55, 16.03)

/10

3. Case-control sample. Data from a case-control study on the relationship between male-pattern baldness and cardiovascular disease is presented. Cases were men less than 55 years of age who were hospitalized for an acute MI (heart attack). Controls were men in the same age range admitted to the same hospital for non-cardiac conditions. Baldness was an ordinal variable graded 1 for no baldness, 2 for moderate baldness, and 3 for extreme baldness. Data were:

Baldness	Cases	Control
1	251	331
2	195	185
3	52	35

(a) Using the no baldness group as the reference group, calculate the **odds ratio associated with moderate baldness**. Show work.

$$\widehat{OR}_2 = \frac{195 \cdot 331}{185 \cdot 251} = 1.39$$

(b) Again, using the no baldness group as the reference group, calculate the **odds ratio associated with severe baldness**. Show work.

$$\widehat{OR}_3 = \frac{52 \cdot 331}{35 \cdot 251} = 1.96$$

(c) Is there evidence of a **dose-response relationship? How so?**

YES. AS BALDNESS GOES UP, RR GOES UP.

(d) Using WinPEPI, compute the ***P*-value for H_0: "no trend in population."**

P = .00055 (WINPEPI REPORTS P = 5.5E-4)

(e) **Interpret** the above result.

SIGNIFICANT TREND

(f) Fill in the blank with the **best possible** term: AGE is an independent risk factor for HEART ATTACKS. BALDNESS is associated with AGE. Therefore, AGE

CONFOUND the association between BALDNESS and HEART ATTACKS.

4 FREE POINTS

10

4. Miscellaneous questions that require little or no calculations:

(a) This is the name of the probability model that is needed when making inferences about proportions when the Normal approximation does *not* hold.

BINOMIAL

(b) The plus-four confidence interval method for proportions is most similar to this method reported by WinPEPI.

WILSON'S

(c) Confidence intervals are a good way to deal with (check all that apply):

☐ information bias ☐ selection bias ☐ confounding ☑ random error

(d) *P*-values are a good way to deal with (check all that apply):

☐ information bias ☐ selection bias ☐ confounding ☑ random error

(e) A case-control study on female breast cancer and family history of breast cancer reports an odds ratio of 1.92 (95% CI: 1.32 - 2.96). *In one sentence, using absolutely no statistical jargon*, summarize the result of this study. Imagine you are reporting for a lay audience. Therefore, the words "significant" and "association" should *not* appear in your response.

BREAST CA ABOUT TWICE AS FREQUENT IN WOMEN WITH FAMILY HISTORY

(f) A study on mortality in a treatment and control group reports a relative risk of 0.91 (95% CI: 0.82 - 1.02, P = 0.12). In one sentence, *using absolutely no statistical jargon*, summarize this result. See instructions for part (e).

MORTALITY SIMILAR IN THE TWO GROUPS.

/10

(f) A Mantel test for trend produces a chi-square statistic of 11.922 [DF = 1]. What is the value of the equivalent z statistic?

$z = \sqrt{11.922} = 3.45$

(g) List the factors that determine the sample size requirements of a cohort study. All 5 factors must be listed to receive credit.

α

$1-\beta$

RATIO n_2/n_1

p_2

p_1 OR RR

(h) Show why a Fisher's test is required to test the association in this table:

	D+	D-	Total
E+	50	34	84
E−	5	5	10
Total	55	39	94

EXPECTED

	D+	D−
E+	49.149	34.851
E−	5.841	4.149

After turning in the exam, please check the calendar for links to a brief recording and for the week's assignment.

10

Ex 2 S10

HS 267 Exam 2 — April, 2010 **Coverage: L2, L3, L4, L5**

Name: KEY

Allowed: SPSS, WinPEPI, calculator, Procedure Notebook (*de novo* contents), Formulas and Tables pamphlet. **Time limit**: 1.25 hours.

Please write neatly.

1. Case-control study: lumbar disk disease [15 pts]. Data[1] from a case-control study on sports participation and lumbar disk disease are:

Crude Data

	Cases	Controls
No sports participation	67	48
Sports participation	220	239

(a) Calculate the crude odds ratio of lumbar disk disease associated with "no sport participation" using these data. Comment *briefly* on this finding.[2] [2 pts]

* OR^ = (67 × 239) / (48 × 220) = 1.52
* Lack of sports participation is associated with a 50% increase in risk.

Data stratified according to smoking status are:

Smokers

	Cases	Controls
No sports participation	~~67~~ 36	~~48~~ 28
Sports participation	~~220~~ 138	~~239~~ 113

Non-smokers

	Cases	Controls
No sports participation	31	20
Sports participation	82	126

[1] Mundt, D. J., Kelsey, J. L., Golden, A. L., Panjabi, M. M., Pastides, H., Berg, A. T., et al. (1993). An epidemiologic study of sports and weight lifting as possible risk factors for herniated lumbar and cervical discs. The Northeast Collaborative Group on Low Back Pain. *Am J Sports Med, 21*(6), 854-860.
[2] Summarize the relationship simply. Do *not* provide unnecessary calculations or information.

2

(b) Calculate the odds ratio for no sports participation and lumbar disk disease within smokers. Include a brief interpretation. [2 pts]

* OR^ = 1.05
* Sports participation has little or no observable effect.

(c) Calculate the odds ratio within non-smokers. Again, include a brief interpretation. [2]

* OR^ = 2.38
* No sports participation more than doubles in risk of lumbar disk disease.

(d) Based on your findings so far, is there evidence of (1) confounding, (2) interaction, or (3) neither confounding and interaction. Explain your response. [2 pts]

* The strata-specific odds ratios are heterogeneous.
* This is evidence of interaction.

(e) Conduct a formal test for interaction. State H_0. Use WinPEPI to calculate the test statistic, its *df*, and *P*-value. Interpret the results. [3 pts]

* H_0: $OR_1 = OR_2$ ("no interaction" OK)
* Chi-sq = 3.69, df = 1, $P = .055$
* The test confirms the "significant" interaction.

(f) Concisely summarize the results of the study for a professional audience.[3] Limit your response to two grammatically correct sentences. [4 pts]

A B C D *** Approach one ("just the facts"): The odds ratio associated with no sports activity is 1.05 for smokers and is 2.38 for non-smokers (*P* for interaction = .055).

Approach two ("narrative"): There is no effect of sports activity on the risk of herniated disks in smokers. In non -smokers, lack of sports activity approximately doubles the risk.

Approach three: "Any other grammatical and logic statement that represents the facts. **

[3] Report only what is most meaningful. Apply judgment.

/13

2. Case-control: oral cancer [15 pts]. Two-by-two data from a case-control study on smoking and oral cancer[4] are:

Crude Data

	Cases	Controls
Smoke +	233	184
Smoke −	9	32

(a) Calculate the crude odds ratio. Comment briefly on this finding.[5] [2 pts]

* OR^ = 4.50
* Large (four-and-half fold) elevation in risk in smokers.

Data stratified according to alcohol use are:

Alcohol Users

	Cases	Controls
Smoke +	225	166
Smoke −	6	12

Non-users of alcohol

	Cases	Controls
Smoke +	8	18
Smoke −	3	20

(b) Calculate the strata-specific odds ratios. [2 pts]

* $OR\hat{}_1 = (225 \times 12) / (166 \times 6) = 2.71$

* $OR\hat{}_2 = (8 \times 20) / (18 \times 3) = 2.96$

[4] Rothman, K., & Keller, A. (1972). The effect of joint exposure to alcohol and tobacco on risk of cancer of the mouth and pharynx. *J Chronic Dis, 25*(12), 711-716.

[5] Do not waste time on unnecessary speculation.

(c) Assess the evidence for interaction. Explain your assessment.[6] [2 pts]

* No good evidence of interaction.
* The odds ratios are relatively homogeneous.

P FOR INTERACTION = .922

(d) Assess the evidence for confounding. Explain your assessment. [2 pts]

* There is good evidence of confounding.
* The crude odds ratio is much larger (4.5) than the strata-specific odds ratios (2.7 and 3.0).

(e) Use WinPEPI to calculate the Mantel-Haenszel adjusted odds ratio and 95% CI. [2]

* $OR\hat{}_{MH} = 2.79$
* (95% confidence interval = 1.22 to 6.38)

(f) Concisely summarize the results for a professional public health audience.[7] [4]

ABCD

Approach one ("just the facts"): The odds ratio of oral cancer associated with smoking adjusted for alcohol consumption is 2.79 (95% CI: 1.22 to 6.38).

Approach two ("narrative"): Smokers have almost three times the risk of oral cancer.

Approach three: Any other grammatical and logical statement that accurately reflects the pertinent facts.

[6] No formal test is requested, although you can perform a test if time permits.
[7] Report only what is most meaningful.

/10

3. Short answer [10 pts]

(a) A case-control study reports an odds ratio of 2.0 (95% CI: 0.25 to 7.28; $P = .45$). In one sentence, using no statistical jargon, summarize the results of the study for a lay audience. Avoid use of the term "significant;" lay audiences do not understand this term.

* This study is too imprecise (or "too small") to make firm conclusions.
Or you can simply say "we cannot make conclusions from this study."

(b) In a study on gender (exposure variable SEX) and obesity (disease variable OBESITY), *explain* how the extraneous variable socio-economic status (SES) could confound the exposure-disease relationship by applying the three properties of confounding learned in class. [3 pts]

* SEX associated with SES
* SES an independent risk factor for OBESITY
* SES not an intermediary in the causal pathway for OBESITY

(c) Why do we avoid Mantel-Haenszel methods when interaction is present?

* Because it would hide the non-uniformity of the effect

(e) Which of the following can be used to control for confounding? Check all that apply.

☑ randomization ☐ increasing the sample size ☑ matching ☑ stratification

* The second choice should NOT be checked

/6

(f) What measures of association are applicable to case-control studies?

* Only the odds ratio

(g) This is the incidence in the exposed group divided by the incidence in the nonexposed group.

* The relative risk

(h) This is the name of the chi-square or z test used with matched-pair binary data.

* McNemar's

(i) This is the name of the method~~s~~ used to summarize data from stratified tables.

* Mantel-Haenszel

Class resumes at 4:55 pm.

4

OPTIONAL TEST

****KEY** HS 267 Exam 2 Addendum — May 12, 2010 Coverage: L2, L3, L4, L5**

Name: KEY 30 OBJECTIVE POINTS

Instructions: Use of the computer is encouraged except when a problem states "show work," in which case you must show the hand calculation. "Use a sharp pencil." This is euphemism for "be accurate." *Accuracy counts.* Use four significant digits on intermediate calculations; round final results appropriately. Do not pursue calculations that are not requested. Keep your answers brief, and do *not* make anything up.
Allowed: SPSS, WinPEPI, hand-calculator, Procedure Notebook (*de novo*), Formulas and Tables. **Time limit**: 1.25 hours (1:30 – 2:45 on 5/13/10).

Please write neatly.

(1) Case-control study of ZDV and HIV following needle sticks. Healthcare workers will occasionally be exposed to HIV through needle sticks and other accidental punctures with sharp objects. A case-control study in the United Kingdom was conducted to determine the effects of administering zidovudine (ZDV) to healthcare workers following such accidental exposure. [5 pts]

Cross-tabulated data are:

	Cases (HIV+)	Controls (HIV−)
ZDV+	9	247
ZDV−	22	432

(a) Calculate the odds ratio of ZDV for HIV. Show work.

✓ **OR = (9* 432) / (22 * 247) = 0.72**

(b) Does ZDV appear to be effective in preventing HIV following needle stick injuries? If so, *to what extent*?

✓ **Yes. The risk of HIV is decreased by ~ 28% with use of ZDV.**

(d) *If* you needed to judge the precision of the estimate, what statistic would you use?

✓ **A (1 – α)100% confidence interval for the odds ratio.**

(e) *If* you were going to test the odds ratio for statistical significance, what null hypothesis would you use?

✓ H_0: $OR = 1$
✓ **+ 1 free point**

/5

(2) Cohort study of back belts for preventing back injuries. Wassell and coworkers (JAMA. 2000 Dec 6;284(21):2727-32) conducted a cohort study to determine the effectiveness of wearing back belts to prevent disabling low-back injuries at work. From April 1996 through April 1998, they interviewed 2939 material handlers who wore a back belt every work day. Within this group, 502 reported back injuries. They also interviewed 2601 material handling employees who chose not to wear a back belt. Among these employees, 455 reported back injuries. [10 pts]

(a) Create a 2-by-2 table for the findings.

	Back Injury Occurrence	No Back Injury	
Use of Back Belts	**502**	**2437**	**2939**
No Use of Back Belts	**455**	**2146**	**2601**

(b) What was the risk of back injury among the workers who used back belts?

p-hat$_1$ = 502 / 2939 = .1708

(c) What was the risk of back injury among the workers who did not use back belts?

p-hat$_2$ = 455 / 2601 = .1749

(d) What is the relative risk of back injury associated with back belt use?

RR = p-hat$_1$/ p-hat$_2$ = (502/2939) / (455/2601) = .1708 / .1749 = 0.98

(e) Interpret the relative risk.

Risks in the groups are about the same; no association between back belt use and back injury occurrence.

(g) Explain how prior back injury could confound the relation between back belt use and back injury occurrence using the properties of a confounder discussed in class.

- **Back belt use associated with prior back injury.**
- **Prior back injury independent risk factor for current back injury.**
- **Prior back injury not in causal pathway**

+ 2 FREE

/10

(3) Cohort **study.** Information is presented on the survival experience of 1430 infants in a cohort study. Simple (crude) data are:

Crude data

	Died	Survived	
Exposed	40	746	786
Non-exposed	12	632	644

(a) Determine the risk of not surviving in the exposed group.

✓ **p-hat1 = 40 / 786 = .05089**

(b) Determine the risk in the non-exposed group.

✓ **p-hat2 = 12 / 644 = .01863**

(c) What is the relative risk associated with exposure?

✓ **RR-hat = .05089 / .01863 = 2.731**

(d) What do you conclude so far?

✓ **Exposure ~triples risk**

Two subgroups contributed to the data. The mortality experience of each subgroup is shown below.

Subgroup 1

	Died	Survived	
Exposed	6	352	358
Non-exposed	8	586	594

Subgroup 2

	Died	Survived	
Exposed	34	394	428
Non-exposed	4	46	50

-- Continued on next page --

4

(e) Calculate the relative risk associated with the exposure within subgroup 1.

✓ **1.244**

(f) Calculate the relative risk associated with the exposure within subgroup 2.

✓ **0.993**

(g) Upon inspection the strata-specific relative risks do not appear very heterogeneous. Now do a formal test for interaction. List H_0, the test statistic and df computed by WinPEPI, and the conclusion. Be aware: this is a cohort study, not a case-control study.

✓ **H_0: $RR_1 = RR_2$**

✓ **Heterogeneity chi-sq = 0.094 DF:1 P = 0.760**

✓ **No significant interaction**

(i) Report the M-H summary relative risk estimate for the exposure and outcome. Include a 95% CI for the summary *RR*.

✓ **Mantel-Haenszel estimator of ratio = 1.108**

✓ **95% CI = 0.54 to 2.28**

(j) Overall, what was the effect of the exposure?

✓ **The exposure had little or no effect**

(h) How do you explain the apparent contradiction between the results and conclusions in analysis (d) and analysis (j)?

✓ **Confounding**

+ 2 FREE ✓✓

/ 11