Vic

EPIDEMIOLOGY 168

Fall 1998

Midterm Examination

19 October 1998

- Please **do not write your name** on this examination. Instead write the FIRST LETTER of your LAST NAME followed by the last four digits of your student ID number in the space provided in the upper-right hand corner of each sheet.

1. Remember:
   - WRITE ALL ANSWERS AND INTERMEDIATE RESULTS ON THE FOLLOWING PAGES
   - **WRITE LEGIBLY**
   - You may use a calculator, an English, foreign language, or medical dictionary.
2. When finished make certain:
   - Your code number appears on all pages.
   - You signed your name and recorded your ID number as it appears on the exam on the sign-out sheet under the pledge: "I have neither given nor received help from others in completing this examination".
3. Exams will be returned in approximately one week.

☐ ID # ___ ___ ___ ___

Epidemiology 168 • Fall 1998

Midterm Exam

1A. Briefly summarize two criteria on which disease classifications are based. Discuss a reason why these two criteria do not always correspond with one another. (3 pts)

1B. List two examples of each of the two types of criteria you mentioned in 1A. (2 pts)

2. Cohort studies can form the framework for efficient substudies, using nested case-control and case-cohort designs. Which of the following best compares and contrasts these nested case control studies and case-cohort studies of this type. (3 pts)

A. Both nested case control and case-cohort studies select controls that are matched on time of case development but only case-cohort studies allow for multiple comparisons with different case groups.
B. Both nested case control and case-cohort studies select controls from the entire baseline cohort, but in case-cohort studies the selection is done at random.
C. In case-cohort studies a single group of controls can be used for comparison with several case groups, whereas in nested case-control studies controls are often matched to cases.
D. In nested case control studies, cases are selected entirely from the non-exposed cohort group, whereas in case-cohort studies cases are drawn from the entire cohort.
E. both c and d

3. Name the three component parts of a incidence measure. (3 pts)

4. Over a ten-year period the number of bicycle injury events in a population increases even as the age-standardized bicycle injury rate decreases in the population. Describe two conditions that could cause this outcome (assume the definition of a bicycle injury and the quality of the data remain constant over the period) (3 pts)

5. Which of the following best describes the condition(s) that are required for the odds ratio (OR) to estimate the risk ratio (RR) in a case-control study of a disease whose duration is related to exposure status. (choose one best answer) (3pts)

A. Incident cases are identified for a defined population at risk.
B. The controls estimate exposure in the base population that gave rise to the cases.
C. The disease outcome is rare in the base population at risk.
D. All of the above.

6. Swaen et al (1998) conducted a study of 6,803 males who worked for at least six months before 1/1/80 at one of nine chemical plants in the Netherlands. The workers were followed for mortality from 1/1/56 until 1/1/96. Before 1/1/80, 2,842 of the workers were occupationally exposed to acrylonitrile and the other 3,961 workers were not exposed to acrylonitrile. After 1/1/80, there was no exposure to acrylonitrile. To measure the association between occupational exposure to acrylonitrile and several outcomes, the investigators calculated standardized mortality ratios (SMRs) for both the exposed and the unexposed workers. Age-interval-specific person-years were generated for exposed and unexposed groups and were multiplied by the mortality rates for the total male population of the Netherlands to generate expected numbers of cause specific deaths. There were 290 deaths due to all causes among men who had ever been exposed and 983 deaths due to all causes among men who had never been unexposed.

6A. What study design did the investigators use? (2 pts)

6B. Compute the (cumulative) crude mortality ratio for exposure to acrylonitirile? State two major problems in using this crude ratio as a measure of the relative mortality risk associated with acrylonitirile exposure. (3 pts)

6C. For brain cancer, the SMR for the exposed workers (SMR=173.9) was more than twice the SMR for the unexposed workers (SMR=85.7). Why are these two SMRs not strictly comparable? (3 pts)

6D. What measure of effect could be calculated to strictly compare all-cause mortality between the exposed and the unexposed group. (3 pts)

ID # ___ ___ ___ ___

7. Generally speaking there are two kinds of strokes, ischemic (blood flow is restricted to brain tissue due to blockage of an artery in or leading to the brain) and hemorrhagic (a vessel in the brain ruptures causing bleeding in the brain). These two pathologic processes are quite different. A panel of experts reviewed the medical records of 525 patients discharged from the hospital with diagnosis codes indicative of a stroke (ICD 430-438). The panel classified strokes as either ischemic or not ischemic. Of the 525 cases, 325 had a discharge diagnosis code for ischemic stroke (ICD code 434). For 85 of these 325 patients, the panel determined that their strokes were not ischemic strokes. All but 20 of the patients with discharge diagnosis codes other than 434 were determined by the panel to had non-ischemic strokes.

Compute the <u>sensitivity</u>, <u>specificity</u>, and <u>positive predictive value</u> of a hospital discharge code for ischemic stroke (ICD code 434) in classifying a patient as truly having had an ischemic stroke. (Assume the diagnosis reached by the panel is correct.)

7A. sensitivity of a 434 code: (2 pts)

7B. specificity of a 434 code: (2 pts)

7C. positive predictive value of a 434 code: (2 pts)

7D. Constructing a receiver/response operating characteristic (ROC) curve may be useful in understanding the implications of using different case definitions or cutpoints. Briefly explain what a ROC curve is and what information it provides. (3 pts)

7E. Suppose that among stroke patients, the sensitivity and specificity for the use of a 434 discharge code to identify ischemic strokes, respectively, 99% and 50%. Which of the following best describes an ischemic stroke case group identified on the basis of a 434 discharge code? (choose one best answer). (3 pts)

A. The case group would be highly homogenous with respect to pathophysiology of stroke.
B. The case group would be highly heterogeneous with respect to pathophysiology of stroke.
D. The case group would be balanced with respect to pathophysiology of stroke.
E. The case group would represent the source population of cases.

7F. In most situations, what two factors are the primary determinants of the positive predictive value of the application of a screening test? (2 pts)

ID # ___ ___ ___ ___

8. A study was conducted to compare the rates of automobile collisions in two cities, one of which used unmarked patrol cars to enforce traffic laws. The researchers were impressed with studies that suggest that the use of cell phones and pagers contribute to auto collisions. They wanted to adjust (standardize) the rates of auto collisions in the two cities for cell phone and pager use. Data on cell phone use and auto collisions in the two cities were collected and presented in the table below.

Cell phone and pager use	Corona del Mar, California			Boulder, Colorado		
	# persons	# accidents	Rate*	# persons	# accidents	Rate*
Heavy	4479	293		100	2	
Moderate	974	27		300	6	
Never	1106	15		8293	145	
Total	6559	335		8693	153	

* per 1000 persons

8A. calculate the crude total and cell phone/pager use specific rates for Corona del Mar and Boulder. How do these two cities compare in crude incidence of auto accidents. (3 pts)

8B. Using the combined number of persons in both areas as a standard, calculate a standardized rate (standardized for cell phone/pager use) for each of the states. Use the direct standardization method. Briefly describe how these standardized rates compare with each other and with the crude rates and note any meaningful differences and their implications for gaining insight into the effect of Corona del Mar's use of unmarked patrol cars. (4 pts)

8C. In general, which of the following best describes a major weakness of both crude and adjusted rates? (3 pts)

A. Both measures hide or obscure the heterogeneity in the population.
B. Both measures are only estimates of the true population rate.
C. Neither measure can be used to determine the magnitude of disease burden in the population.
D. None of the above.

ID # ___ ___ ___ ___

9. The Minnesota Heart Health Program is an example of a community intervention trial to evaluate the effectiveness of an educational intervention program. Which of the following best describes the unit of assignment, the unit of observation, and the unit of analysis in this this type of study (in this order)? (3 pts)

A. group, person, group
B. person, group, group
C. group, group, group
D. person, person, group

10. Indicate next to each statement below whether you consider it to be TRUE or FALSE. (2 points each)

_____ A. As described in class, a randomized clinic trial is an example of a prospective dynamic cohort study.

_____ B. A disadvantage of the cohort design compared to a case control study is that in a cohort study one needs to follow a large number of participants if the disease is rare.

_____ C. Ecological studies cannot directly assess causal inference because they measure disease and exposure in a person at the same point in time.

_____ D. A case report is a type of descriptive study that is commonly conducted, partially because an appropriate control group is easily defined.

_____ E. A risk difference is determined by the absolute difference in two incidence rates, whereas the relative difference is considered an attributable risk.

_____ F. As an estimate of a relative risk, an odds ratio is a measure of association that can be used to determine the magnitude of an association between exposure and an outcome.

_____ G. Case control studies have several crucial advantages that relate to their efficiency for studying rare conditions and those with prolonged induction and their efficiency in examining many exposure and outcomes.

_____ H. The decision to use an incidence density measure or a cumulative incidence as a measure of the strength of association may depend on the objectives of the study. Cumulative incidence is preferred if estimating individual risk is the main objective.

_____ I. A standardized mortality ratio (SMR) can be determined using indirect adjustment. Because rates from a standard population are used, SMR's from two study populations can be compared as long as the rates in the standard population are stable.

_____ J. Comparability between cases and controls is a important step in constructing a case-control study. It should be possible to detect exposure in controls to the same extent as in cases. It is also critical that controls have similar motivation and availability as cases. These two conditions are best met when controls are selected from the general population.

ID # ___ ___ ___ ___

11. Attributable measures are used by researchers to assess the public health impact of a detrimental exposure. Given data from a cohort study on the incidence of stroke (see below), estimate the attributable risk proportion among the exposed (physically inactive). Explain your answer in one sentence.

Physical activity level	Did develop a stroke	Do not develop a stroke	Person years (PY)	Incidence per 1,000 PY
ACTIVE	45	5,955	43,200	
INACTIVE	135	13,865	100,800	
Total	180	19,820	144,000	

11A. attributable risk proportion (INACTIVITY) (3 pts)

Explain:

11B. Data from the U.S. National Health And Nutrition Examination Survey (NHANES) estimate the prevalence of a physically active lifestyle (at least 30 minutes of moderate activity 3 days per week) as 27%. Using this information and your answer to part (A), estimate what we can hope to accomplish with programs to get people to be physically active in the total population. In one sentence explain your answer. (3 pts)

Explain:

11C. What assumption is required to interpret the measures from part A and B as indicative of potential for stroke prevention. (3 pts)

ID # ___ ___ ___ ___

12. Suppose that in 1998 researchers hypothesized that communication ability and skill in young adulthood are related subsequent risk of Alzheimer's Disease. To test this hypothesis, the researchers evaluated hand written essays completed by a group of 350 nuns joining a single religious sect in 1930. By careful review of these writing samples, the researchers categorized the nuns as having either a high error profile (N=150) or a low error profile (N=200). Through surveillance of death certificates and other methods the researchers ascertained vital status and cause of death for each nun through 1998. An accounting of all deaths produced the table below.

Cause of Death and Year by Handwriting Profile Status

Cause of Death	High error profile: # of Deaths	High error profile: Year of Death	Cause of Death	Low error profile: # of Deaths	Low error profile: Year of Death
Alzheimer's Disease	2	1980	Alzheimer's Disease	1	1985
Alzheimer's Disease	5	1985	Alzheimer's Disease	3	1990
Alzheimer's Disease	6	1990	Alzheimer's Disease	4	1995
Alzheimer's Disease	5	1995	Heart Disease	8	1980
Heart Disease	10	1980	Heart Disease	10	1990
Heart Disease	15	1995	Other	20	1960
Other	25	1960	Other	10	1970
Other	30	1970			

12A. Describe the type of study design used in this example. (2 pts)

12B. Compute the incidence density rate of Alzheimer's disease death for those with a high error profile and for those with a low error profile. (4 pts) Show your work.

12C. Compute the incidence density ratio for Alzheimer's disease death associated with a high error communication profile. Explain, in two sentences or less, what this value means. (4 pts)

12D. Using data from this study compute an odds ratio for the association of a high error communication profile with death from Alzheimer's disease. Show a clearly labeled 2x2 table. (3 pts)

12E. Explain in two sentences or less why the incidence density and the odds ratios are (or are not) similar. (3 pts)

ID #____ ____ ____ ____

Epidemiology 168 • Fall 1998

KEY to Midterm Exam

1A.Manifestational criteria: disease definition and classification based on observable characteristics, such as symptoms, signs, history, labloratory findings, response to treatment, prognosis.
Causal criteria: disease definiiton and classification based on the cause of the condition,

1B.Manifestational criteria: Examples are cancers, arthritis, cholescystitis, schizophrenia, depression, addiction, insomnia, . . .
Causal criteria : microbial diseases for which the pathogen has been identified (syphilis, TB, malaria, yellow fever, influenza, etc.), lead poisoning, birth trauma,

2.(C)- Other choices are incorrect because controls in case-cohort studies are not matched to cases (a), contrrols are selected at random with both designs (b), and cases must be selected without regard to exposure (d).

3. New cases or events, populaiton at risk or source population, passage of time

4. The size of the population may have grown (number increases even though rate does not); the age distribution of the populaiton may have changed (e.g., influx of families with small children, outmigration of families with older children), so that age-standardized rate may not change but a greater proportion of the population may be in the higher risk age range (assuming that younger children have higher injury rates).

5.(D)- All of the above - use of prevalent cases requires that duration is not related to exposure, controls should provide estimate of exposure in study base, and although OR can estimate IDR without rare disease assumption, that assumption is required to estimate RR.

6A. A (retrospective) cohort study.

6B.CIR = (290/2,842) / (983/3,961) = 0.411
A cumulative measure ignores possible differences in length of follow-up between groups being compared. A crude measure ignores possible differences in the age distributions between men who have been exposed and men who have not.

6C. SMRs are an indirect method of standardization, since they are based on weighted averages for which the weights are taken from the population whose SMR is being computed rather than from a "standard" population. Unless the age (and in this case, age-calendar year interval) distributions for the populations whose SMR's are being computed are the same, then the weighted averages that make up the SMR's are based on different sets of weights and are not strictly comparable. Since the age-interval distributions of the exposed and unexposed workers may well differ, their SMR's are not strictly comparable.

6D.Directly-standardized rates are based on weights from a "standard" population, and are therefore strictly comparable. [However, if the number of brain cancers is small (as is likely), the rate estimates will be imprecise. ~~However, unless~~ the age-interval distributions for exposed and unexposed men are examined, it would their SMR's should not be compared.

7A. sensitivity= (325-85) correctly identified ischemic strokes /(325-85+20) ischemic strokes = 240 / 260 = 92.3%

7B. specificity = 20 / (525-325) = 20 / 200 = 0.10, Specificity = 0.90

7C. positive predictive value of a 434 code= (325-85) correctly identified ischemic strokes / 325 strokes with ICD code 434 = 73.8%

7D. An ROC curve plots the value of sensitivity and specificity for each case definition or cutpoint. Examining the ROC curve shows the trade-off between sensitivity and specificity that is available for the diagnostic test or measurement method. [The area between the identity diagonal (slope = 1.0) and the ROC curve serves as a measure of accuracy that takes into account both sensitivity and specificity, with the assumption that the costs of false negatives and false positives are the same.

7E. (B) - due to the low specificity (50%), half of hemmorhagic strokes in the patient group will be classified as ischemic strokes.

7F. Specificity and prevalence of the condition

8A. Corona del Mar has a 2.9 times higher crude accident rate than Boulder.
Corona del Mar = 51.1/1000 and Boulder = 17.6/1000

8B. Adjusted rates=
Corona del Mar: (4579 x .0654) + (1274 x .0277) + (9799 x .0136)/15,652 = 29.9/1000
Boulder: (4579 x .0200) + (1274 x .0200) + (9799 x .0178)/15,652 = 18.6/1000

The cell phone/pager adjusted auto accident rate for Corona del Mar was 1.6 times that of Boulder. A portion of the difference seen in the crude rates was due to differences in the distribution of use of cell phones and pagers between the two cities.

The standard weights are the sum of the population sizes for the two cities. The weighted rates are the rates for each city, weighted (multiplied) by the standard weights. The total of the weighted rates is the directly standardized rate. A problem in using the directly standardized rates is that there are small numbers of cellular phone and pager users in Boulder.

The higher crude rate in Corona del Mar reflects the much higher use of cellular phones and pagers, which is associated with a much higher accident rate. The difference is reduced for the standardized rates, since these control for the different distributions of cellular phones and pagers between the two cities. However, this is a situation where it is essential to examine the specific rates, since Boulder has lower accident rates among cellular phone and pager users but a higher rate among never-users.

Since the rates in never users are quite similar, Corona del Mar is likely to make its greatest impact on accident rates by getting motorists to reduce cellular phone and pager use while driving or finding some way to such use safer (promote the use of "designated drivers"!?).

8C.(A) Both measures obscure heterogeneity (variation) in rates across subgroups.

9.(A) Community intervention trials of this type assign groups to treatments and collect measurements from individuals. The unit of analysis must be the same as the unit of assignment (GROUP) or both (i.e., using mixed models).

T?

ID #____ ____ ____ ____

10.A (F), B (T), C)F), D (F), E (F), F (T), G (F), H (T), I (F) , J (F)

11A. ARP = (I1 - I0) / I1 = (RR-1) / RR = (1.34-1.04) / 1.34 = 0.30 / 1.34 = 22% (after rounding) The "I can't remember formulas" method:
ARP = attributable cases / all exposed cases = attributable cases / 135
Attributable cases = attributable risk * Exposed PY = (1.34-1.04)*100,800 = 30.24
ARP = 30/135 = 22% (after rounding)

Explain: Based on these data, 22% (about one in five) strokes in people who are physically inactive can be attributed to their physical activity; in other words, if physically inactive people became physically active early enough in their lives, thei stroke incidence would be reduced by 22%

11B. A key point to note here is that 27% is the prevalence of physically active people, whereas the exposure is physical inactivity, whose prevalence is therefore 100% - 27% = 73%

PARP = p1(RR-1) / [1 + p1(RR-1)] = 0.73(1.286-1) / [1 + 0.73(1.286-1)]
= (0.73 x 0.286) / (1 + 0.73 x 0.286) = 0.209 / 1.209 = 17%
(The formula PARP = (I - I0) / I can also be used by first estimating the crude population incidence, I, as a weighted average of the incidences in exposed and unexposed, weighting by the prevalence of exposure, e.g.: I = (0.73)(1.34) + (0.27)(1.04) = 1.26, so PARP = (1.259 - 1.04) / 1.259 = 17%

The "I can't remember formulas" method:

PARP = Attributable cases / All cases
Attributable cases are (1.34-1.04) x number of exposed person-years. Since we do not know the population size, represent it by n. Based on the NHANES data, 27% of people are physically active, so there are 0.73n physically inactive people (in one year, 0.73 person-years). So: Attributable cases = (1.34-1.04)(0.73) = 0.219.
All cases are exposed cases + unexposed cases. Since we do not know the population size, let it be represented by n. Based on the prevalence of physically active people, there are 0.73n phyisically inactive and 0.27n physically active people (or person-years, if we assume a one-year period). So the total number of cases = exposed cases + unexposed cases = 0.73(1.34) + 0.27(1.04) = 1.259

Therefore, PARP = 0.219n/1.259n = 17%

Note that these measures can be computed more precisely by using the original number of cases and person-years and not rounding intermediate results, but two significant figures is adequate for the actual result, and in this case the answer does not change.

Explain:
Seventeen percent of all strokes in the population are attributable to physical inactivity; if everyone were physically active, there would be 17% fewer strokes.

11C. Attributable risk measures assume that the relationship is causal (i.e., that physical inactivity does in fact cause an ncrease stroke risk). Some of the above interpretations may also require that the process be reversible, so that changing to a physically active lifestyle brings risk down to the level of someone who was not inactive. Another assumption is that the rates and rate ratio observed in the cohort study hold ofr the entire population. Also, we have ignored the effects of other factors, most notably age.

12A.This is a retrospective cohort study (researchers developed the hypothesis in 1998).

12B.High: (2 + 5 + 6 + 5)/8021 = 2.24 per 1,000 py. Low: (1+3+4) / 12,287 = 0.651 per 1,000 py

12C. IDR= ID High / ID low = 2.24/0.651 = 3.4. Nuns with a high error communications profile are 3.4 times more likely to die from Alzheimer's Disease than nuns with a low error profile.

12D.

	Alzheimer's Disease	
Handwriting Profile	AD Yes	AD No
High error	18	132
Low error	8	192

odds ratio = $\frac{(18)\ (192)}{(8)\ (132)} = 3.27$

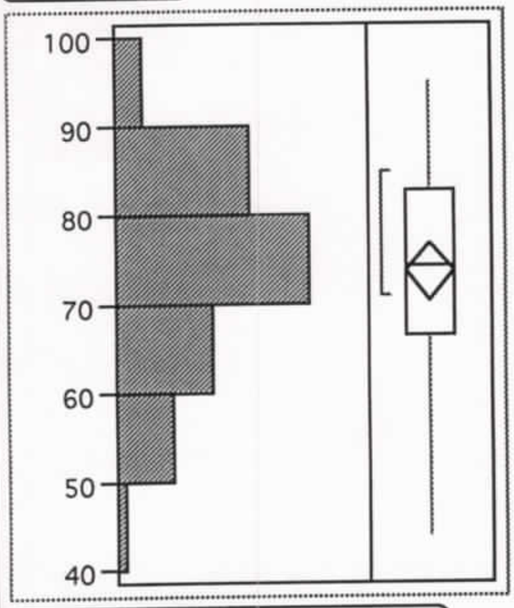
12E. The two are similar because the condition is fairly rare and the overall follow up period is long.

Vic -

1998 Midterm.

I think this is a good distribution and med 75 is appropriate.

midterm



Quantiles

maximum	100.0%	95.000
	99.5%	95.000
	97.5%	95.000
	90.0%	88.000
quartile	75.0%	83.000
median	50.0%	74.500
quartile	25.0%	66.500
	10.0%	56.500
	2.5%	46.625
	0.5%	44.000
minimum	0.0%	44.000

Moments

Mean	73.70370
Std Dev	11.65159
Std Error Mean	1.58558
Upper 95% Mean	76.88397
Lower 95% Mean	70.52344
N	54.00000
Sum Weights	54.00000

epid168 grade sort

Rows	initial	ID	midterm
1	p	18	76
2	?	99	62
3	j	220	87
4	b	264	44
5	s	345	68
6	b	624	74
7	g	654	52
8	g	663	63
9	r	694	85
10	k	915	85
11	h	982	95
12	h	1211	95
13	f	1235	74
14	o	1413	79
15	f	1752	72
16	p	1974	77
17	h	2115	61
18	s	2334	77
19	b	2335	70
20	p	2440	59
21	z	2761	54
22	f	3085	51
23	g	3525	71
24	m	3534	71
25	k	3671	67
26	f	4174	80
27	w	4180	71
28	g	4246	69
29	j	4301	74
30	p	4305	75
31	r	5191	79
32	n	5395	86
33	n	5691	81
34	g	5970	83
35	m	6111	55
36	j	6275	76
37	a	6957	73
38	o	7475	83
39	s	7578	93
40	s	7602	61
41	l	7655	58
42	l	7926	86
43	c	7983	75
44	f	8259	84
45	p	8541	89
46	j	8806	65
47	a	8879	68
48	a	9003	79
49	b	9380	71
50	f	9640	63
51	z	9729	85
52	y	9903	89
53	j	9951	79
54	d	9978	81