Edits for "Epidemiology Qualifying Exam Summary Notes"
by Chia-Yen Chen, John Jackson, Chien-Chang Lee, Mitchell Machiela, and Andreas Neophytou

Updated: 4.27.13. All corrections updated except where noted in red. EMurray

General Notes

- Modeling is spelled with one 'l'
- Parameter (not paramater)

Table of Contents

- Update page numbers in Table of Contents
- Page 4: "Chapter 6. Concepts & Definitions"
    - Spelling Error: "1. Phiolosophical Viewpoints of Causality"
- Page 5: "Chapter 16. Models for Statistical Inference"
    - "1. stratified analysis using regression models"--> lacks capitalization

Chapter 1. Epidemiological Concepts

- Headings misspell 'epidemiological'

Chapter 2. Measures of Frequency and Occurrence

- Page 9: 1.2 Estimator
    - "biased estimator" is missing a "not" in the definition
- Page 14: 1.3 Risk-Set
    - "unadjusted analysis: Kaplan-Mier"--> should be Meier
    - "Proportional" misspelled
- Page 15: P / (1 - P) should be p / (1 - p)
    - P is used here for prevalence, so instead of changing to 'p' I added "where P is prevalence"
- Page 17: 3 Estimation
    - Kaplan-Meir--> should be Meier
- Page 21 (a) The formula for risk should have a negative sign inside exp()
    - Page 19: PR(Y=1|X) should be Pr(Y=1|X)
- Page 21: "poissson" has an extra 's'

Chapter 3. Measures of Association

- Page 25: 3.2
    - Final bullet "counterfactual notation" is missing '=1' in some of the probabilities
- Page 28: 3.5
    - Should the interpretations for the attributable rate percents be in terms of rates instead of fractions of disease?
        - I'm not sure if this should be changed or not. I've left it as is for now.
- Page 28: 4. Preventable Measures
    - Should read "for when exposure is believed to protect against disease
- Page 29: 5. Risk Ratio
    - should read 'preventative [0, 1)'
- Page 29: 6. Incidence Rate Ratio
    - should read 'preventative [0, 1)'

- Page 29: 7. Odds Ratio
    - should read ‘preventative [0, 1)’
- Page 30: Flow Chart
    - open and closed cohort should be flipped

Chapter 4. Measures of Classification

- Page 33: Formula for PPV is incorrect

Chapter 5. Study Designs

- Page 40: 50 yr olds is not a closed cohort
- Page 43: **stratified sampling fractions**: we can allow sampling fractions to vary by covariate level, thus within strata that define the sampling fraction, the rate ratio is consistent.
- Page 43: pseudo-random sampling: choose controls through a process that is arguably random in nature (e.g. next live birth)
- Page 45: an advantage of case-control over cohort is by studying fewer people, can put more effort/resources into ascertainment and measurement of confounders for better accuracy.
- Page 54: Mantel-Hansel misspelled.
- Page 58: Closed cohort study Assumptions: defined by “membership defining event”

Chapter 6. Concepts & Definitions

- Page 67, and forward: I think that individual counterfactuals are better represented with a subscript *i* on the *Y* rather than the *a*
- Page 70: partial exchangeability
- Page 76: under “test for homogeneity of ratio measures” one of the denominators for the variance of the log RR should read N *subscript 0,i*
- Page 77: the use of IPW and Standardization as adjustment methods.... can get $E[Y^{A=a}]$ and $E[Y^{A=a}|M=m]$
- Page 78: the causal effect of A on Y...
- Page 81:
    - 6.1.2. ‘Outcome’ misspelled.
    - ‘Doomed’ misspelled in table
    - ‘(2) there are three sufficient causes for Y’

Chapter 7. Bias

- Page 93: (1) ‘a variable is a function of its parents and a random error term’
- Page 97: D-separation
    - a set of graphic rules to decide if two variables in a DAG are independent (d-separated) or not...
    - (1) A path is blocked if two arrowheads on a path collide.
- Page 99: Exchangeability: lack of paths between A and Y nodes, other than those originating from A, that...
- Page 100: (2) confounding by indication of channeling (figure 7.19 part (b) or (b))
- Page 102: 2.5 **Structural vs Statistical Definition of Confounding**: structural definition of confounding shows that **statistical criteria is NOT sufficient to define confounding**
- Page 103: 2.5, there is unmeasured confounding between A and Y (residual and unmeasured confounding)
- Page 103: under 2.7 heading, “observariobal study”--> “observational study”
- Page 112: (3) Conditioning on a common cause of some trait...

- Page 120 “Idependence”--> “Independence”
- Page 120 - (2) Non-differentiality. Second equality is incorrect.
- Page 122: 4.4: For non-dichotomous treatment....as long as A*-A association is non-monotonic

Chapter 8. Advanced Topics in Causal Inference

- Page 133: 1.3: we need an assumption to link **causal DAG** to **statistical data**; otherwise, a causal DAG is just a graph of no practical use
- Page 134:
    - = the non-descendants of a given...
    - 1.3: Markov factorization implies additional statistical independence
- Page 134: conditional independence... if Y and A **are** d-separated, marginal independence... if Y and A **are** d-separated without **conditioning on** any other...
- Page 134: D-Connection (3): need assumption of **faithfulness** to say that two variables are dependent if they are d-connected
- Page 136: 2.1 - to separate direct (not through intermediate...
- Page 137: 2.1: in the presence of covariates that affect the intermediate and the outcome regardless of association with exposure partial exchangeability does not hold
- Page 137: *Note*:_confounding (need space between colon and “confounding)
- Page 137: Figure 8.4 - U or L?
- Page 137: “If the covariates are affected by exposure or data are **unavailable, G-computation** methods must be used
- Page 139: Direct effects of MSMs: no effect of A0 on Y without intervening on A1=.....; theta1 and theta3 represent direct effect of A0, while theta2 represents indirect effect
- Page 139: Direct effect in SNMs
- Page 141: this estimator identifies the causal effect if the following assumptions hold
- Page 141: solution of **“**carrying last measurement forward**”**
- Page 143: create a pseudo-population that has...
- Page 143: artificial censoring vs. **modeling** effect of observation plan
- Page 144: “exchangeability:_at every time t, **one** can **achieve**...

Chapter 10. Traditional Methods

- Page 156: Table 10.3 title should read RGB not RBG
- Page 157: 1.1.1 count data refers to (1) number of cases and non-cases collected in a closed cohort study
- Page 157: 1.1.1. 2 of a, (N1-a), b, (N0-b) for closed cohort study
- Page 157: 1.1.1. thus, we pretend...for closed cohort study....
- Page 158: construct test statistics: Z=..... follows normal distribution asymptotically
- Page 158: 1.1.2. (1) number of cases (counts) and person-times collected in a closed cohort study
- Page 158: construct test statistics: Z=..... follows normal distribution asymptotically
- Page 159: construct test statistics t=....follows normal distribution asymptotically
- Page 159: 1.1.4. construct test statistics: t follows normal distribution asymptotically
- Page 164: 2. Stratification
    - if the stratum-specific effect measures are all the same, the effect measure for an entire population...
- Page 166: under frequency matching “can be used to create a matched population)

- Page 168: e.g. for 1:1 individually matching on gender only, may perform stratified analysis on gender instead of stratified analysis on matching set
- Page 169: ...of the estimator and when bi > 0 it contributes to the denominator if both...
- Page 169: the same informative risk sets are the ones contributing to the RRMH variance
- Page 169: matched data: risk set sampling assures...are defined at the time each case occurs
- Page 169: Consider figure 10.1 where
- Page 170: for matched cohort data...
- Page 170: consider figure 10.2 where...
- Page 170: in matched data analysis informative risk sets are those with discordant...
- Page 171: McNemar vs Mantel-Haenszel

Chapter 11. G-Methods

- Page 173: ...where the weights are the stratum-specific probabilities of the covariate L...
- Page 174: Indirect standardization: weighted average of stratum specific rate ratios using the distribution of expected exposed cases...
- Page 180:
    - (3) evolve over time in a deterministic way
    - (1) time-dependent confounders present
- Page 185: more efficient than unstabilized weights (W)
- Page 188: Called nested model because structural model can be nested in the exposure model
- Page 189: instrumental variables can be used to estimate causal effects without exchangeability
- Page 190: ...under additional assumptions without these assumptions...

Chapter 12. Modeling Concepts

- Page 195: "Unbiased: in repetitions of the study the 95% CI..."
- Page 197:
    - The unspecified parameters are allowed to follow any distribution
    - Parametric: function of data; synonym: unsaturated
- Page 198: Random-Effects Models: synonyms: hierarchical
- Page 198:
    - the assumptions made in the modeling strategy should reflect one's prior beliefs
    - Under 2.2, Ex. for binary exposure, outcome, and 10 binary covariates, the saturated model has $2^{10} = 1025$ parameters....Even if assume no effect modification → $2^9 = 513$ parameters"
        - I don't think this correction is right, so I've left this as is (added text underlined below). There are 10 covariates and 1 exposure, so there are $2^{11} = 2048$ parameters to estimate if we assume all possible interactions as in the text, and $2^{10} + 1 = 1025$ if we assume no effect modification for the effect of exposure, but all possible interactions between the covariates
- Page 199: 2.3: Sparse Data: Highly stratified data, but can obtain useful confidence intervals
- Page 199: Under 2.3, "usually prior beliefs for risk factors of disease are more 'sharp' than beliefs for predictors of exposure"
- Page 199: 2.4 Bias
- Page 200: (5) if statistical tests are used to verify assumptions,...part on the strength of one's beliefs

- Page 205: “You can infer effect modification on the additive scale if it is qualitative (change in sign)”

Chapter 13: Time

- Page 209: units of person x time
- Page 210: unnecessary space: “average rate_, based on an average population_,”
- Page 211: Period Effect, definition: “change in disease frequency that is specific to a calendar time”
- Page 212: -also viewed as an interaction between *age* and *calendar* time
- Page 213: **Multiple events per subject:** For example, time to recurrent … which the same subject...
- Page 215: age is often the time scale (two words)
- Page 216: 5.1 Induction time: residual effect: “subsequent changes in disease incidence attributable to exposure
- Page 217: unnecessary space: “chronic exposures_,”
- Page 217: assume neglibile induction period. (don’t need period)
- Page 217: in this situation time of exposure-time at risk of its effect are
- Page 217: Immortal person time
- Page 218: 5.2 (2): “often not much power to determine which is exposure definition is more biologically relevant”
- Page 218: Composite measures.... “simultaneous interpretation”

Chapter 14: Continuous exposures & Confounders

- Page 223: under Score test: the exposure A can be replaced by any ordinal score vector …. to different exposure ….

Chapter 15: Missing Data

- Page 233: Efficiency decreased because sample size is reduced by (1-f)%
- Page 233: Unbiased in MAR if missingness is multiplicative
- Page 233: In general, if each cell has a fraction...then complete case method is unbiased
- Page 233: For continuous outcome, even though....
- Page 233: 3. Missing indicator: Recode missing data as zero....then use it for stratified analysis or control in regression model
- Page 233: For missing data on many covariates....must have variation in the outcome
- Page 234: 4.1: Single variable: then run weight regression of of Y on A,L
- Page 235: 5.2.1:Underestimates variability in Xj and weakens
- Page 238: 6. Maximum Likelihood: can deal with MNAR

Chapter 16: Models for Statistical Inference

- Page 243:1:. exposures and confounders are both included in the model as covariates
- Page 246: 4.4: B can be interpreted as log odds ratio...or log odds ratio associated with a one unit change of X (continuous variable controlling for other variables)
- Page 248: What & Where is the Offset? : we add it to the model to give a rate ratio interpretation to the parameters even though we’re modeling the average count of events
- Page 251: 6.4: the incidence rate is small (less than 0.1)...
- Page 252: 7.4: Note that the regression equation...as the effects of matching factors on the outcome Y cannot be computed. The stratum-specific effects of the matching factors are included in the intercept terms.

- Page 252: 7.5 practical issues: the conditional maximum likelihood....multiplicative intercept models for rates and odds
- Page 253: 8.1 counting process: in the Anderson-Gill formulation, the time....(not to include_ left hand side...)