# Final Examination
# PB HLTH 250C: Advanced Epidemiologic Methods
# Katherine Rose Wolf
# May 13, 2020



PB HLTH 250C: Final Exam

**Due by 6pm, Friday 15 May 2020**

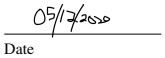
**This final exam is take-home and there are 9 questions.** You may use any PB HLTH 250C class materials from this semester (e.g. notes, readings, previous homework assignments), any PB HLTH 252 materials, or any published materials (books, journal articles) but **you must not discuss this exam with anyone except the course instructor and GSI.** This includes, but is not limited to communications in-person, *via* phone, Skype, Zoom, text messaging, internet bulletin board, social media, Owl Post, Floo Network, etc. . . You are also specifically not allowed to access exams from previous offerings of PB HLTH 250C. If you have any questions regarding this please see the GSI or course instructor.

**After you have submitted your exam do not discuss it amongst your peers until after grades are posted.**

**Read all questions carefully before answering.** Please maintain numbering on sub-questions, type your responses, **do not submit unformatted computer output or code** and **please keep answers brief.** Report measures of association to **two** decimal places. Submit a pdf to Gradescope with the following statement at the beginning, and sign below it:

On my honor, I have neither given nor received any assistance in the taking of this exam, as stipulated above, and that I will refrain from discussing it until grades are posted.

Signature



Date

# Questions

## Bias analysis

**Answer the following *TRUE*/*FALSE* questions regarding bias analysis in general and *provide a 1-2 sentence justification* (please read carefully). (5 points each)**

1. **Deterministic bias analysis incorporates uncertainty in the bias parameters into your analysis.**

   FALSE. Deterministic bias analysis specifies biasing relationships as if they were known with certainty (lecture 8, slide 9), assigning one fixed value to each bias parameter that is not presumed to coem from a distribution.

2. **Probabilistic bias analysis yields a range of bias-corrected measures of association.**

   TRUE. Probabilistic bias analysis yields a distribution of bias-corrected estimates (that we can then summarize concisely with an estimate of the central tendency and interval) (lecture 10, slide 6). (As Dr. Bradshaw uses the term, the distribution accounts for uncertainty in the bias parameters but does not incorporate random error in the original estimate of the target parameter (lecture 10, slide 7).)

3. **The gamma distribution is a good choice for the distribution of the bias parameter for a proportion/prevalence/probability.**

   FALSE. A Gamma-distributed variable has no upper bound, i.e., a $\theta \sim Gamma(a, b)$ has the range $\theta \in (0, \infty)$, whereas proportions/prevalences/probabilities can only range from 0 to 1. Better choices might be a uniform, beta, trapezoid, or triangle distribution specified to range from 0 to 1.

4. **The normal distribution is a good choice for the distribution of the bias parameter for a log-relative risk.**

   TRUE. A log relative risk can theoretically range from $-\infty$ to $\infty$, as does the normal distribution. read Greenland for this

5. **One reason quantitative bias analysis is important is that systematic errors can be larger than random errors.**

   TRUE. Not only can systematic errors be larger than random errors, but also they do not necessarily shrink as sample size increases, unlike random errors.

## Paper evaluation

**A paper by Keil et al. (2014) in *Environmental Health* examined the association between exposure to imidacloprid (a common flea and tick medication for pets) and autisum spectrum disorder (ASD) among children. Read the accompanying paper, paying particular attention to the methods and results sections, and answer the following questions regarding the methodological approach and presentation of results *(please keep answers brief)*:**

6. **The authors mention using "3 jointly estimated models to simultaneously model the 'true' exposure and estimate its association with ASD." For the analysis in this paper, write out each of these models as described by the authors.[1] Consider only the non-differential misclassification scenario (group 2). Give definitions for each of the covariates and parameters in your models. Specify any distributional assumptions for the outcomes on the models, but here you do not need to specify the priors on the model parameters: *Hint: refer to misclassification example from the Bayesian Bias Analysis lecture.***

   a. **Exposure model (10 points)**

$$\begin{aligned} logit(\pi_{a,i}) &= \alpha_1 + \alpha_2 d_{1,i} + \alpha_3 d_{2,i} + \alpha_4 r_i + \alpha_5 t_i + \alpha_6 f_i \\ &\quad + \alpha_7 s_i + \alpha_8 g_i + \alpha_9 b_{1,i} + \alpha_{10} b_{2,i} + \alpha_{11} b_{3,i} + \alpha_{12} b_{4,i} \\ a_i &\sim Bernoulli(\pi_{a,i}) \end{aligned}$$

   b. **Measurement model (10 points)**

$$\begin{aligned} p_{a_i^*} &= a_i Se + (1 - a_i)(1 - Sp) \\ a_i^* &\sim Bernoulli(p_{a^*,i}) \end{aligned}$$

   c. **Outcome model (10 points)**

$$\begin{aligned} logit(\pi_{y,i}) &= \beta_1 + \beta_2 d_{1,i} + \beta_3 d_{2,i} + \beta_4 race_i + \beta_5 t_i + \beta_6 f_i \\ &\quad + \beta_7 s_i + \beta_8 g_i + \beta_9 b_{1,i} + \beta_{10} b_{2,i} + \beta_{11} b_{3,i} + \beta_{12} b_{4,i} + \beta_{13} a_i \\ y_i &\sim Bernoulli(\pi_{y,i}) \end{aligned}$$

   where

   - $i$ indexes the study participants;
   - $d_i$ is a categorical covariate for maternal education, factored into the binary covariates
     - $d_{0,i}$, where $d_{0,i} = 1$ indicates that participant $i$'s mother had a college degree and $d_{0,i} = 0$ otherwise; $d_{0,i}$ serves as the reference, thus not appearing in the model but rather subsumed in the intercept term ($\alpha_1$ in the exposure model or $\beta_1$ in the outcome model);

[1] Note correction at top of page 3. Text should read: 'probability of reported exposure, given "true" exposure and case/control status...'

  - $d_{1,i}$, where $d_{1,i} = 1$ indicates that participant $i$'s mother had a high school education; and
  - $d_{2,i}$, where $d_{2,i} = 1$ indicates that participant $i$'s mother had at least some college;
- $r_i$, a binary covariate for race ethnicity, where $r_i = 1$ indicates that participant $i$ is not non-Hispanic/Latinx white, and $r_i = 0$ indicates that participant $i$ is white and not Hispanic or Latinx;
- $t_i$, an ordinal integer covariate indicating the parity of participant $i$;
- $s_i$, a binary covariate for sex, where $s_i = 1$ indicates that participant $i$ is female and $s_i = 0$ that participant $i$ is male;
- $f_i$, a binary covariate for pet ownership during pregnancy, where $f_i = 1$ indicates yes and $f_i = 0$ indicates no;
- $g_i$ is an ordinal integer covariate for the matching factor of maternal age at the interview;
- $b_i$ is a five-category categorical covariate for region of birth, factored into the five binary indicator covariates
  - $b_{0,i}$, where $b_{0,i} = 1$ if participant $i$ was recruited from unnamed Regional Center 0 and $b_{0,i} = 0$ otherwise; it serves as the reference, thus not appearing in the model but rather is subsumed in the intercept term ($\alpha_1$ in the exposure model or $\beta_1$ in the outcome model);
  - $b_{1,i}$, where $b_{1,i} = 1$ indicates that participant $i$ was recruited from unnamed Regional Center 1 and $b_{1,i} = 0$ otherwise;
  - $b_{2,i}$, where $b_{2,i} = 1$ indicates that participant $i$ was recruited from unnamed Regional Center 2 and $b_{2,i} = 0$ otherwise;
  - $b_{3,i}$, where $b_{3,i} = 1$ indicates that participant $i$ was recruited from unnamed Regional Center 3 and $b_{3,i} = 0$ otherwise; and
  - $b_{4,i}$, where $b_{4,i} = 1$ indicates that participant $i$ was recruited from unnamed Regional Center 4 and $b_{4,i} = 0$ otherwise; and
- $a_i$ is a binary variable for the exposure, where $a_i = 1$ indicates that the authors categorized participant $i$ as exposed to imidacloprid, where they defined exposure as the mother reporting any household usage of sprays, dusts, powders, or skin applications for fleas or ticks on pets from 3 months before conception until birth, and $a_i = 0$ otherwise.

7. **The authors describe a sequence of analyses where each one treats the sensitivity and false positive rate (1-specificity) as fixed. In the case of non-differential misclassification (group 2), sensitivity ranges from 0.70-0.95 and false positive probability (1-specificity) from 0.00-0.20 for both cases (ASD) and controls (TD). Instead of specific fixed scenarios, describe one possible set of prior distributions for sensitivity and specificity that would be consistent with these ranges of values (5 points)** ***Hint: Assume there is no prior probability outside of the stated ranges. You should specify two reasonable distributions of your choice (one for each bias parameter), and include specific values for the hyperparameters.***

8. **The model above assumes non-differential exposure misclassification.**

   a. **Modify and present the approopriate sub-model from question 6 to accommodate *differential* misclassification (this will only involve one of (a), (b), or (c)). Make sure to define each of the necessary bias parameters in this new sub-model. (5 points)**

   b. **Assuming this modification, specify the necessary prior distributions for the bias parameters consistent with the group 3 scenario (see Figure 2). (5 points)**

## Analytic plan

9. **Given the analysis and methodological issue that you outlined on the midterm, *briefly* outline an plan for the analysis. Consider writing this in the format you would for a MPH capstone proposal or dissertation prospectus. Be specific about 1) target parameter, 2) modeling forms, 3) covariates included, and 4) how you will inform any priors (e.g. for Bayesian analysis or probabilistic bias analysis). Include an expression for the model form (you can denote sets of covariates in vector notation for brevity). (30 points) LIMIT YOUR ANSWER TO ONE PAGE OR LESS.**

## Bonus (3 points)

**Include (with your answers here) documentation that you have completed the online course evaluation for PBHLTH 250C. (And thank you!) (Please do not include any information on your responses to the questions.)**

Berkeley

Sign Out

**Fillout Task List**

**Task Owner:** Katherine Wolf
**Project Title:** Spring 2020 Evaluations
**Category:** 2020
**Subcategory:** Spring

| Subject | Due date | Status |
|---|---|---|
| PB HLTH 252D LEC 001 CAUSAL INFERENCE I | Sunday, May 10, 2020 | Completed |
| PB HLTH 252D LEC 001 CAUSAL INFERENCE I (EVAL FOR GSI) | Sunday, May 10, 2020 | Completed |
| PB HLTH 250C LEC 001 ADV EPI MTHDS | Sunday, May 10, 2020 | Completed |
| PB HLTH 250C LEC 001 ADV EPI MTHDS (EVAL FOR GSI) | Sunday, May 10, 2020 | Completed |
| PB HLTH 255D LEC 001 SOCIAL EPI METHODS | Sunday, May 10, 2020 | Completed |
| IAS 250 SEM 001 GRADUATE STDYS IAS | Sunday, May 10, 2020 | Completed |

Mobile Version | Standard Version