ADVANCED METHODS FOR THE DESIGN AND ANALYSIS OF COHORT STUDIES

Course No.	340.728.01
Dates	1st Term (2018/2019)
Classroom	W5030
Class meets	Mondays, 8:30 - 9:20 a.m. Tuesdays & Thursdays, 8:30 - 10:20 a.m.

Instructors:

Alvaro Muñoz, PhD (amunoz@jhu.edu)
Christopher Cox, PhD (ccox18@jhu.edu)
Derek Ng, PhD (dng@jhu.edu)

Lecturers:

Emily Gurley, PhD
Michael Schneider, MS

Lab Instructor:

Michael Schneider, MS (mschnei4@jhu.edu)

Teaching Assistants:

Daniel Antiporta, PhD student (dantipo1@jhu.edu)
Yun Soo Hong, PhD student (hong.yunsoo@jhu.edu)
Pei-Lun Kuo, PhD student (perrykuo@jhu.edu)

For instructor and TA office hours—see next page

Office hours for AMDACS (2018 – 2019)

YSH: Yun Soo Hong; PK: Pei-Lun Kuo; CC: Chris Cox; MS: Mike Schneider;
DA: Daniel Antiporta; AM: Alvaro Muñoz; DN: Derek Ng

Date	Day	Time	Room Number	Team Members
9/11	Tuesday	1:30 – 3:00PM	E6130	YSH & PK & CC
9/14	Friday	1:30 – 3:00PM	W6015	MS & DA & AM
9/18	Tuesday	1:30 – 3:00PM	E6130	YSH & PK & CC
9/21	Friday	1:30 – 3:00PM	W6015	MS & DA & AM
9/25	Tuesday	1:30 – 3:00PM	E6130	YSH & PK & CC
9/28	Friday	1:30 – 3:00PM	W6015	MS & DA & AM
10/2	Tuesday	1:30 – 3:00PM	E6130	YSH & PK & CC
10/5	Friday	1:30 – 3:00PM	W6015	MS & DA & AM
10/9	Tuesday	1:30 – 3:00PM	E6130	YSH & PK & DN
10/12	Friday	1:30 – 3:00PM	W6015	MS & DA & AM
10/16	Tuesday	1:30 – 3:00PM	E6130	YSH & PK & DN
10/19	Friday	1:30 – 3:00PM	W6015	MS & DA & AM
10/23	Tuesday	1:30 – 3:00PM	E6130	YSH & PK & DN
10/26	Friday	1:30 – 3:00PM	W6015	MS & DA & AM

Tuesday, September 4

8:30 – 10:20 **Introduction: Design and overview of analytical methods for cohort studies and analytical methods for incomplete observations in cohort studies** ***A. Muñoz***

Objectives: To describe the basic concepts that define different studies in epidemiologic research. To provide an overview of analytical methods for cohort studies.
To describe the incompleteness of censored observations and late entries in data from cohort studies.
To demonstrate how standard Kaplan-Meier methods are equivalent to completing the incomplete data.
To show how to extend methods to incorporate unseen individuals due to truncation implicit in late entries.

Reference: Muñoz A, Nieto FJ. Cohort Studies. In: Oxford Textbook of Public Health, 6th edition. Detels R, Gulliford M, Karim QA, Chuan Tan C (Eds). Oxford, UK, Oxford University Press, 2015.

Thursday, September 6

8:30 – 10:20 **Description of the data sets from the Chronic Kidney Disease in Children Study and the Multicenter AIDS Cohort Study** ***M. Schneider A. Muñoz***

Objective: To describe data conducive to implementation of methods to be discussed in the course towards the writing of report in the form of "Methods and Results" sections of paper which will be the basis for evaluation of students.

Monday, September 10

8:30 – 9:20 **Validation of procedures to simplify the design and conduct of cohort studies Part I: optimization of instruments** ***D. Ng***

Objectives: To describe enhancements of the Bland-Altman plot to assess agreement of continuous biomarkers.

References: Schwartz GJ, Furth S, Cole SR, Warady BA, Muñoz A. Glomerular filtration rate via plasma iohexol disappearance: pilot study for chronic kidney disease in children. Kidney Int 2006;69:2070-7.
Ng DK, Schwartz G J, Jacobson L P, Palella FJ, Margolick, J B, Warady BA, Furth SL, Muñoz A. Universal GFR determination based on two time points during plasma iohexol disappearance. Kidney Int 2011;80:423-30.
Schwartz GJ, Abraham A, Furth SL, Warady B, Muñoz A. Optimizing iohexol plasma disappearance curves to measure the glomerular filtration rate in children with chronic kidney disease. Kidney Int 2010;77:65-71.

Tuesday, September 11

8:30 – 10:20 **Estimation and testing of relative hazards: proportional hazard regression methods and extension to allow for non-proportionality** ***C. Cox***

Objective: Review the Cox proportional hazards model with special attention to time dependent covariates, both discrete and continuous, for the case of non-proportional hazards.

Reference: Uno H, Claggett B, Tian L, Inoue E, Gallo P, Miyata T, Schrag D, Takeuchi M, Uyama Y, Zhao L, Skali H, Solomon S, Jacobus S, Hughes M, Packer M, Wei L-J. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. J Clinical Oncology 2014; 22:2380-85.

Thursday, September 13

8:30 – 10:20 **Measures to compare time-to-event data: relative percentiles and relative hazards; and taxonomy of hazard functions of the generalized gamma distribution** ***M. Schneider***

Objectives: To describe the differences between the commonly used relative hazards and the relative percentiles or times as measures for comparing survival functions. To describe the power of the general gamma distribution as a framework to incorporate most types of hazard functions in epidemiologic research.

References: Schneider MF, Gange SJ, Williams CM, Anastos K, Greenblatt RM, Kingsley L, Detels R, Muñoz, A. Patterns of the hazard of death after AIDS through the evolution of antiretroviral therapy: 1984-2004. AIDS 2005;19:2009-18.
Cox C, Chu H, Schneider M, Muñoz A. Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. Stat Med 2007;26:4352-74.

Monday, September 17

8:30 – 9:20 **Laboratory 1: semi-parametric regression models with emphasis on time dependent relative hazards: Modeling incident CHD: The Atherosclerosis Risk in Communities (ARIC) Study, 1987-94** ***M. Schneider***

In this lab, we consider two important risk factors for incident coronary heart disease, BMI and smoking, both treated as categorical variables. Kaplan-Meier survival curves are compared using the log rank test. Univariate and multivariate Cox proportional hazards are also fitted to the data. Results show evidence of non-proportional hazards; this question is examined using time-varying effects in the Cox model.

Tuesday, September 18

8:30 – 9:45 **Parametric survival analysis based on the generalized gamma distribution** ***C. Cox***

Objectives: To indicate how to incorporate risk factors into parametric regression models. To illustrate the use of relative percentiles (times) as a measure of the effect of risk factors.

Reference: Cox C, Chu H, Schneider M, Muñoz A. Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. Stat Med 2007;26:4352-74.

9:50 – 10:20 **Student presentations**

Thursday, September 20

8:30 – 9:45 **Parametric survival analysis based on the generalized gamma distribution (cont.)** ***A. Muñoz***

Objective: To present and run Stata and R code to reproduce Table II and Figure 5 (relative times) and Figure 6 (relative hazards) in Statistics in Medicine 2007 paper by Cox, Chu, Schneider, Muñoz.

Reference: Cox C, Chu H, Schneider M, Muñoz A. Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. Stat Med 2007;26:4352-74.

9:50 – 10:20 **Student presentations**

Monday, September 24

8:30 – 9:20 **Laboratory 2: Generalized Gamma models with age as the time scale** ***D. Antiporta***

Objective: Using data from the Epidemiologic Catchment Area Program with more than 25 years of follow-up for 15,440 individuals with 6,924 recorded deaths, data were analyzed using age as the time scale and parametric approaches to quantify the heterogeneity of relative times. The pitfalls of proportional hazards assumption are shown for the description of the effect of race on survival.

Reference: Eaton WW, Roth KB, Bruce M, Cottler L, Wu LT, Nestadt G, Ford D, Bienvenu J, Crum RM, Rebok G, Anthony J, Muñoz A. The Relationship of Mental and Behavioral Disorders to All-cause Mortality in a 27-year Follow-up of Four Epidemiologic Catchment Area Samples. Am J Epidemiol 2013;178:1366-77.

9:30 – 10:20 **Student presentations (if needed)**

Tuesday, September 25

8:30 – 9:45 **Random effects models for longitudinal data** ***C. Cox***

Objectives: To describe models that allow for different individuals to have different intercepts and slopes.
To characterize the decline of $CD4^+$ lymphocytes in HIV infected individuals according to baseline HIV-RNA.

Reference: Mellors JW, Muñoz A, Giorgi JV, Margolick JB, Tassoni CJ, Gupta P, Kingsley LA, Todd JA, Saah AJ, Detels R, Phair JP,Rinaldo CR, Jr. Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection. Ann Intern Med 1997; 126:946-54.

9:50 – 10:20 **Student presentations**

Thursday, September 27

8:30 – 9:45 **Empirical Bayes estimates from random effects models for longitudinal data** ***A. Muñoz***

Objectives: To cast the random effects models from the perspective of empirical Bayes methods. To derive and calculate the posterior distribution of intercepts and slopes of individuals' regression lines.

Reference: Mellors JW, Margolick JB, Phair JP, Rinaldo CR, Detels R, Jacobson LP, Muñoz A. Prognostic value of HIV-1 RNA, CD4 Cell count slope for progression to AIDS and death in untreated HIV-1 infection. JAMA 2007;297:2349-50.

9:50 – 10:20 **Student presentations**

Monday, October 1

8:30 – 9:20 **Laboratory 3: Linear mixed models for effects of proteinuria on longitudinal GFR decline in children with CKD** ***YS. Hong***

Using random effects models to describe the effect of presence of protein in the urine on decline of kidney function, it is found that the association is strong but it is different according to the type of CKD diagnosis. The differences are not only on the estimates of decline but also in the variance components.

9:30 – 10:20 **Student presentations (if needed)**

Tuesday, October 2

8:30 – 9:45 **Models for joint analysis of longitudinal and survival data** ***C. Cox***

Objectives: To understand the structure of a joint model for two different outcomes, with particular attention to joint modeling of longitudinal and survival data using a shared parameter model. To illustrate the use of shared parameter models to adjust for informative dropout in a longitudinal cohort study.

Reference: Kogon A, Pierce CB, Cox C, Brady TM, Mitsnefes M, Warady BA, Furth SL, Flynn JT. Nephrotic-range proteinuria is strongly associated with poor blood pressure control in pediatric chronic kidney disease. Kidney Int 2014;85:938-944.

9:50 – 10:20 **Student presentations**

Thursday, October 4

8:30 – 9:45 **Models for joint analysis of longitudinal and survival data (cont.)** ***A. Muñoz***

9:50– 10:20 **Student presentations**

Monday, October 8

8:30 – 9:20 **Laboratory 4: Joint models to describe the predictive value of proteinuria on GFR decline adjusting for time to ESRD (RRT or 50% decline in GFR)** ***C. Cox***

Standard methods for dealing with missing data, such as multiple imputation, assume that that data are missing at random (MAR). In some cohort studies, however, dropouts occur because of disease progression, and in this situation, there may be informative missingness, in which the probability of dropout depends on unobserved data. This lab considers models for longitudinal data with informative dropout. To account for informative dropout a joint model is used, which includes a survival model for time-to-dropout. The longitudinal and survival models are linked by shared random effects, which account for the informative missingness.

9:30 – 10:20 **Student presentations (if needed)**

Tuesday, October 9

8:30 – 9:45 **Inference for mutually exclusive competing events through a mixture of generalized gamma distributions** ***A. Muñoz***

Objectives: Time-to-event data with two or more types of endpoints are found in many epidemiological settings. Instead of treating the times for one of the endpoints as censored observations for the other, we present an alternative approach where we treat competing events as distinct outcomes in a mixture of two distributions.

References:
Cole SR, Li R, Anastos K, Detels R, Young M, Chmiel JS, Muñoz A. Accounting for leadtime in cohort studies: evaluating when to initiate HIV therapies. Stat Med 2004;23:3351-63.

Checkley W, Brower RG, Muñoz A. Inference for mutually exclusive competing events through a mixture of generalized gamma distributions. Epidemiology 2010;21:557-65.

9:50 – 10:20 **Student presentations**

Thursday, October 11

8:30 – 9:45 **The perilousness of proportional hazards in the competing risks setting** ***A. Muñoz***

Objectives: To characterize the tethering of relative hazards under the assumption of proportionality in the competing risks setting.
To show the perilousness of analysis where relative hazards change over time and results are reported under inappropriate proportionality assumptions.
To divulge the lack of specificity of cause-specific hazards.

References:
Muñoz A, Abraham AG, Matheson M, Wada N. Non-proportionality of hazards in the competing risks framework. In: Risk Assessment and Evaluation of Predictions. Lee MLT, Gail M, Pfeiffer R, Satten G, Cai T, Gandy A (Eds). Springer, 2013.

Muñoz A, Mongilardi N, Checkley W. Multilevel competing risks in the evaluation of nosocomial infections: time to move on from proportional hazards and even from hazards altogether. Crit Care 2014;18:146.

9:50 – 10:20 **Student presentations**

Monday, October 15

8:30 – 9:20 **Laboratory 5: Comparison of methods for the analysis of competing risks data** ***P. Kuo***

Objectives: To reproduce the analysis of the Fluid management Trial data in Checkley/Muñoz*, Epidemiology* 2010. Stata software to fit mixtures of generalized gamma distributions is introduced. Results from mixture models are compared to standard semiparametric approaches comparing cause-specific hazards and sub-hazards of the cumulative incidences.

Reference: Checkley W, Brower RG, Muñoz A. Inference for mutually exclusive competing events through a mixture of generalized gamma distributions. Epidemiology 2010;21:557-65.

9:30 – 10:20 **Student presentations (if needed)**

Tuesday, October 16

8:30 – 10:20 **STUDENT PRESENTATIONS**

Thursday, October 18

8:30 – 9:45 **Validation of procedures to simplify the conduct of cohort studies Part II: Longitudinal studies and epidemiologic inference** ***D. Ng***

Objectives: To identify and address challenges with instrument validation in long term longitudinal studies using agreement analyses
To present graphical methods to illustrate bivariate relationships for agreement
To consider different methods to assess agreement when multiple instruments purport to measure the same underlying construct

References:
Ng DK, Schwartz GJ, Schneider MF, Furth SL, Warady BA. Combination of pediatric and adult formulas yield valid glomerular filtration rate estimates in young adults with a history of pediatric chronic kidney disease. Kidney Int. 2018.

Ng DK, Schwartz GJ, Warady BA, Furth SL, Muñoz A. A comparison of relationships of measured iohexol GFR and estimated GFR with CKD-related biomarkers in children and adolescents. American Journal of Kidney Diseases, 2017, 6386(17)30637-6. PMID: 28549535

Li X, Buechner JM, Tarwater PM, Muñoz A. A Diamond-Shaped Equiponderant Graphical Display of the Effects of Two Categorical Predictors on Continuous Outcomes. *The American Statistician*. 2003;57(3):193-199. http://www.jstor.org/stable/30037269.

9:50 – 10:20 **Student presentations**

Monday, October 22

8:30 – 9:20 **Time-varying coefficient of determination to quantify the explanatory power of biomarkers on continuous outcomes** ***D. Ng***

Objectives: Epidemiologists are often interested in how much variability in an outcome is explained by a predictive biomarker. In the context of longitudinal repeated measurements, the explanatory power may change over time and can provide valuable insight into the epidemiology of the disease process. The purpose of this lecture is to describe time-varying coefficients of determination, their derivation from the variance components of linear mixed effects models and to demonstrate the limitations of time-fixed estimates. The lecture provides an applied example in which a classical risk factor and a novel risk factor for kidney disease progression are compared, and how underlying diagnosis modifies these relationships.

Reference: Ng DK, Portale AA, Furth SL, Warady BA, Muñoz A. Time-varying coefficient of determination to quantify the explanatory power of biomarkers on longitudinal GFR among children with chronic kidney disease. Annals of Epidemiology. 2018

9:30 – 10:20 **Student presentations (if needed)**

Tuesday, October 23

8:30 – 9:45 **Inverse probability methods to characterize racial differences in the setting of competing risks** ***D. Ng***

Objectives: To introduce and illustrate the use of inverse probability methods for inference in complex data from cohort studies.

Reference: Ng DK, Moxey-Mims M, Warady BA, Furth SL, Muñoz A. Racial differences in renal replacement therapy initiation among children with a non-glomerular cause of chronic kidney disease. Ann Epidemiol 2016;26:780-87

9:50- 10:20 **Student presentations**

Thursday, October 25

8:30 – 10:20 **Life after AMDACS: Quantification of the reduction of age at first acute respiratory infection by exposure to particulate matter** ***E. Gurley***

Objectives: To show the usefulness of the Generalized Gamma model to determine the magnitude by which exposure to particulate matter reduces the life span free of acute lower respiratory infections

Reference: Gurley ES, Salje H, Homaira N, Ram PK, Haque R, Petri WA, Bresee J, Moss WJ, Luby SP, Breysee P, Azziz-Baumgartner E. Indoor exposure to particulate matter and age at first acute lower respiratory infection in a low-income urban community in Bangladesh. Am J Epidemiology 2014; 179: 967-73. [Selected as one of the 10 best articles published in AJE in 2014].