

## PHA 6805

### Applied Data Analysis, Interpretation and Reporting of Findings in Pharmacy

#### Instructors:

Jingchuan Serena Guo, MD, PhD  
Assistant Professor, Department of Pharmaceutical Outcomes & Policy  
University of Florida College of  
Pharmacy E-mail: [guoj1@ufl.edu](mailto:guoj1@ufl.edu)  
Office Hour: By appointment

Hui Shao, MD, PhD  
Assistant Professor, Department of Pharmaceutical Outcomes & Policy  
University of Florida College of  
Pharmacy E-mail: [hui.shao@ufl.edu](mailto:hui.shao@ufl.edu)  
Office Hour: By appointment

#### Teaching Assistant:

Dawei Guan, MD MS  
PhD Candidate, Department of Pharmaceutical Outcomes & Policy  
University of Florida College of  
Pharmacy E-mail: [daweiguan@ufl.edu](mailto:daweiguan@ufl.edu)  
Office Hour: By appointment

**Class Meetings:**       Lecture: Monday, 3:00-4:30pm, HPNP 2309  
                                  Lab: Wednesday, 8:30am-10:00am, HPNP 2309

**Course Prerequisites:**   STA 6166 and 6167, or permission of the instructor.

#### Recommended Reference Textbooks:

Gelman A, Hill J. Data Analysis Using Regression and Multilevel/Hierarchical Models. 2007; New York: Cambridge University Press.

Gillingham M. SAS Programming with Medicare Administrative Data. 2014. SAS® Press.

Delwiche and Slaughter. The Little SAS® Book: A Primer, Fifth Edition. SAS® Press.

**Recommended Reference Materials:** SAS – IDRE Stats – UCLA website (includes many tutorials and code for statistical analyses. Simple Google of, “logistic regression SAS ucla” will pull up many great links, for example.)

#### Course Description:

The purpose of this course is to develop the skills students need to ask good research questions, plan and design studies, present a clear and reproducible protocol, develop

analytical datasets from raw data, test hypotheses that answer the research questions, and report, interpret, and discuss the findings. The course consists of two components: 1) lectures, readings, and activities focused on the conceptual and practical aspects of data analysis; and 2) hands-on laboratory and supplemental lectures regarding types of data, analysis and management of data, and reporting of methods and results.

A major assumption of this course is that students already have taken statistics or biostatistics courses and have introductory SAS programming skills. The primary focus of this course is **NOT** to re-teach statistics or programming, but to ensure that the student can correctly apply logic to programming challenges and apply appropriate statistical techniques that match the research question.

### **Course Structure:**

The course will consist of two different sections: a) lectures over data analysis and concepts and b) experiential lab and project workshop sessions. Schedules of each are below. A project will encompass the entire class and be evidence for concepts learned throughout the course. A final protocol, report, and presentation will be required and described by the instructor. Two exams will assess your conceptual knowledge and class presentations and peer grading will evaluate your ability to work and communicate in a team environment.

### **Research dataset(s)**

Each student is required to have a dataset (selected in consultation with their academic advisor and the course instructors) to work with. The instructors will work with students who do not have appropriate data sets available. In general, administrative claims data, including both MarketScan and Medicare are NOT encouraged or supported for this course. Examples of available and “easy to use” databases are Nationwide Readmissions Database, NHANES, NAMCS, BRFSS, MEPS, and so on. We have access to such databases on our research servers and many can be freely accessed and downloaded to your personal computer without relying on ResVault or other online resources.

### **Student Evaluation:**

- Full study protocol development
- Student presentation of assignments and project progress reports
- Peer evaluations of code, presentations, and overall class participation
- Final study report including SAS code
- Final exam

Final grades will be assessed on the basis of the following course requirements:

**Project Protocol Development:** 15 points

**Class Presentations:** 20 points

**Participation in Teamwork: (Anonymous Peer Evaluation):** 15 points

**Final Study Report:** 30 points

**Final Exam:** 20 points

*Students who audit or retake the course are expected to follow all class regulations and complete all work mentioned above.*

Scores from each of the assignments will be combined to calculate the final grade. Letter grades will be assigned according to the following scheme. Grades will be given only with whole numbers; thus, no discussion about “rounding up” or extra credit will be entertained.

Points earned	100-93	92-90	89-87	86-83	82-80	79-77	76-73	72-70	69-67	66-63	60-62	<60%
Letter grade	A	A-	B+	B	B-	C+	C	C-	D+	D	D-	E

**Attendance and Participation:** Students are fully responsible for any materials missed during any absence. Your attendance is mandatory AND your participation is necessary. In-person attendance is expected and required until further notice. Zoom connection is provided for those with concerns regarding in-person attendance or those who are self-isolating. All lectures will be recorded.

**Academic Dishonesty:** Familiarize yourself with the University's policy regarding academic dishonesty. See the Statements regarding the Student Conduct Code on page 11 of the 1999-2000 Graduate Catalog. This policy will be strictly enforced. The University's conduct regulations are available on the Internet at <https://www.dso.ufl.edu/sccr>.

\*Class schedules and topics are subject to change based on course needs

\*\*All deliverables are due at 11:59 pm on the day listed unless noted otherwise

**Lecture Session, Monday 3-4:30pm**

Date	Topic	Instructor	Deliverable**
W2	August 29	Intro to MEPS and NHANES database	Guo, Shao
W3	September 5 (holiday)	No class	
W4	September 12	Data Types Overview: Administrative Claims data; EHR; Survey data	Guo
W5	September 19	Analytic Cohort Curation	Zhiying “Julie” Zhaoa
W6	September 26	Reporting scientific data, best practices, journal formatting	Guo

W7	October 3	Linear regression	Guo	
W8	October 10	Logistic regression	Guo	
W9	October 17	Survival analysis	Guo	
W10	October 24	Incorporating time, time series, and repeated measures	Guo	
W11	October 31	Peeking the Natural of Statistics through the Lens of Simulation (Part 1)	Shao	
W12	November 7	Peeking the Natural of Statistics through the Lens of Simulation (Part 2)	Shao	
W13	November 14	Marginal Effect and Model Interpretation	Shao	
W14	November 21	GLM: Poisson Regression	Shao	
W15	November 28	Generalized Linear Model (Part 2)	Shao	
W15	December 5	Exam	Guo/ Shao	
W16	December 12	Student presentations		

**Lab: Wednesday, 8:30am – 10am**

Date		Topic	Instructor	Deliverable**
W1	August 24	Class introduction, SAS Studio Sign-up	Guo/Shao	Team up
W2	August 31	Data management & efficiency	Carl Henriksen	Project idea submitted to faculty
W2	September 7	Introduction to SQL and SAS macros	Carl Henriksen	
W4	September 14	Student project concept	Guo, Shao	

		presentations (5 mins)		
W5	September 21	Analytic cohort curation	Zhiying “Julie” Zhao	
W6	September 28	Creating Table 1,.	Guo	Analytical protocol due
W7	October 5	Linear regression	Guo	
W8	October 12	Logistic regression	Guo	
W9	October 19	Survival analysis	Guo	
W10	October 26	Incorporating time, time series, and repeated measures	Guo	
W11	November 2	Exercise: Using sampling to generate population estimates	Shao	
W12	November 9	Exercise: the impact of endogeneity in regression estimates	Shao	
W13	November 16	Generalized Linear Model (Part 1)	Shao	
W14	November 23 (holiday)		No class	
W15	November 30	Generalized Linear Model: Final Mission	Shao	
W15	December 7	Prepare for the Presentation and final report		Peer evaluation due (Programming and draft final report)
W16	December 14 (TBD)	Final report due		