BIOS 777: Precision Medicine (Fall 2025)

Instructor: Professor Michael R. Kosorok, 3105H McGavran-Greenberg (McGG).
Grader (for computer programming assignments): Yating Zou.
Additional Graders (for assistance with lecture content): Molly Hoch and Hong Niu.
Class Information: on the canvas website.
Time and location: Mondays and Wednesdays, 9:05-10:20am, in room 228 Rosenau Hall.
Instructor office hours: Mondays 3:00-4:30PM. Room McGG 3105H.
Grader office hours: Tuesdays 2:00-3:00PM. Room McGG 1302. Only available on the following days: September 9, 16, 23 and 30; October 14, 21 and 28; and November 4, 11 and 18.
Prerequisites: BIOS 661 and 663 or equivalent.

Summary: Precision medicine seeks to maximize the quality of healthcare by individualizing the healthcare process to the uniquely evolving health status and circumstances of each patient. This endeavor spans many scientific disciplines, including biomedical science, genetics, statistical science, computer science, machine learning, psychology, operations research, and many other areas, all in support of evidence-based, data-driven decision support.

In this course, we will address precision medicine (PM) from a statistical and machine learning perspective with numerous examples of application. We will develop a working knowledge of the following inter-related areas in the context of precision medicine and precision health: dynamic treatment regimes, causal inference for precision medicine; basic machine learning tools including support vector machines, random forests and neural nets; the single decision setting, outcome weighted learning and extensions of outcome weighted learning; the multi decision setting, reinforcement learning, inverse reinforcement learning, sequential multiple assignment randomized trials (SMARTs), mobile health, micro-randomized trials, and Markov decision processes; advanced machine learning including deep learning, generative AI, and large language models; relevant statistical inference issues; and several topics on the frontiers of the area. There will be homework assignments, short papers and projects but no exams.

Course requirements: The grade for the course will depend on regular homework, consisting of six short assignments (S1–S6), three computer programming assignments (C1–C3), two mini-projects (P1–P2), and a final project (FP). All assignments must by typed (not handwritten) in either latex (preferred) or other word processing platform. Page limits given in the assignment descriptions do not include figures or plots. Each short assignment is worth 5 points, totaling 30 points; each computing assignment is worth 10 points, totaling 30 points; each mini-paper is worth 10 points, totaling 20 points; and the final project is worth 20 points. The six short assignments, two mini-projects and final project are described below. Details for each of the computer programming assignments will be provided at least two weeks before their due date, where the due date is given in the following schedule. Thus the total points possible are 100. All components of the final project, and any late assignments which have been approved for being turned in late, are all due by Wednesday, December 1, 2021, by 5:00pm. Grades will be determined on the basis of the number of points achieved (out of 100 possible): H (90–100%), P (80–90%), L (70–80%), and F (less than 70%). There will be no in-class exams.

Schedule of topics, readings, and due dates given below. The assignments are due on Wednesdays at the beginning of class.

Section	Date	Topic	Readings	Due
1		*Overview and Background*		
1.1	Aug. 18	Introduction to PM	Kosorok and Laber (2019)	
	20	”		
1.2	25	Causal Inference	Hernán & Robins (2020) Ch. 6, 15 Little & Rubin (2000)	
1.3	27	Machine Learning	Hastie et al (2011) Ch. 12, 15	S1
	Sep. 3	”	Zhu & Kosorok (2012), Wager & Athey (2018)	
2		*Single Decision DTRs*		
2.1	8	Regression and Policy Methods	Zhang et al (2012), Athey and Wager (2018)	
2.2	10	Outcome Weighted Learning	Zhao et al (2012), (2019)	S2
2.3	17	Extensions	Chen et al (2016)	S3
	22	”	Cui et al (2018), Laber and Staicu (2018)	
	24	”	Kallus (2017), Zhang et al (2018)	P1
3		*Multi-Decision DTRs*		
3.1	29	Reinforcement Learning	Schulte et al (2014)	
	Oct. 1	Q-learning	Zhao et al (2011), Moodie et al (2012)	C1
	Oct. 6	”	Goldberg and Kosorok (2012), Zhao et al (2015)	
3.2	8	SMART Designs	Kidwell (2016), Hibbard et al (2018)	S4
	13	”	Rose et al (2019+)	
3.3	15	Micro-Randomized Trials	Klasnja et al (2015)	P2
3.4	20	Markov Decision Processes	Ertefei and Strawderman (2018)	
	22	”	Luckett et al (2020)	S5
4		*Deep Learning*		
4.1	27	Introduction to DNNs	LeCun et al (2015)	
4.2	29	Generative AI without LLMs	Goudet et al (2018)	C2
4.3	Nov. 3	Large Language Models (LLMs)	Wilson (2014), Ch. 2, Herlands et al (2018)	
	5	”	Liang et al (2018), Wang et al (2018)	(FP Topic)
4.4	10	LLM Applications		
5		*Advanced Topics*		
5.1	12	Multiple Utilities/Inverse RL	Butler et al (2018),	S6
	17	”	Luckett et al (2021)	
5.2	19	Value Function Inference	Laber and Qian (2018),	C3
	24	”	Jiang et al (2021), Lu et al (2025+)	Slides
6		*Conclusion*		
6.1	Dec. 1	Student Presentations		
	3	”		FP

Short Assignments:

S1: Due Wednesday, August 27. Please define precision medicine as well as its main goals and features as though your were writing to a friend or relative who is not a professional scientist. Please include important concepts such as heterogeneity of patients and treatments. Please restrict the answer to 1 page or less.

S2: Due Wednesday, September 10. Please describe random forests to a non-expert, explaining how they are computed and used. Discuss strengths and weaknesses relative to deep neural nets for both classification and regression. Please use 1–2 pages.

S3: Due Wednesday, September 17. Please compare and contrast the advantages and disadvantages of policy learning approaches to estimating individualized treatment rules versus other approaches. Provide several examples of settings where a policy learning approach should be used. Include citations. Please use 1–2 pages.

S4: Due Wednesday, October 8. Please define finite horizon, off-policy reinforcement learning for a friend or relative who is not a professional scientist but who has had an introductory course in statistics which included the basic concepts of regression modeling. Explain why off-policy learning is important to biomedical research on humans and describe how to perform Q-learning. Please use 1–2 pages.

S5: Due Wednesday, October 22. Please define infinite horizon, off-policy reinforcement learning for a non-expert. Give examples of biomedical research settings where such methodology is needed, including when a Markov Decision Process setting is appropriate. Compare and contrast Q-learning and V-learning in this context, and give examples where reinforcement learning is not needed for analysis of data from micro-randomized trials. Please use 1–2 pages.

S6: Due Wednesday, November 5. Please describe deep learning, the transformer architecture, and large language models to a non-expert. Explain the advantages and disadvantages of using large language models compared to more traditional machine learning approaches. Also describe at least two potential dangers from misuse of artificial intelligence. Please use 1–2 pages.

Computer Programming Assignments: These will be provided on the class web site and/or in class at least two weeks before they are due. We will be using R (see `https://www.r-project.org/`) as our main programming language as well as the package DynTxRegime (see `https://www2.cscc.unc.edu/impact7/DynTxRegime`).

Mini-Projects and Final Project:

P1: Due Wednesday, September 24. For this project, select a realistic disease setting and write the statistical part of a grant proposal to NIH for a reanalysis of a clinical trial for the purpose of discovering an individualized treatment rule for a single decision setting, including the background, specific aims (goals of the study), study design, analysis plan and (optionally) sample size justification. Please keep within 2–3 pages total.

P2: Due Wednesday, October 15. For this project, select a realistic disease setting and write the statistical part of a grant proposal to NIH for a SMART study to estimate a sequence of individualized treatment rules for a multi-decision setting, including the background, specific aims (goals of the study), study design, analysis plan and (optionally) sample size justification. Please include a schematic of your proposed SMART design. Please keep within 2–3 pages total.

FP: For this project, choose a recent statistical article on precision medicine and/or precision health (published 2021 or later, either peer-reviewed or on arXiv), that is not one of the required readings for the course, and do one (or both) of the following: (1) verify the proof of a non-trivial theorem in the article and describe an open theoretical research question inspired by the theoretical aspects of the paper, and/or (2) identify a practical limitation in the proposed methodology and describe an open methodological research question inspired by the paper. Please keep your paper within 2–3 pages. Also, prepare 5 slides summarizing something interesting about your study of this paper and give a five minute presentation to the class towards the end of the semester (the time slot for this will be assigned to you). The paper, slides, and presentation are all required components of the final project. In addition, you need to select your paper and send the citation to me sometime between October 22 and November 5. I will be keeping track of selected papers to avoid duplication. If your paper has already been selected, you will need to select a different paper prior to November 10. I will be keeping a running list of topics to assist with this process. Slides are due on Monday, November 24. The paper (and any other outstanding assignments) will be due Wednesday, December 3.

Required Reading References: The following papers and books are required reading for the class. The class period for which the reading needs to be completed by is given in the schedule above. For books, it is only necessary to read the chapters designated in the schedule.

1. Athey S and Wager S (2021). Policy learning with observational data. *Econometrica* 89:133–161.

2. Butler EL, Laber EB, Davis SM, and Kosorok MR (2018). Incorporating patient preferences into estimation of optimal individualized treatment rules. *Biometrics* 74:18–26.

3. Chen G, Zeng D, and Kosorok MR (2016). Personalized dose finding using outcome weighted learning (with discussion and rejoinder). *Journal of the American Statistical Association* 111:1509–1547.

4. Cui Y, Zhu R, and Kosorok MR (2017). Tree based weighted learning for estimating individualized treatment rules with censored data. *Electronic Journal of Statistics* 11:3927–3953.

5. Ertefaie A and Strawderman RL (2018). Constructing dynamic treatment regimes over indefinite time horizons. *Biometrika* 105:963–977.

6. Goudet O, Kalainathan D, Callou P, Guyon I, Lopez-Paz D, and Sebag M (2018). Causal generative neural networks. arXiv:1711.08936v2 [stat.ML]

7. Goldberg Y and Kosorok MR (2012). Q-learning with censored data. *Annals of Statistics* 40:529–560.

8. Hastie T, Tibshirani R, and Friedman J (2011). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd Ed., Springer: New York.

9. Herlands W, Neill DB, Nickisch H, and Wilson AG (2018). Change surfaces for expressive multidimensonal changepoints and counterfactual prediction. arXiv:1810.11861v2 [stat.ML]

10. Hernán MA and Robins JM (2020). *Causal Inference: What If.* Boca Raton: Chapman & Hall/CRC.

11. Hibbard JC, Friedstat JS, Thomas SM, Edkins RE, Hultman CS, and Kosorok MR (2018). LIBERTI: A SMART study in plastic surgery. *Clinical Trials* 15:286–293.

12. Jiang X, Nelson AE, Cleveland RJ, Beavers DP, Schwartz TA, Arbeeva L, Alvarez C, Callahan LF, Messier S, Loeser R, and Kosorok MR (2021). A precision medine approach to determine optimal exercise and weight loss treatments for overweight and obese adults with knee osteoarthritis. *Arthritis Care & Research* 73:693–701.

13. Kallus N (2018). Balanced policy evaluation and learning. *Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS)* 8909–8920.

14. Kidwell KM (2016). DTRs and SMARTs: definitions, designs, and applications. In: Kosorok MR and Moodie EEM (Eds), *Adaptive Treatment Strategies in Practice: Planning Trials and Analyzing Data for Personalized Medicine*, 7-23. ASA-SIAM: Alexandria, VA, and Philadelphia.

15. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, and Murphy SA (2015). Micro-randomized trials: An experimental design for developing just-in-time interventions. *Health Psychology* 34:1220–1228.

16. Kosorok MR and Laber EB (2019). Precision medicine. *Annual Reviews of Statistics and Its Application* 6:263–286.

17. Laber EB and Qian M (2018). Generalization error for decision problems. arXiv:1812.08696 [stat.ME]

18. Laber EB and Staicu AM (2018). Functional feature construction for personalized dynamic treatment regimes. *Journal of the American Statistical Association* 113:1219–1227.

19. LeCun Y, Bengio Y, and Hinton G (2015). Deep learning. *Nature* 521:436–444.

20. Liang M, Ye T, and Fu H (2018). Estimating individualized optimal combination therapies through outcome weighted deep learning algorithms. *Statistics in Medicine* 37:3869–3886.

21. Little RJ and Rubin DB (2000). Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Annual Reviews of Public Health* 21:121–45.

22. Luckett DJ, Laber EB, Kahkoska AR, Maahs DM, Mayer-Davis E, and Kosorok MR (2020). Estimating dynamic treatment regimes in mobile health using V-learning. *Journal of the American Statistical Association* 115:692–706.

23. Luckett DJ, Laber EB, Kim S, and Kosorok MR (2021). Estimation and optimization of composite outcomes. *Journal of Machine Leaerning Research* 22(167):1–40.

24. Moodie EEM, Chakraborty B, and Kramer MS (2012). Q-learning for estimating optimal dynamic treatment rules from observation data. *Canadian Journal of Statistics* 40:629–645.

25. Rose EJ, Laber EB, Davidian M, Tsiatis AA, Zhao YQ, and Kosorok MR (2019+). Sample size calculations for SMARTs. arXiv:1906.06646v1 [stat.ME]

26. Schulte PJ, Tsiatis AA, Laber EB, and Davidian M (2014). Q- and A- learning methods for estimating dynamic treatment regimes. *Statistical Science* 29:640–661.

27. Wager S and Athey S (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association* 113:1228–1242.

28. Wang L, Laber EB, and Witkiewitz K (2018). Sufficient Markov decision processes with alternating deep neural networks. arXiv:1704.07531v2 [stat.ME]

29. Wilson AG (2014). *Covariance Kernels for Fast Automatic Pattern Discovery and Exploration with Gaussian Processes*. PhD Dissertation. University of Cambridge.

30. Zhang B, Tsiatis AA, Laber EB, and Davidian M (2012). A robust method for estimating optimal treatment regimes. *Biometrics* 68:1010–1018.

31. Zhang Y, Laber EB, Davidian M, and Tsiatis AA (2018). Interpretable dynamic treatment regimes. *Journal of the American Statistical Association* 113:1541–1549.

32. Zhao YF, Zeng D, Socinski MA, and Kosorok MR (2011). Reinforcement learning strategies for clinical trials in non-small cell lung cancer. *Biometrics* 67:1422–1433.

33. Zhao YQ, Laber EB, Ning Y, Saha S, and Sands BE (2019). Efficient augmentation and relaxation learning for individualized treatment rules using observational data. *Journal of Machine Learning Researcy* 20:1–23.

34. Zhao YQ, Zeng D, Rush AJ, and Kosorok MR (2012). Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association* 107:1106–1118.

35. Zhao YQ, Zeng D, Laber EB, and Kosorok MR (2015). New statistical learning methods for estimating optimal dynamic treatment regimes. *Journal of the American Statistical Association* 110:583-598.

36. Zhu R and Kosorok MR (2012). Recursively imputed survival trees. *Journal of the American Statistical Association* 107:331–340.

Supplemental References: The following references are not part of the required reading but may be helpful:

1. Chakraborty B and Moodie EEM (2013). *Statistical Methods for Dynamic Treatment Regimes: Reinforcement Learning, Causal Inference, and Personalized Medicine.* Springer: New York.

2. Dawid AP (2015). Statistical causality from a decision-theoretic perspective. *Annual Reviews of Statistics and Its Application* 2:273–303.

3. Goodfellow I, Bengio Y, and Courville A (2016). *Deep Learning.* MIT Press: Cambridge, MA.

4. Kosorok MR and Moodie EEM (Eds) (2016). *Adaptive Treatment Strategies in Practice: Planning Trials and Analyzing Data for Personalized Medicine.* ASA-SIAM: Alexandria, VA, and Philadelphia.

5. Rasmussen CE and Williams CKI (2006). *Gaussian Processes for Machine Learning.* MIT Press: Cambridge, MA.

Attendance and Participation: Your attendance and active participation are an integral part of your learning experience in this course. If you are unavoidably absent, please notify the course instructor (and Teaching Assistant if one is assigned). No right or privilege exists that permits a student to be absent from any class meetings, except for these University Approved Absences:

1. Authorized University activities
2. Disability/religious observance/pregnancy, as required by law and approved by Accessibility Resources and Service (`https://ars.unc.edu`) and/or the Equal Opportunity and Compliance Office (`https://eoc.unc.edu/what-we-do/accommodations/`).
3. Significant health condition and/or personal/family emergency as approved by the Office of the Dean of Students (`https://odos.unc.edu/`), Gender Violence Service Coordinators (`https://womenscenter.unc.edu/resources/gender-violence-services/`), and/or the Equal Opportunity and Compliance Office (`https://eoc.unc.edu/what-we-do/accommodations/`).

Code of Conduct: To ensure effective functioning of the UNC Code of Conduct (see `https://studentconduct.unc.edu/student-code-of-conduct/`), students are expected to:

- Conduct all academic work within the letter and spirit of the Honor Pledge and University policy, which prohibits the giving or receiving of unauthorized aid in all academic processes.
- Learn the recognized techniques of proper attribution of sources used in written work; and to identify allowable resource materials or aids to be used during completion of any graded work.
- You are welcome and encouraged to discuss course materials and work on assignments in groups to help each other understand the concepts better, but all assignments (short assignments, programming assignments, mini and final projects) need to be written and completed individually on your own and with your own words. Plagiarism is not allowed in this course.

If you have any questions about your rights and responsibilities, consult the UNC Code of Student Conduct (`https://studentconduct.unc.edu/student-code-of-conduct/`) or review the following resources: Honor System (`https://studentconduct.unc.edu/honor-system`; UNC Librarys plagiarism tutorial (`https://guides.lib.unc.edu/plagiarism`); UNC Writing Centers handout on plagiarism (`https://writingcenter.unc.edu/tips-and-tools/plagiarism/`).

Artificial Intelligence (AI) Use Policy: Carolina students are expected to follow these AI guidelines:

- AI should help you think, not think for you. You may be able to use these tools to brainstorm ideas, research topics, and analyze problems, but you must decide whats appropriate and accurate.
- Engage responsibly with AI. You must evaluate AI-generated outputs for potential biases, limitations, inaccuracies, false output, and ethical implications. Do not put personal or confidential data into these tools.
- The use of AI must be open and documented. You should declare, explain, and cite any use of AI in the creation of your work using applicable standards (e.g., APA, MLA, course guidelines). Understand that you are ultimately 100% responsible for your final product.
- Follow strictly at all times the UNC Code of Conduct.

Gillings Community: Shared values such as mutual respect, intellectual humility, interdisciplinary collaboration, and commitment to public health ethics form the foundation of our teaching and learning, research, public health practice, professional networks, and community engagement. In this spirit, we strive to foster an environment across the school that welcomes and values all individuals while supporting their development and success.

Additional campus resources include: UNC Student Affairs (`https://studentaffairs.unc.edu/`); including the LGBTQ Center (`https://lgbtq.unc.edu/`); Non-Discrimination Policies at UNC Chapel Hill (`https://eoc.unc.edu/our-policies/policy-statement-on-non-discrimination/`); Ombuds (`https://ombuds.unc.edu/`); Prohibited Discrimination, Harassment, and Related Misconduct at UNC Chapel Hill (`https://eoc.unc.edu/our-policies/ppdhrm/`).

Technical Support: The best way to help prevent technical issues from causing problems for assignments and quizzes is to submit them at least 24-36 hours before the due date and time. Your instructor cannot resolve technical issues, but its important to notify them if you are experiencing issues. If you have problems submitting an assignment or taking a quiz in Canvas, immediately do the following:

1. Contact the UNC Information Technology Services (ITS) department with the time you attempted to do your course action and what the course action was.

2. Email your instructor with the information you sent to ITS and what time you sent the information.

The ITS department provides technical support 24-hours per day, seven days per week. If you need computer help, please contact the ITS Help Desk by phone at +1-919-962-HELP (4357), or by online help request (`https://tdx.unc.edu/TDClient/33/Portal/Home/`).

Additional Resources and Policies: Please see the Syllabus Expectations, Policies and Resources webpage (`https://sph.unc.edu/resource-pages/syllabus-expectations-policies-and-resources/`) for additional important information and resources on the following topics:

- Equal Opportunity and Compliance Accommodations
- University Attendance Policy
- Course Communication Expectations
- Counseling and Psychological Services (CAPS) at UNC Chapel Hill
- Appropriate Use of Course Resources
- Policy on Non-Discrimination
- Student Feedback and Equity Concerns
- Safety and Emergency Information at Gillings
- Title IX and Related Resources

Syllabus Changes: The instructor reserves the right to make changes to the syllabus including project due dates and test dates. These changes will be announced as early as possible.