MODULE SPECIFICATION

Academic Year (student cohort covered by specification)	2021-22
Module Code	2493
Module Title	Analysis of Electronic Health Record data
Module Organiser(s)	Professor Elizabeth Williamson Professor Ruth Keogh
Faculty	Epidemiology & Population Health
FHEQ Level	Level 7
Credit Value	**CATS:** 15 **ECTS:** 7.5
HECoS	100246: 100366: 101031
Term of Delivery	Term 2
Mode of Delivery	For 2021-22, this module is currently planned as a mixture of online and face to face teaching. There will be a combination of live and interactive activities (synchronous learning) as well as recorded or self-directed study (asynchronous learning).
Mode of Study	Full-time
Language of Study	English
Pre-Requisites	Statistics for Health Data Science (or equivalent) Programming (R Sessions) Epidemiology for Health Data Science (or equivalent) A knowledge of linear regression, analysis of variance, logistic regression, maximum likelihood estimation and simple methods of analysing quantitative and categorical data is essential (t-test, RR, OR). Understanding of basic epidemiological design is important. Familiarity with R is needed.
Accreditation by Professional Statutory and Regulatory Body	Not currently accredited by any other body.
Module Cap (Indicative number of students)	Maximum of 30 students
Target Audience	This is a recommended module for the MSc Health Data Science.

Module Description	Electronic Health Record (EHR) data are complex, messy and large. Data are ongoingly recorded over time, for the purposes of clinical care. As such, they raise complex analytical challenges. This module provides the key analytic skills needed to analyse and interpret electronic health record data. This module will tackle key analytic concepts and techniques required to analyse EHR data. This will include: creating data sets for analysis from EHRs using various study designs; codelist creation and use; measurement in EHR (exposures, outcomes, covariates) including drug exposure data; a review of the causal inference framework and important causal analytic methods, including regression and propensity score analysis; developing prediction models; methods for handling longitudinal and other hierarchical data; and survival analysis. The module includes an in-depth exploration of the last topic, survival analysis. These methods are widely used in health data science and particularly in the analysis of EHR data. The methods covered also apply to other data sources, including longitudinal cohort data, patient disease registries, and randomized trials.
Duration	5 weeks at 2.5 days per week
Timetabling slot	Slot D2
Last Revised (e.g. year changes approved)	September 2021

Programme(s) This module is linked to the following programme(s) *(Lead programme first)*	**Status** *(Compulsory/Recommended Option)*
MSc Health Data Science	**Recommended**

Module Aim and Intended Learning Outcomes

Overall aim of the module
The overall module aim is to: • Equip students with the necessary skills to analyse data arising from Electronic Health Records

Module Intended Learning Outcomes
Upon successful completion of the module a student will be able to: 1. Demonstrate knowledge and understanding of the unique features of Electronic Health Record (EHR) data and why these features lead to analytic complexities 2. Critically apply a range of analytic methods to EHR data and demonstrate an understanding of their purposes 3. Demonstrate an understanding of the assumptions underlying a range of relevant methods to handle confounding, and critically apply the methods to EHR using R 4. Analyse and interpret EHR data using the most appropriate method of analysis and evaluate its effectiveness.

Indicative Syllabus

Session Content
The module is expected to cover the following topics: • Measurement in Electronic Health Record (EHR) data, including use of codelists and obtaining drug exposure data • Causal analysis, review of basic concepts • Propensity score analysis, regression, and related approaches • Handling longitudinal and other hierarchical data • Functions used in the description and analysis of survival data, including hazard and survivor functions • Non-parametric and parametric estimation of survival curves • The theory and use of proportional hazard models, including parametric models and the Cox model

Teaching and Learning

Notional Learning Hours

Type of Learning Time	Number of Hours	Expressed as Percentage (%)
Contact time	**50**	**33**
Directed self-study	**30**	**20**
Self-directed learning	**20**	**14**
Assessment, review and revision	**50**	**33**
Total	**150**	**100**

Teaching and Learning Strategy
The teaching and learning strategy is structured around a combination of lectures, delivered either in person or via a series of short pre-recorded videos. Each section of lecture material will be followed by computer or non-computer practical sessions. These practical sessions ensure that students have the opportunity to apply the concepts and methods covered by lecture content to real data scenarios. Electronic Health Records are protected by stringent information governance rules and so real patients' records will not be shared with students, but datasets used in this course will be designed to closely mimic real Electronic Health Record data drawn from real analyses undertaken by staff in the faculty. Students are provided with detailed solutions to the tasks set in practical sessions, enabling them to check their understanding of the material. The assessment task, which comes towards the end of the module, is the point at which students demonstrate a consolidation of their learning across the whole module.
Indicative Breakdown of Contact Time:

Type of delivery	**Total (hours)**
Lecture	*20*
Seminar	*3*
Tutorial	
Computer Practical	*27*
Laboratory Practical	
Fieldwork	
Project Supervision	
Total	**50**

Assessment

Assessment Strategy
Formative multiple-choice questions will be used in the earlier part of the module to assess understanding of key concepts. A second type of formative assessment will be undertaken in the practical sessions, in which students will write a methods and results section about an analysis they have undertaken, with peer marking of the report. The summative assessment will consist of an analysis of electronic health record data to answer a given health data science question. Students will select appropriate techniques from the module to address the question, apply them to the data, and write a short report on the results and interpretation. Resit/deferred/new attempts: the task will be similar to the original assessment but the dataset and question will be different.

Summative Assessment

Assessment Type *(delete as appropriate)*	**Assessment Length (i.e. Word Count, Length of presentation in minutes)**	**Weighting** (%)	**Intended Module Learning Outcomes Tested**
Coursework	**4-5 pages**	**100**	**1-5**

Resitting assessment
Resits will accord with the LSHTM's Resits Policy For individual students resitting a group assessment there will be an approved alternative assessment as detailed below.

Assessment being replaced	**Approved Alternative Assessment Type**	**Approved Alternative Assessment Length** (i.e. Word Count, Length of presentation in minutes)
NA	NA	NA

Resources

Indicative reading list *(if applicable)*

Lash TL, Vanderweele TJ, Haneuse S, Rothman KJ. Modern Epidemiology. Fourth edition. Wolters Kluwer. 2021

Collett D (2003): "Modelling Survival Data in Medical Research"
Cox DR and Oakes D (1984): "Analysis of survival data"
Marubini and Valsecchi (1995): "Analysing Survival Data from Clinical Trials and Observational Studies", Machin D., Cheung Y.B. and Parmar M.K.B: "Survival Analysis. A practical approach (2006).
Aalen, Borgan, Gjessing. (2008) "Survival and Event History Analysis". Springer

Other resources

We do not recommend any specific textbook relating to analysis of EHR data, but there are many useful publications. Here are a few that are relevant to the course materials:

Okoli, G.N., Myles, P., Murray-Thomas, T. et al. Use of Primary Care Data in Research and Pharmacovigilance: Eight Scenarios Where Prescription Data are Absent. Drug Saf (2021). https://doi.org/10.1007/s40264-021-01093-9

Pye, SR, Sheppard, T, Joseph, RM, et al. Assumptions made when preparing drug exposure data for analysis have an impact on results: An unreported step in pharmacoepidemiology studies. Pharmacoepidemiol Drug Saf. 2018; 27: 781– 788. https://doi.org/10.1002/pds.4440

Watson J, Nicholson BD, Hamilton W, et al. Identifying clinical features in primary care electronic health record studies: methods for codelist development BMJ Open 2017;7:e019637. doi: 10.1136/bmjopen-2017-019637

Teaching for Disabilities and Learning Differences

- Lectures will be recorded in line with the LSHTM's policy on Lecture Recording.
- The module manual will be made available in advance of the start of the module and will be produced in accessible format.
- Slides will be made available in advance of each lecture or seminar and produced in accessible format.
- All material will be made available through Moodle.

The module-specific site on Moodle provides students with access to lecture notes and copies of the slides used during the lecture prior to the lecture (in pdf format). All lectures are recorded and made available on Moodle as quickly as possible. All materials posted up on Moodle areas, including computer-based sessions, have been made accessible where possible.

The LSHTM Moodle has been made accessible to the widest possible audience, using a VLE that allows for up to 300% zoom, permits navigation via keyboard and use of speech recognition software, and that allows listening through a screen reader. All students have access to "SensusAccess" software which allows conversion of files into alternative formats.

For students who require learning or assessment adjustments and support this can be arranged through the Student Support Services – details and how to request support can be found on the LSHTM Disability Support pages.