MODULE SPECIFICATION

Academic Year (student cohort covered by specification)	2020-21
Module Code	2450
Module Title	Advanced Statistical Modelling
Module Organiser(s)	Edmund Njeru Njagi and Nick Jewell
Faculty	Epidemiology & Population Health
FHEQ Level	Level 7
Credit Value	**CATS:** 15 **ECTS:** 7.5
HECoS Code	101031 : 101034
Term of Delivery	Term 3
Mode of Delivery	For 2020-21 this module is delivered online. Teaching will comprise a combination of live and interactive activities (synchronous learning) as well as recorded or self-directed study (asynchronous learning). We do not yet know whether or not there will be any on-campus activities during Term 3. This decision will be made in February.
Mode of Study	Full-time
Language of Study	English
Pre-Requisites	This module is intended for students with an understanding of probability (Module 2038), linear regression models, generalized linear models and linear regression models for hierarchical data to the level provided by the modules Generalized Linear Models (2462) and Analysis of Hierarchical & Other Dependent Data (2465). Students require an understanding of matrix algebra and calculus (to first year undergraduate mathematics level). The module Analysis of Hierarchical & Other Dependent Data (2465) is strongly recommended for prior study.
Accreditation by Professional Statutory and Regulatory Body	None
Module Cap (Maximum number of students)	25 (numbers may be capped due to limitations in facilities or staffing)
Target Audience	This module is intended for students with an understanding of linear regression models for hierarchical data and generalized linear models to the level provided by the modules Generalized

	Linear Models (2462) and Analysis of Hierarchical & Other Dependent Data (2465).
Module Description	A two-part module, on methods for causal inference (Part I), and population-averaged and subject-specific models for discrete dependent data (Part II).
Duration	5 weeks at 2 days per week
Timetabling slot	Slot E.
Last Revised (e.g. year changes approved)	October/2020

Programme(s) This module is linked to the following programme(s)	**Status**
MSc Medical Statistics	Recommended

Module Aim and Intended Learning Outcomes

Overall aim of the module
The overall module aim is to: introduce recent developments in statistical modelling, focussing in the first part of the module on statistical models for drawing causal inferences, and in the second part of the module on models for discrete dependent data.

Module Intended Learning Outcomes
Upon successful completion of the module a student will be able to: - Understand the nature of the scientific enquiry / practical problem that led to the developments and for which they are appropriate.   - Understand the concepts on which they are based and their relationship to each other and to the approaches/techniques encountered earlier in the MSc Medical Statistics programme.   - Develop the skills to use the methods creatively and independently in practical problems.   - Develop the capacity to review critically their own and others' statistical modelling work in these settings.

Indicative Syllabus

Session Content
Part I: - Causal languages in statistics including counterfactual reasoning   - Graphical models for statistics and causality   - Causal inference using parametric statistical models: traditional regression

Session Content
- Causal inference using semiparametric statistical models: propensity scores, inverse probability weighting and doubly robust estimation   - Causal mediation analysis   **Part II:**   - Revision of linear and generalized linear models   - Modelling dependent non-normal data: subject specific and marginal models   - Generalised estimating equations and empirical (sandwich) variance estimates   - Likelihood based analyses for subject-specific models

Teaching and Learning

Notional Learning Hours

Type of Learning Time	Number of Hours	Expressed as Percentage (%)
Contact time	42.5	28
Directed self-study	0	0
Self-directed learning	57.5	39
Assessment, review and revision	50	33
Total	**150**	**100**

Student contact time refers to the tutor-mediated time allocated to teaching, provision of guidance and feedback to students. This time includes activities that take place in face-to-face contexts such as lectures, seminars, demonstrations, tutorials, supervised laboratory workshops, practical classes, project supervision as well as where tutors are available for one-to-one discussions and interaction by email. Student contact time also includes tutor-mediated activities that take place in online environments, which may be synchronous (using real-time digital tools such as Zoom or Blackboard Collaborate Ultra) or asynchronous (using digital tools such as tutor-moderated discussion forums or blogs often delivered through the School's virtual learning environment, Moodle).

The division of notional learning hours listed above is indicative and is designed to inform students as to the relative split between interactive (online or on-campus) and self-directed study.

Teaching and Learning Strategy
A combination of lectures and computer-based practical sessions. Practical sessions after the lectures will give students the opportunity to apply, on relevant data sets, the statistical methods learnt during lectures. Solutions to the practical tasks will be explained during the various sessions and through solutions documents.

Assessment

Assessment Strategy
The assessment for this module has been designed to measure student learning against the module intended learning outcomes (ILOs) as listed above. Formative assessment methods may be used to measure students' progress. The grade for summative assessment(s) only will go towards the overall award GPA. Part 1: The assessment will be a project assignment (data analysis). It will involve the application, interpretation and comparison of different statistical approaches to addressing particular causal questions using a dataset from an observational study. Part 2: The assessment will be a project assignment (data analysis). It will involve the use, interpretation and comparison, of several alternative statistical analyses for a given longitudinal data set from a medical setting.

Summative Assessment

Assessment Type	Assessment Length (i.e. Word Count, Length of presentation in minutes)	Weighting (%)	Intended Module Learning Outcomes Tested
Coursework (Part 1)	Report, maximum of three pages (11pt font), including figures and tables.	50	All
Coursework (Part 2)	Report, maximum of three pages (11pt font), including figures and tables.	50	All

Resitting assessment
Resits will accord with the LSHTM's [Resits Policy.](#) Resit/deferred/new attempts - The tasks will be similar to the original assessments, but involving a different data set. The next assessment deadline will be during mid/late September of the current academic year.

Resources

Indicative reading list

Part I:

1. Course handbook will be distributed that covers most of the material for Part I

Part II:

1. Agresti, A. (2002). *Categorical Data Analysis*. 2nd ed. New Jersey: Wiley.
2. Fitzmaurice, G.M., Laird, N.M., and Ware, J.H. (2011). *Applied Longitudinal Analysis*. 2nd Ed. New York: Wiley.
3. Hosmer, D.W., Stanley, L., and Hosmer, D.W. (2000). *Applied logistic regression*. New York: Wiley.
4. Molenberghs, G., and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer.
5. Rabe-Hesketh, S., and Anders. Skrondal. (2012). *Multilevel and Longitudinal Modeling Using Stata*. Volume 2: *Categorical responses, counts, and survival*. 3rd ed. College Station, Tex: Stata Press Publication.

Other resources

Module Information can be found on the Virtual Learning Environment (Moodle) containing information about each session and key references for the module. Copies of all the slides used, pre-recorded videos, practicals and their solutions; datasets required for the practicals and assessment exercise will be available on Moodle.

Teaching for Disabilities and Learning Differences

The module-specific site on Moodle gives students access to lecture notes and copies of the slides used during the lecture. Where appropriate, lectures are recorded and made available on Moodle. All materials posted on Moodle, including computer-based sessions, have been made accessible where possible.

LSHTM Moodle is accessible to the widest possible audience, regardless of specific needs or disabilities. More detail can be found in the Moodle Accessibility Statement which can also be found within the footer of the Moodle pages. All students have access to "SensusAccess" software which allows conversion of files into alternative formats.

Student Support Services can arrange learning or assessment adjustments for students where needed. Details and how to request support can be found on the LSHTM Disability Support pages.