MODULE SPECIFICATION

This module is a compulsory module of the new MSc Health Data Science. The module specification has been provisionally agreed as part of the validation process for the MSc. Module Organisers are currently developing the details of the teaching to ensure the best possible learning experience and therefore some changes may still be made. We anticipate that the final module specifications will be published by the end of the summer.

1. Overview

Academic Year (student cohort covered by specification)	2020-21			
Module Code	2489			
Module Title	Statistics for Health Data Science			
Module Organiser(s)	Nick Jewell and Kathleen O'Reilly			
Faculty	Epidemiology & Population Health			
FHEQ Level	Level 7			
Credit Value	CATS	15	ECTS	7.5
HECoS Code	101031			
Term of Delivery	Term 1			
Mode of Delivery	Face-to-face			
Mode of Study	Full-time			
Language of Study	English			
Pre-Requisites	None, over and above the pre-requisites for the programme MSc Health Data Science			
Accreditation by Professional Statutory and Regulatory Body	None			
Module Cap (Maximum number of students)	In first year of delivery: max 20 students.			
Target Audience	This is a compulsory module for the programme MSc Health Data Science			

Module Description	This module provides an introduction to the key statistical concepts and methods for health data science. Topics covered include probability, initial data description and exploration, statistical inference, regression, and Bayesian analysis. These topics provide the framework needed for subsequent modules. The module places a focus on learning through practical examples and incorporates directed learning, lectures, group discussion, and computer practical exercises.
Duration	15 x 0.5 day sessions
Timetabling slot	Term 1
Last Revised (e.g. year changes approved)	December 2019

2. Programme(s) that this module is part of

Programme This module is linked to the following programme(s)	**Status**
MSc Health Data Science	Compulsory

3. Module Aim and Intended Learning Outcomes

Overall aim of the module
The overall module aim is to: • introduce the motivation and critical thinking towards solving a question in health science through interrogation of data and drawing conclusions from evidence; • introduce the principles of probability, regression modelling and statistical inference within frequentist and Bayesian settings.

Module Intended Learning Outcomes
Upon successful completion of the module a student will be able to: 1. evaluate the application of different probability distributions to model health data (including Poisson, Binomial and Normal);   2. critically analyse frameworks for frequentist and Bayesian inference and evaluate their strengths, limitations and differences;   3. examine the concepts of sampling variability, estimators, bias, confidence intervals and credible intervals;   4. examine the theoretical basis of linear regression and generalized linear models;   5. assess the application of regression modelling to address specific health data science questions;   6. critically evaluate strengths and limitations of different statistical methods, including regression models, within a health data science project;   7. draw conclusions from the results of a data analysis and justify those conclusions, appropriately acknowledging uncertainty in the results.

4. Indicative Syllabus

Session Content
The module is expected to cover the following topics: • Exploratory data analysis and the "problem solving" cycle   • Conditional probability, Bayes theorem, binary/discrete distributions   • Distributions for continuous variables   • Sampling distributions and the central limit theorem   • Likelihoods and maximum likelihood estimation   • Frequentist inference   • Bayesian inference   • Linear regression modelling   • Generalized linear models

5. Teaching and Learning

Notional Learning Hours		
Type of Learning Time	**Number of Hours**	**Expressed as Percentage (%)**
Contact time	37.5	25
Directed self-study	52.5	35
Self-directed learning	34.5	23
Assessment, review and revision	25.5	17
Total	150	100

Teaching and Learning Strategy
Each session will consist of an interactive session between students and lecturer, either as a lecture or seminar. Students will be expected to read directed material ahead of the lectures/seminars and the learning outcomes will be consolidated during this time. This module will use a mixture of teaching techniques, including traditional lecture and flipped classroom. Practical sessions will be provided that consist of problem-based learning exercises to further consolidate the learning outcomes into useable skills when facing a health data question. The practicals will include computer exercises and group discussion.
Indicative Breakdown of Contact Time:

Type of delivery	Total (hours)	
Lecture	7.5	
Seminar	7.5	
Tutorial	*7.5*	
Computer Practical	15	
Laboratory Practical	0	
Fieldwork	0	
Project Supervision	0	
Total	**37.5**	

6. Assessment

Assessment Strategy
Assessment will consist of a formative assignment during the module, and an assessed exam (open book) at the end of the module.   The formative assessment will feature multiple choice questions incorporated into the practicals, and will be self-assessed.   The main summative assessment will incorporate a mixture of multiple choice questions and traditional exam questions. The exam will be of length 2 hours and questions will be mapped to the intended learning outcomes through questions specific to each outcome.

Summative assessment			
Assessment Type	**Assessment Length (i.e. Word Count, Length of presentation in minutes)**	**Weighting (%)**	**Intended Module Learning Outcomes Tested**
Timed Test (in-module test e.g. MCQ)	2 hours	100	1-7

Resitting assessment		
Resits will accord with the LSHTM's [Resits Policy in Chapter 8a, PGT Regulations, of the Academic Manual](#) 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)
N/A – no group assessment		

7. Resources

Indicative reading list

Huebner et al. A Contemporary Conceptual Framework for Initial Data Analysis. Observational Studies 4 (2018) 171-192. https://obsstudies.org/contemporary-conceptual-framework-initial-data-analysis/

Essential Medical Statistics, Kirkwood B. R. and Sterne J. A. C. Wiley-Blackwell, 2nd Edition,
2003.

A First Course in Probability. Ross S. Pearson, 8th Edition, 2008.

An Introduction to Medical Statistics, Bland J. M. OUP, 3rd Edition, 2000.

Harrell, F.E. Regression Modeling Strategies. Springer. [Extensive coverage of practical
strategies for modelling data].

Dobson, A.J and Barnett, A.G. (2008) An Introduction to Generalized Linear Models, Third Edition. Chapman & Hall.

David Lunn et al "The BUGS Book: A Practical Introduction to Bayesian Analysis" Chapman & Hall (2013)

Emmanuel Lesaffre and Andrew B. Lawson "Bayesian Biostatistics" Wiley (2012)

David Spiegelhalter et al "Bayesian approaches to clinical trials and health-care evaluations" Wiley (2004)

Other resources
Module information, including timetables, lecture notes, practical instructions and key literature for each session will be made available via the Virtual Learning Environment (Moodle).

8. Teaching for Disabilities and Learning Differences

- Lectures will be recorded using Panopto 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.