



MODULE SPECIFICATION

Academic Year (student cohort covered by specification)	2026-27
Module Code	2450
Module Title	Causal Inference and Missing Data
Module Organiser(s)	Clémence Leyrat, Jonathan Bartlett
Faculty	Epidemiology & Population Health
FHEQ Level	Level 7
Credit Value	CATS: 15 ECTS: 7.5
Term of Delivery	Term 3
Mode of Delivery	For 2026-27 this module will be delivered face-to-face. The recordings of lectures will be available on Moodle after lectures take place.
Mode of Study	Full-time
Language of Study	English
Pre-Requisites	This module is intended for students with an understanding of probability (Module 2038: Foundations of Medical Statistics) and regression models to the level provided by the module 'Statistical models for discrete outcomes' (2462). Students should be competent in using R for statistical analysis.
Accreditation by Professional Statutory and Regulatory Body	None
Module Cap (indicative number of students)	35 (numbers may be capped due to limitations in facilities or staffing)
Target Audience	This module is intended for statisticians and epidemiologists who want to learn about methods for causal inference and missing data.
Module Description	This first part of this module covers the concepts and methods of causal inference from observational data based on the concept of potential outcomes. It will cover causal language, assumptions sufficient to estimate causal effects, graphical representation of causal assumptions, and statistical methods for estimation of causal effects. The second part will explore the impacts of missing data on the validity and efficiency of statistical inferences. It will introduce



	the methods of multiple imputation and inverse probability weighting for handling missing data and their practical implementation.
Duration	5 weeks at 2 days per week
Timetabling slot	Slot E
Last Revised (e.g. year changes approved)	June 2023

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

Module Aim and Intended Learning Outcomes

Overall aim of the module
The overall module aim is to: introduce students to concepts and statistical methods for i) drawing causal inferences from observational data; ii) handling missing data.

Module Intended Learning Outcomes
Upon successful completion of the module a student will be able to: - Understand the concept of potential outcomes and the assumptions under which causal effects can be estimated from observational data. - Apply suitable statistical methods to estimate causal effects from observational data and be able to critically evaluate their suitability in applied settings. - Understand the impacts of missing data on statistical analyses and assumptions under which valid inferences can be obtained when some data are missing. - Apply principled statistical methods for handling missing data and be able to critically evaluate their suitability in applied settings.

Indicative Syllabus

Session Content
Causal inference: - Causal languages in statistics including counterfactual reasoning - Graphical models for statistics and causality

Session Content

- Emulated target trials
- Causal inference using parametric statistical models: traditional regression
- Causal inference using semiparametric statistical models: propensity scores, inverse probability weighting and doubly robust estimation
- Causal mediation analysis
- Causal inference for time-dependent confounding

Missing data:

- Impacts of missing data and missingness mechanisms
- Complete case analysis and ad-hoc methods for missing data
- Multiple imputation for single and multiple variables
- Handling missing data in longitudinal analyses

Teaching and Learning

Notional Learning Hours

Type of Learning Time	Number of Hours	Expressed as Percentage (%)
Contact time	40	27
Directed self-study	0	0
Self-directed learning	60	40
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.

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

Teaching and Learning Strategy

A combination of lectures and computer-based practical sessions, with content informed by current research developments. 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

Teaching and Learning Strategy

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. The grade for summative assessment(s) only will go towards the overall award GPA.

The summative assessment will consist of a statistical data analysis to address a causal question using data which has missing values. Students will write a structured report based on their analyses.

Summative Assessment

Assessment Type	Assessment Length (i.e. Word Count, Length of presentation in minutes)	Weighting (%)	Intended Module Learning Outcomes Tested
Coursework	Structured report, maximum of five pages, including figures and tables.	100	All

Resitting assessment

Resits will accord with [Chapter 8a](#) of the LSHTM Academic Manual.

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

1. Hernan MA, Robins JM. *Causal Inference: What If (Monographs on Statistics and Applied Probability)*. CRC Press. <https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/>
2. Carpenter JR, Bartlett JW, Morris T, Wood A, Quartagno M, Kenward MG. (2023) *Multiple imputation and its application*. 2nd ed. Chichester, West Sussex: Wiley.
3. Little RJA, Rubin DB. (2002) *Statistical analysis with missing data*. 2nd ed. Hoboken, New Jersey: Wiley
4. Van Buuren S. *Flexible imputation of missing data*. 2nd ed. CRC Press. <https://stefvanbuuren.name/fimd>

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, links to lecture recordings, practicals and their solutions, datasets required for the practicals and assessment 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](#).