



## MODULE SPECIFICATION

<b>Academic Year (student cohort covered by specification)</b>	2025-26
<b>Module Code</b>	2487
<b>Module Title</b>	Health Data Management
<b>Module Organiser(s)</b>	Orlagh Carroll and Kevin Van Zandvoort
<b>Faculty</b>	Epidemiology & Population Health
<b>FHEQ Level</b>	Level 7
<b>Credit Value</b>	<b>CATS: 15</b> <b>CATS: 7.5</b>
<b>HECoS Code</b>	100755
<b>Term of Delivery</b>	Term 1
<b>Mode of Delivery</b>	For 2025-26 this module will be delivered by predominantly face-to-face teaching modes.  Where specific teaching methods (lectures, seminars, discussion groups) are noted in this module specification these will be delivered by predominantly face-to-face sessions. There will be a combination of live and interactive activities (synchronous learning) as well as recorded or self-directed study (asynchronous learning).
<b>Mode of Study</b>	Full-time
<b>Language of Study</b>	English
<b>Pre-Requisites</b>	None, over and above the prerequisites for the programme MSc Health Data Science
<b>Accreditation by Professional Statutory and Regulatory Body</b>	None
<b>Module Cap (indicative number of students)</b>	30 students
<b>Target Audience</b>	This module is compulsory for the MSc in Health Data Science.
<b>Module Description</b>	This module is an introduction to the theory and practise of health data management for use in data science. The module will use a project life cycle workflow to introduce each successive step. Students will develop skills in a variety of

	<p>software and languages that support each stage in the process. The components included will be:</p> <ul style="list-style-type: none"> <li>• Data standards, and Relational Data, its uses and storage <ul style="list-style-type: none"> <li>• Data Collection – demonstrated using Open Data Kit (ODK) tools</li> <li>• Databases, query languages, and particular implementing technologies - demonstrated using SQLite and PostgreSQL</li> <li>• Data cleaning as exemplified using R/Python</li> <li>• Accessing data in a relational database directly from R/Python</li> <li>• The use of a relational database platform (MIMIC)</li> <li>• Data pipelines, automated report generation and the presentation of data (data visualisation)</li> </ul> </li> </ul>
<b>Duration</b>	8 days, each day 3 x 90 minute blocks; most days 1 seminar, 1 group practical session, 1 individual practical session
<b>Timetabling slot</b>	Term 1
<b>Last Revised (e.g. year changes approved)</b>	2024

<b>Programme(s)</b>	<b>Status</b>
This module is linked to the following programme(s)	
MSc Health Data Science	Compulsory

## Module Aim and Intended Learning Outcomes

<b>Overall aim of the module</b>
<p>The overall module aim is to:</p> <ul style="list-style-type: none"> <li>• Enable students to develop a robust data management infrastructure to support health informatics research projects. Successful students will develop core technical skills and conceptual frameworks, and demonstrate these in integrated assessments of typical data management tasks.</li> </ul>



### Module Intended Learning Outcomes

Upon successful completion of the module a student will be able to:

1. identify key attributes of data collection systems for subsequent practical use of yielded data.
2. conceptualise complex data in relational terms.
3. apply standard query language to extract data views from large scale databases
4. fuse data from multiple sources into an integrated view
5. implement core principles of effective data and process documentation

### Indicative Syllabus

#### Session Content

The module is expected to cover the following topics:

- data capture using ODK
- data standard formats, and relational data schema
- relational database usage, particularly focusing on SQL databases
- data visualisation and data analysis pipelines;
- data and process documentation

### Teaching and Learning

#### Notional Learning Hours

Type of Learning Time	Number of Hours	Expressed as Percentage (%)
Contact time	36	24
Directed self-study	37	25
Self-directed learning	37	25
Assessment, review and revision	40	26
<b>Total</b>	<b>150</b>	<b>100</b>

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

The learning activities will roughly follow a seminar, group work, then individual work model. The overall schedule of the module will follow the lifecycle of data, focusing the “management” portion of the cycle, from collection to dissemination. Students will work with, and be evaluated on, real world data and data management tasks – i.e. they will have to deal with messy data and ambiguous analytical demands.

## 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 are used to measure students’ progress. The grade for summative assessment only will go towards the overall award GPA.

The assessment for this module in term 1 will be online.

The formative assessments will involve exercises built into the course materials to test and reinforce knowledge.

The summative assessment will entail extracting data from multiple sources, combining that data, presenting an analysis of that data, document the process, and critique data source quality.

### Summative Assessment

Assessment Type	Assessment Length (i.e. Word Count, Length of presentation in minutes)	Weighting (%)	Intended Module Learning Outcomes Tested
Coursework	Data analysis workflow (order 1000 lines of code), compiled report from that workflow (max 3 pages), critical feedback actions (max 3 pages)	100	all

### Resitting assessment

Resits will accord with [Chapter 8a](#) of the LSHTM Academic Manual. The resit assessment will be the same format as the main assessment, but with different required statistics to be reported.



## Resources

### Indicative reading list

SQL in 10 minutes a day, Sams Teach Yourself, Ben Forta, Fifth edition.

The Book of OHDSI. Observational Health Data Sciences and Informatics. [Free online]  
<https://ohdsi.github.io/TheBookOfOhdsi/>

### 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).

## Teaching for Disabilities and Learning Differences

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](#).

A set of lecture notes will be produced to accompany the core material, adhering to LSHTM accessibility guidelines.

Any screencasts will be transcribed and the transcription provided to accompany the core material.

Panopto will be used to capture lectures for later review by students in line with the LSHTM's policy on Lecture Recording.