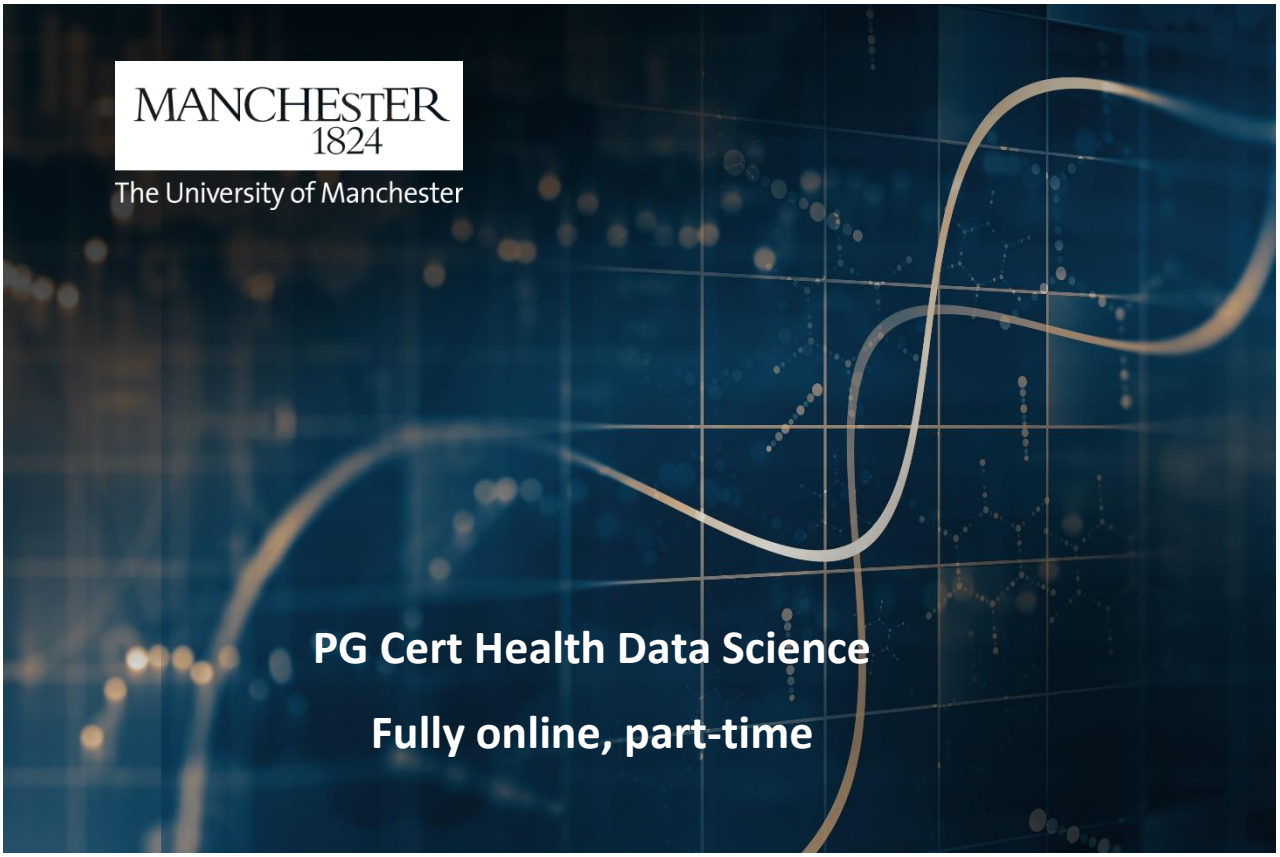

Indicative Course Content – September 2021

Contents

Unit start dates – September 2021 intake

- Unit 1: Monday, 6 September 2021
- Unit 2: Monday, 22 November 2021
- Unit 3: Monday, 28 February 2022
- Unit 4: Monday, 30 May 2022

Teaching and learning overview

This course is delivered entirely via distance learning, including assessment. Each unit runs over 10 weeks, with a nominal 15 hours per week of student work. The times specified will vary each week. For example, there might be no private study in the first week before assessments have been set. There might be a larger private study requirement later on in the unit when students will need to prepare for their assessments.

A typical week consists of:

- An overview of the material, presenting the learning objectives for the week.
- Explanatory material (approximately 3 hours of student activity per week) in the form of video lectures, papers, articles, the course text, and links to further resources.
- Exercises (approximately 4 hours per week). These are formative, with feedback for them given in the tutorial.
- Discussion (Approximately 2 hours per week). Students are encouraged to discuss the exercises and material in the forums where tutors will facilitate peer learning, providing feedback or input where necessary.
- Formative Questionnaire. This is to gather students' questions and highlight misconceptions ready for the tutorial.
- Tutorial (1 hour per week). Student will video conference with their tutor in groups of 6-8 to discuss and give and receive feedback.

Private study (approx. 5 hours per week):

- Revision
- Coursework
- Further practice (after the tutorials)
- Independent/further study

Unit 1 – Mathematics & Statistics for Health Data Science

Unit aims

The unit aims to introduce fundamental statistical principles used in health data science. The focus is on gaining a holistic understanding of statistical principles on a conceptual (rather than theoretical) level, and to be exposed to exemplar techniques/methods that are widely used when analysing health data. There will be a strong focus on embedding and motivating the techniques from an applied perspective. On completion the student should be aware of a range of techniques and have experience of writing statistical scripts and executing them in R/Python.

Unit description

The unit will begin with some required mathematical and probability theory, then will focus on the three key 'data science tasks' of description, prediction, and causal inference. Throughout, there will be a focus on articulating appropriate scientific/research questions, coding in R, and communicating results effectively (including through visualisations). Watch the Unit 1 Spotlight video.

Indicative content:

- Mathematics and Probability Theory: Rules of probability, conditional probability, discrete probability distributions, continuous probability distributions, statistical inference.
- Description: The three 'data science' tasks, summary statistics, visualisations for data exploration, confidence intervals, hypothesis testing.
- Prediction and Causal Inference: Linear and logistic regression, assessing goodness of fit, confounding, regression for causal estimation, understanding the difference between prediction and causal inference.

Learning outcomes

Category	Learning outcome
Knowledge and understanding	▪ Demonstrate a critical understanding of the steps required to address a scientific question using health data. ▪ Discuss and justify the importance of probability as a measure of uncertainty. ▪ Explain the key discrete and continuous probability distributions. ▪ Demonstrate a critical understanding of the principles of statistical inference. ▪ Explain key ideas and concepts in statistics. ▪ Explain the distinctions between description, prediction, and causal inference. ▪ Explain the principles and goals of regression analysis for prediction, and for causal inference. ▪ Describe the concept of confounding and its relevance in causal inference. ▪ Discuss how to examine adequacy of fitted models, including goodness of fit and plausibility of required assumptions.
Intellectual skills	▪ Formulate a clear statement of a research question or hypothesis. ▪ Identify statistical methods to be applied to health data to address a stated question or hypothesis. ▪ Evaluate scientific and clinical literature. ▪ Critically appraise methods for analysing health data.
Practical skills	▪ Apply statistical methods and knowledge, including of regression modelling, to health data to address a stated research question or hypothesis. ▪ Assess uncertainty present in a problem, and reflect this in the recommendations and/or decision making. ▪ Execute and write statistical scripts to compute statistical measures and analyse data. ▪ Prepare a data set for analysis.

Transferable skills and personal qualities	▪ Present complex ideas in simple terms. ▪ Demonstrate statistical programming skills. ▪ Demonstrate experience of solving applied problems. ▪ Work in partnership with colleagues to create reports suitable for presentation to stakeholders.

Assessment

Task	Weighting
Contribution to discussion boards	20%
Individual presentation describing a key concept in statistics or probability, and an example of its application.	40%
Short written assignment demonstrating a critical analysis of health data to address a research question.	40%

Suggested reading materials

Introductory Reading (less technical)

- Spiegelhalter, D. (2019). The Art of Statistics: Learning from Data. Penguin UK.
- Pearl, J., & Mackenzie, D. (2018). The book of why: the new science of cause and effect. Basic Books.

Detailed Reading (specific chapters indicated)

- Grinstead, C. M., & Snell, J. L. (2012). Introduction to probability. American Mathematical Soc.. (Chapters 1-9).
- Kirkwood, B. R., & Sterne, J. A. (2010). Essential medical statistics. John Wiley & Sons. (Chapters 1-21).
- Steyerberg, E. W. (2019). Clinical prediction models (2nd ed.). New York: Springer. (Chapters 1-3, 4.1-4.2, 11, 15, 17).
- Hernán MA, Robins JM (2020). Causal Inference: What If. Boca Raton: Chapman & Hall/CRC (Chapters 1-10).

Unit 2 – Data Engineering for Health Data Science

Unit description and aims:
The unit aims to equip the student with fundamental techniques in data engineering, with a particular focus on programming in a transparent and reproducible fashion (using Python), use of version control, and interacting with relational databases. The unit covers three key areas: fundamentals of computer programming in Python, use of version control, and relational databases using SQL.

Indicative content:
- Programming in Python: Variables and data types, control structures, data manipulation, visualisations, programming best practice.
- Version control in Git: Creating a repository, git commands, collaborating,
- Relational databases and SQL: Extracting and summarising data, understanding joins.

Learning outcomes

Category	Learning outcome
Knowledge and understanding	▪ Demonstrate a critical understanding of the basic principles of the Python programming language. ▪ Explain the features of Python that support object-oriented programming. ▪ Create functioning and well documented Python code. ▪ Explain the need for repositories to manage code. ▪ Construct a Git repository to manage Python code. ▪ Demonstrate a critical understanding of relational databases, their advantages and disadvantages. ▪ Explain the function of the main SQL commands. ▪ Explain the conceptual distinction between different database joins.
Intellectual skills	▪ Demonstrate a capacity to think about programming problems in an abstract way and make critical judgement regarding programming decisions. ▪ Demonstrate the ability to solve programming problems in an efficient and effective manner. ▪ Planning, conducting and report on a programming problem.
Practical skills	▪ Apply programming skills to address a stated objective. ▪ Extract a data set ready for analysis. ▪ Write clear programming scripts that are reproducible.
Transferable skills and personal qualities	▪ Present complex ideas in simple terms. ▪ Demonstrate problem solving and logical skills. ▪ Work in partnership with colleagues to create reports suitable for presentation to stakeholders.

Assessment

Assessment task	Weighting
Contribution to discussion boards	20%
Individual presentation describing a key concept in data engineering, and an example of its application.	40%
Short written assignment demonstrating critical use of data engineering techniques to address a scientific objective.	40%

Suggested reading materials
https://software-carpentry.org/reading

Unit 3 – Systems and Technologies in Health Data Science

Unit aims

This unit aims to examine health systems and technologies and evaluate their impact on the delivery of healthcare, and to understand different approaches and methodologies to systems development, and how and when to apply them, in an international context.

Unit description

This unit will provide an opportunity for students to develop an understanding of health systems and technologies, and their implementation. In order to understand a health system or technology a health data scientists needs to be aware of the core issues and methodologies in developing, deploying and managing a healthcare system, and the impact to healthcare delivery. It is also paramount to understand how the data is created, collected, stored and retrieved so as to be able to use it. In particular, the unit will cover the framework for handling patient data in a confidential and secure manner to ethical and quality standards that are appropriate for a modern health service.

Indicative content:

- Health Information Systems and Technologies – social and mobile health technologies; architectures; networks; internet; cloud;
- Data storage and retrieval technologies – databases and Standard Querying Language (SQL)
- Information governance and security
- Data models and architecture
- System testing and quality assurance: ISO-standards
- System design cycle and methodologies – agile; waterfall
- Description of main requirements analysis methodologies
- Evaluation of informatics systems design/development
- Critical appraisal and understanding failure and success factors of historical and current projects

Learning objectives

Category	Objective
Knowledge and understanding	▪ Demonstrate a critical understanding of the historical context of health information/IT systems ▪ Explain the different approaches to systems development and evaluate how they fit in the broader cycle of development, deployment and maintenance of healthcare systems. ▪ Discuss issues surrounding system quality assurance ▪ Review future thinking around healthcare systems ▪ Discuss the range of technologies available to produce and store data ▪ Understand the role of the patient in the health system development ▪ Discuss and understand concepts in data modelling
Intellectual skills	▪ Determine the impact of introducing a new system/technology in the healthcare system ▪ Develop and communicate a strategy to introduce system/technology to address a healthcare issue ▪ Critically appraise a system/technology or data modelling plan and make suggestions for improvement ▪ Apply systems methodologies to a healthcare problem ▪ Design approaches and queries to retrieve data ▪ Apply information governance and security guidelines to information collection and use in healthcare systems
Practical skills	▪ Perform and communicate the results of a systems requirements gathering exercise

	▪ Perform systems methodologies to design a technological solution to a healthcare problem
Transferable skills and personal qualities	▪ Work collaboratively within a team ▪ Present ideas and work in a verbal and written format ▪ Understand about resource allocation and project planning ▪ Work through the problem-solving cycle

Assessment

Task	**Weighting**
Contribution to discussion boards	20%
Group presentation describing a health system or technology, its uses, strengths and weaknesses	20%
Written report detailing a plan and issues surrounding a system development	60%

Suggested reading materials

- A systems development life cycle approach to patient journey modelling projects.
- Curry JM, McGregor C, Tracy S. Stud Health Technol Inform. 2007; 129 (Pt 2):905-9.

Unit 4 – Applied Health Data Science

Mandatory pre-requisite units:

- Unit 1 - Maths and Statistics for Health Data Science
- Unit 2 - Data Engineering for Health Data Science
- Unit 3 - Systems and Technologies in Health Data Science

Aims

This unit aims to develop and integrate the strong healthcare acumen, problem-solving skills, communication and influencing skills with the technical and methodological skills required for a health data scientist working in an international context. It will provide the student with an understanding of the strategic importance of business intelligence and decision support in health.

Unit description

To improve the delivery of healthcare there is a need to maximise the potential of health data by turning it into useful intelligence to provide insights into past, current and future healthcare delivery. The role of the health data scientist is to be at the forefront of this and to influence decision-making for healthcare delivery by deriving understanding and significance from data. To be able to do this it is critical to understand the decision-making governance and cycle; know how to elicit understanding of the clinical and public health objectives and processes; understand the different data sources available to support decision-making, and which are appropriate to use; understand which techniques and methodologies are most appropriate to investigate data; and know how to communicate and visualise results /ideas to various stakeholders (often from a non-technical background; including the public); and determine how this will impact on healthcare. This unit will integrate technical and methodological skills with each of these issues and apply them to uses of 'big data' in healthcare.

Indicative content

- Governance - Decision-making process
- Data – sources, quality, technical, ethical and legal issues; linking data sources
- Data visualisation – techniques, software and presentation style
- Communication/presentation styles – requirements elicitation;
- Organisational and change management
- Risk management

Learning outcomes

Category of outcome	**Students should/will (please delete as appropriate) be able to:**
Knowledge and understanding	▪ Demonstrate a critical understanding of the pathway from data collection, interpretation, analysis, visualization and use. ▪ Evaluate the advantages and the technical, ethical and legal problems associated with the use of health data ▪ Explain the range of data sources available for analysis and discuss their characteristics ▪ Explain how large datasets are created and used to support planning/commissioning, research/funding decisions ▪ Understand the strategic importance of business intelligence, knowledge management, and decision support in healthcare ▪ Demonstrate a critical understanding of barriers when working with health data
Intellectual skills	▪ Identify and appraise data sources used to support healthcare decision making

	▪ Critically analyse a healthcare problem and provide a suitable strategy to address it
Practical skills	▪ Elicit information from various stakeholders using various practices ▪ Conduct a risk analysis focussed on a healthcare problem
Transferable skills and personal qualities	▪ Communicate effectively both in written and verbal format to both non-technical and technical audiences (including the public) ▪ Work effectively as a group

Assessment

Task	**Weighting**
Contribution to discussion boards	20%
Group presentation, presenting results of a health data analysis as if to a stakeholder.	20%
Short written report based on addressing a health (data) problem covering the entire health data science pathway from data collection, interpretation, analysis, visualization and how the results would be used.	60%