Imperial College London

Programme Specification (Master's Level)

MSc Health Data Analytics and Machine Learning

This document provides a definitive record of the main features of the programme and the learning outcomes that a typical student may reasonably be expected to achieve and demonstrate if s/he takes full advantage of the learning opportunities provided. This programme specification is primarily intended as a reference point for academic and support staff involved in delivering the programme and enabling student development and achievement, for its assessment by internal and external examiners, and in subsequent monitoring and review.

Programme Information				
Award(s)	MSc			
Programme Title	Health Data Analytics and Machine Learning			
Programme code	A3BJ			
Awarding Institution	Imperial College London			
Teaching Institution	Imperial College London			
Faculty	Medicine			
Department	• School of Public Health (lead department) • Data Science Institute			
Associateship	n/a			
Mode and Period of Study	1 calendar year full-time (12 months)			
Cohort Entry Points	Annually in October			
Relevant QAA Benchmark Statement(s) and/or other external reference points	Medicine and Mathematics.			
Total Credits	ECTS:	90	CATS:	180
FHEQ Level	Level 7 - Master's			
EHEA Level	2nd cycle			
External Accreditor(s)	None			
Specification Details				
Student cohorts covered by specification	2021-22			
Person Responsible for the specification	Jo Tite j.tite@imperial.ac.uk			
Date of introduction of programme	October 2018			

Date of programme specification/revision	September 2021

Description of Programme Contents

The MSc in Health Data Analytics is a 12 month full-time taught programme which aims at building a solid and common background in Health Data Analytics for students:

- with strong methodological backgrounds in mathematics and statistics who are willing to move to an applied field with vast potential for developing novel approaches to produce high impact results;
- with a more applied background in Epidemiology, Medical Sciences and Biology, willing to learn about how to handle high dimensional data and how to fully exploit these rich datasets in order to produce public health relevant knowledge.
- with another quantitative background willing to enrich their methodological portfolio with cutting edge methods that are useful to exploit complex and high dimensional data.

The main objective of the programme is to ensure that students are able to use appropriate cutting edge quantitative methods to fully exploit complex and high dimensional data. Students successfully completing the programme will have acquired a strong methodological background that is necessary to perform the in-depth analysis of medical and epidemiological high throughput data sets. This knowledge and experience will place students, following graduation, in an ideal position to potentially:

- pursue academic careers (through a PhD for instance) in many fields: molecular, environmental, chronic disease and computational epidemiology, public health, population sciences;
- work as an expert analyst in industries (pharma, food, cosmetics…) where these big data are also accumulating;
- engage in large data companies (e.g. Google, Deepmind-health) which are developing in the field of health and need trained data scientists with experience in exploring and mining medical and epidemiological data experience.

Programme Structure:

The programme is a full-time 12 month taught masters programme, which runs from October-September. The programme comprises nine taught core modules (60 ECTS credits) and one core 4-month research project module (30 ECTS credits), totalling 90 ECTS credits for the final award.

Term 1

During term 1, the programme will share two core modules with the existing MSc Epidemiology and Master of Public Health to create a solid and common epidemiological foundation for all students, a third core module bespoke to the MSc in 'Clinical Data Management', a ten-week core module in 'Molecular Epidemiology', and a five week core module Translational Data Science Part I as follows:

- Introduction to Statistical Thinking and Data Analysis
- Principles and Methods of Epidemiology
- Clinical Data Management
- Molecular Epidemiology
- Translational Data Science Part I

In addition, and in order to facilitate the integration of students with heterogeneous backgrounds, during term one the programme will offer optional additional workshops on mathematical and statistical principles, and, epidemiological and biological principles.

Term 2

The second term comprises four core taught modules and a series of research seminars at which attendance is mandatory, and focuses on more sophisticated statistical methods to perform the in-depth analysis of high throughput data, health data at the population level and machine learning approaches to explore 'big data', as follows:

- Machine Learning
- Computational Epidemiology
- Translational Data Science Part II
- Population health
- Research seminar series

The second term also includes the selection and development of core research project planning as part of the 4-month research project.

Term 3

The third term of the programme focuses predominately on the continuation and development of the core 4-month research project-based module dedicated to the application of the taught advanced techniques to real data sets. Throughout this project, students will be guided through all the steps of the process of conceiving and delivering (and assessing) a high standard scientific publication (term 3).

Teaching will be led by School of Public Health (SPH) in collaboration with the Data Science Institute (DSI) within Imperial College London. Building upon the complementarity of the existing international networks in both entities, the programme is designed to attract:

- Students with methodological background willing to move to a promising applied field
- Students with a medical/biological/epidemiological background willing to reinforce their analytical skills.

Teaching staff from both departments includes international experts with strong methodological background and expertise in the application of these approaches to large-scale medical and clinical data. Both departments have a large collaborative network with industries (e.g. through EiT – European Institute of Innovation and Technology).

Teaching throughout the programme will include project-based work for each specific module and for the 4-month research project. These will be based on real data and will address real scientific questions from research staff within School of Public Health and Data Science Institute and, where appropriate, industrial partners.

Learning Outcomes
Upon successful completion of the programme, students should be able to: 1. Understand in-depth the nature, the complexity and the richness of Medical/Clinical/Epidemiological/population-based 'big data'; 2. Independently and originally design, maintain, and query novel Medical/Clinical databases, to link and harmonise developed and existing databases and ensure that they comply with legal, ethical, security and privacy standards to ensure a secure use of medical and clinical data; 3. Independently design, implement and interpret results from established statistical models used to perform high throughout data profiling (univariate approaches, dimensionality reduction techniques, and variable selection); 4. Demonstrate a systematic understanding, knowledge and expertise in in machine learning approaches and their utility in medical data analytics and to be able to design, implement and interpret results from established machine learning approaches (including decision trees, SVM, and Bayesian/neural networks; 5. Comprehensively understand and identify the best (set of) statistical and/machine learning method(s) to address a specific research question considering the nature of the data, their specifics, be able to quantitatively compare the performances of different possible techniques; 6. Develop an advanced technical and applied knowledge ensuring full awareness of pressing research questions in Medical data analytics and proposed methodological solutions, including data integration, complex data bases structures; 7. Generate original thinking on how to use and combine existing techniques to address in the simplest and most interpretable way pressing questions in the field such as data integration, results visualisation and interpretation; 8. Identify original and relevant research questions from existing data sets and solicit appropriate techniques to address them; 9. Produce a sound scientific work from real data addressing an original research question and (i) using appropriate statistical approaches, and (ii) developing visualisation resources ensuring clear and convincing presentation of the work.
The Imperial Graduate Attributes are a set of core competencies which we expect students to achieve through completion of any Imperial College degree programme. The Graduate Attributes are available at: www.imperial.ac.uk/students/academic-support/graduate-attributes

Entry Requirements	
Academic Requirement	Normally at least a 2:1 in a science based UK bachelors honours degree or medical degree or equivalent qualification in (maths, statistics, epidemiology biology).
Additional Requirements	None

Applicants who do not meet the academic requirements above but who have substantial relevant academic or professional experience may be admitted following completion of a 'Special Qualifying Exam' (SQE).	
Depending on the strength of the application, including personal statement, references and those that do not meet the academic requirements for entry, some applicants may be invited to attend a post-application interview to assess the suitability of entry to the programme.	
English Language Requirement	IELTS 7 with a minimum of 6.5 in each element or equivalent.
Learning & Teaching Strategy	
Scheduled Learning & Teaching Methods	• Lectures and associated practicals • Seminars and technical tutorials • Small group tutorials • Group work sessions • Computer based practical workshops • Group revision sessions
E-learning & Blended Learning Methods	• The Imperial VLE blackboard will be used for access to lecture materials, assessment information and practical questions/answers
Project Learning Methods	• Group mini research projects including an oral presentation of the results and/or production of a written report. • Preparation and peer-evaluation of a scientific paper and corresponding presentation • The final individual research project
Placement Learning Methods	• None
Assessment Strategy	
Assessment Methods	• Written examinations • Oral presentations • Online quizzes • Short answer essay questions • Real case analyses reports • Final research paper • Research plan • Project journal and log
Academic Feedback Policy	
The programme complies with the College policy and will provide students with feedback within a timely and appropriate turnaround time. Students will be provided with a bespoke assessment	

schedule confirming all submission deadlines, marking periods and feedback points to manage their expectations and ensure feedback can inform the next assessment.

Feedback will take different forms according to the type of assessment:

- Exams and written reports : students will be given written feedback, and given the opportunity to discuss their provisional marks and possible improvements with the module leaders;
- Oral presentations: comments and immediate feedback will be given during the presentation and further discussion will be organised once provisional marks are communicated to students;
- Final research project: students be will given written feedback for their final project/presentation and also receive oral feedback following the oral presentation.

Further exchanges will also take place throughout the programme through:

- Questions during or after the lectures;
- Interaction with tutors during practicals and 1-2-1.

Re-sit Policy

Students will be permitted to re-enter a failed examination or resubmit a piece of failed coursework on a single occasion. Examination re-sits may only be available at the next available sitting (i.e. in the next following academic year).

Mitigating Circumstances Policy

The College's Policy on Mitigating Circumstances is available at: www.imperial.ac.uk/registry/exams

Programme Structure

Full-time	Pre-session	Term One	Term Two	Term Three	Term Four
Core Modules	0	4	4	0	0
Projects	0	0	0	1	

Assessment Dates & Deadlines

Written Examinations	End of Term 1 modules (January) End of Term 2 modules (April/May)
Coursework Assessments	Continuous
Project Deadlines	End of Term 1 modules (January) End of Term 2 modules (April/May) Term 3 research project (Late August)
Practical Assessments	Continuous Research Project Oral (Mid/late September)

Assessment Structure

Programme Component	ECTS	% Weighting
Introduction to Statistical Thinking and Data Analysis	7.5	8.33%
Principles and Methods of Epidemiology	7.5	8.33%
Clinical Data Management	7.5	8.33%
Molecular Epidemiology	7.5	5.56%
Computational Epidemiology	7.5	8.33%
Machine Learning	7.5	8.33%
Translational Data Sciences	7.5	8.33%
Population Health	5	5.56%
Research seminar series	2.5	5.56%
4-month Research project	30	33.3%
Total	90	100%

Marking Scheme

The MSc can be awarded as a Pass, Merit or Distinction.

In order to be awarded a **Pass for the Masters award**, a student must:

- Achieve a mark of at least 40% in each assessment
- Achieve an aggregate pass mark of at least 50% in each module
- Students may be condoned in a maximum of 15 ECTS for the Masters award (i.e. excluding the Research Project) with an aggregate mark of at least 40% in each module providing the overall aggregate mark for the programme is at least 50%
- Achieve a mark of at least 50% in the Research Project

In order to be awarded a **Merit for the Masters award**, a student must:

- Achieve a mark of at least 40% in each assessment
- Achieve an aggregate mark of at least 60% in each module
- Students may be condoned in a maximum of 15 ECTS for the Masters award (i.e. excluding the Research Project) with an aggregate mark of at least 40% in each module providing the overall aggregate mark for the programme is at least 60%
- Achieve a mark of at least 60% in the Research Project

In order to be awarded a **Distinction for the Masters award**, a student must:

- Achieve a mark of at least 40% in each assessment
- Achieve an aggregate pass mark of at least 70% in each module

- Students may be condoned in a maximum of 15 ECTS for the Masters award (i.e. excluding the Research Project) with an aggregate mark of at least 40% in each module providing the overall aggregate mark for the programme is at least 70%
- Achieve a mark of at least 70% in the Research Project

Exit award

In exceptional circumstances, students may be permitted to be awarded a PG Certificate as an exit award, subject to examination board approval and on successful completion of a minimum of 30 ECTS in the taught modules.

- A **Postgraduate Certificate** (upon passing and completing the four core modules during term one) totalling 27.5 ECTS modules, subject to exam board approval. The modules required are:
    - Introduction to Statistical Thinking and Data Analysis (7.5 ECTS)
    - Principles and Methods of Epidemiology (7.5 ECTS)
    - Molecular Epidemiology (5 ECTS)
    - Clinical Data Management (7.5 ECTS)

- A **Postgraduate Diploma** (upon passing and completing the first four core modules during term one, and, four core modules from term totalling 60 ECTS modules, subject to exam board approval. The modules required are those detailed above, and:
    - Machine Learning (7.5 ECTS each)
    - Translational Data Sciences (7.5 ECTS)
    - Population Health (5ECTS)
    - Research seminar series (5 ECTS)
    - Computational Epidemiology (7.5 ETCS)

In order to be awarded a **PG Certificate and PG Diploma exit award**, a student must:

- Achieve a mark of at least 40% in each assessment
- Achieve an aggregate pass mark of at least 50% in each module
-

Indicative Module List												
Code	Title	Core/ Elective	Year	L&T Hours	Ind. Study Hours	Place-ment Hours	Total Hours	% Written Exam	% Course-work	% Practical	FHEQ Level	ECTS
PUBH97052	Introduction to Statistical Thinking and Data Analysis	Core	1	45	142.5	0	187.5	40%	40%	20%	7	7.5
PUBH97002	Principles and Methods of Epidemiology	Core	1	35	152.5	0	187.5	35%	65%	0%	7	7.5
PUBH97027	Clinical Data Management	Core	1	60	127.5	0	187.5	70%	30%	0%	7	7.5
PUBH97060	Molecular Epidemiology	Core	1	40	147.5	0	125	100%	0%	0%	7	7.5
PUBH97028	Computational Epidemiology	Core	1	40	147.5	0	187.5	50%	0%	50%	7	7.5
PUBH97029	Machine Learning	Core	1	40	147.5	0	187.5	50%	0%	50%	7	7.5
PUBH97030	Translational Data Sciences	Core	1	40	147.5	0	187.5	0%	80%	20%	7	7.5
PUBH97061	Population Health Science	Core	1	30	95	0	125	100%	0%	0%	7	5
PUBH97062	Research Seminar Series	Core	1	20	20	0	40	100%	0%	0%	7	2.5
PUBH97031	Research project	Core	1	50	950	0	1000	0%	80%	20%	7	30

Supporting Information
The Programme Handbook is available at: TBC
The Module Outlines are available at: TBC
The College's entry requirements for postgraduate programmes can be found at: www.imperial.ac.uk/study/pg/apply/requirements
The College's Quality & Enhancement Framework is available at: www.imperial.ac.uk/registry/proceduresandregulations/qualityassurance
The College's Academic and Examination Regulations can be found at: http://www3.imperial.ac.uk/registry/proceduresandregulations/regulations
Imperial College is an independent corporation whose legal status derives from a Royal Charter granted under Letters Patent in 1907. In 2007 a Supplemental Charter and Statutes was granted by HM Queen Elizabeth II. This Supplemental Charter, which came into force on the date of the College's Centenary, 8th July 2007, established the College as a University with the name and style of "The Imperial College of Science, Technology and Medicine". http://www.imperial.ac.uk/admin-services/secretariat/college-governance/charters-statutes-ordinances-and-regulations/
Imperial College London is regulated by the Higher Education Funding Council for England (HEFCE) http://www.hefce.ac.uk/reg/of/

Term 1	05-Oct	12-Oct	19-Oct	26-Oct	02-Nov	09-Nov	16-Nov	23-Nov	30-Nov	07-Dec	14-Dec			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Monday	Induction	**Core Module 1. Statistical Thinking and Data Analysis** (With MSc Epi)										**Independent Study, Assessment Preparation & revision**		
Tuesday		**Core Module 2. Principles and Methods of Epidemiology** (With MSc Epi & MPH)												
Weds AM		**Optional workshops and Independent Study**												
Weds PM		**Optional core mathematics classes**												
Thursday		**Core Research Project Planning and Independent Study**						**Core Module 4 Part I. Molecular epidemiology**						
Friday		**Core Module 3. Clinical Data Management**												

Term 2	11-Jan	18-Jan	25-Jan	01-Feb	08-Feb	15-Feb	22-Feb	01-Mar	08-Mar	15-Mar	22-Mar					
	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Monday	**Exam Week 1**	**Core Module 5. Machine Learning**										**Core Research Project Planning, Assessment Preparation and Independent Study**				
Tuesday		**Core Module 4 Part II. Population science**				**Core Research Project Planning and Independent Study**										
Weds AM		**Core Module 8. Health Data Research Seminar Series**														
Weds PM		**Independent Study & revision**														
Thursday		**Core Module 6. Computational Epidemiology**														
Friday		**Core Module 7. Translational Data Science**														

Term 3	03-May	10-May	17-May	24-May	31-May	07-Jun	Jun-Sept		06-Sep	20-Sep
	31	32	33	34	35	36	37-48	48	49-52	
Monday	**Exam Week 2**	**Full-time research project (30 ECTS)**							**Dissertation submission & final oral exam**	
Tuesday										
Wednesday										
Thursday										
Friday										