



Job Description

Research Fellow – Health Data Scientist **Grade: 7**

Department: Institute of Health Informatics

Location: 222 Euston Road, London/Hybrid

Reports to

Dr Ana Torralbo, Senior Data Scientist.

Context

UCL established the Institute of Health Informatics (IHI) in August 2014 in the Faculty of Population Health Sciences, under the leadership of Professor Harry Hemingway.

The Institute of Health Informatics is part of the [UCL Faculty of Population Health Sciences](#),

The aim of the Institute of Health Informatics is to conduct quality research that leverages big data and health and bioinformatics approaches to improve health at local, national and international levels. Our cutting edge programmes of research span the areas of Discovery Science, Precision Medicine, Learning Health Systems, Public Health and Citizen Driven Health. These programmes of research are underpinned by: excellent secure data infrastructure; adding value to key linked data resources to make data research-ready; development of multidisciplinary methodological expertise; engagement with patients, the public and health services; strong local, national and international academic partnerships and excellent multidisciplinary training programmes. The Institute leads major programmes in the NIHR UCLH/UCL Biomedical Research Centre, CALIBER, Health Data Research UK (HDR-UK), Genomics England Electronic Health Records Clinical Interpretation Partnership (GeCIP), Big Data@Heart among others.

Main purpose of the job

The Research Fellow will join a multidisciplinary team of researchers (e.g. computer science, health data science, epidemiology, clinicians) to develop next generation methods and tools for defining disease phenotypes in complex health data.

Building on our highly successful collaboration with our industry partner GlaxoSmithKline which generated state-of-the-art disease definitions in the UK Biobank, we are extending our partnership to develop cutting edge EHR phenotyping infrastructure across major biobanks (UK Biobank, Genes and Health and Our Future Health).

The project will deliver two major work packages: (1) To develop and apply novel disease and clinical biomarker phenotypes that are reproducible and replicable across multiple EHR-linked biobanks; and (2) To establish and implement a bi-directional knowledge exchange to drive disease severity and progression phenotyping projects.

This role offers a unique opportunity to be involved in working with world-leading experts in the field of phenomics and computational medicine in academia and industry to deliver novel data-derived knowledge and infrastructure to support cutting-edge research. The role also provides valuable learning around industry consultancy through the knowledge exchange programme.

The Research Fellow will work closely with the project team consisting of other health data scientists, a project manager and a project administrator to deliver on creating and evaluating novel methods for defining phenotyping algorithms in contemporary data sources and data modalities.

Duties and responsibilities

- Work within relevant secure data environments to create and evaluate phenotyping algorithms leveraging controlled clinical terminologies (e.g. ICD-10, SNOMED CT) on multiple linked electronic health record sources
- Under supervision, use Python (or R) to develop detailed analytic plans and undertake statistical analyses for evaluating phenotypes and comparing results across data sources e.g. comparing prevalence and incidence rates among different groups, and other epidemiological measures with external datasets
- Carry out data management tasks to prepare data for analyses, including data extraction and data quality assurance
- Generate efficient and shareable code which can be used in different datasets
- Participate in weekly internal project meetings to plan and review progress against project deliverables and prepare documents for discussion at the meetings
- Prepare research reports and present findings at meetings with the project funder and at conferences
- Author publications and/or conference abstracts to present results
- Engage with our industry partner scientists to help plan, and participate in, bi-monthly knowledge exchange sessions based on the priorities and questions driven by our industry partner scientists, for example providing expert knowledge around development of new methods
- Contribute to teaching and supervision of MSc students in the Institute as appropriate

Person Specification

Criteria	Essential or Desirable	Assessment method (Application/Interview)
Qualifications, experience and knowledge		
PhD in a relevant discipline (e.g. computer science, health informatics, bioinformatics) or similar level of experience in an academic and/or industrial setting	E	A/I
Good knowledge of the Python scientific programming stack (e.g. pandas, numpy, jupyter) and best practices for software development (e.g. version control, unit tests)	E	A/I
Experience of manipulating and analysing large, high-dimensional electronic health record datasets and practical knowledge of controlled clinical terminologies (e.g. SNOMED, ICD-10, dm+d)	E	A/I
Experience creating and evaluating phenotyping algorithms for identifying clinically important features in large complex datasets	E	A/I
Experience working with relational databases (e.g. SQL, sqlite) for manipulating electronic health record and clinical data	E	A/I
A working understanding of information governance, privacy and security issues	D	A/I
Experience working with large cohorts and datasets (e.g. UK Biobank or Genes and Health) and electronic health records from routine sources	D	A/I
Skills and abilities		
Ability to manage own time	E	A/I
Ability to clearly present the methods and results of statistical analyses verbally and in writing	E	A/I
Commitment to high quality academic research	E	A/I
Personal attributes		
Self-motivated and ability to work on own initiative	E	A/I
Willingness and ability to work collaboratively with colleagues	E	A/I
Commitment to continuous professional development	E	A/I