Gibran Hemani

Senior Research Fellow, University of Bristol

8 Picton Mews, Bristol, BS6 5PB, United Kingdom

+44(0)7930951876 | g.hemani@bristol.ac.uk | gibranhemani.com | explodecomputer

~

Present Appointment

MRC Integrative Epidemiology Unit, University of Bristol — UK
Senior research fellow — Jan-18 to present

Previous Appointments

MRC Integrative Epidemiology Unit, University of Bristol — UK
Research fellow — Jan-14 to Dec-17

Queensland Brain Institute, University of Queensland — Australia
Post doctoral statistical geneticist — Jan-12 to Dec-13

Academic Qualifications

University of Bristol — Bristol
Associate of the Higher Education Academy, UK — Sep-19 to Mar-20

University of Edinburgh — Edinburgh
PhD in Quantitative Genetics — Oct-07 to Aug-11

University of Nottingham — Nottingham
BSc (hons) 2:1 — Sep-03 to Jun-06

Special Awards, Honours and Distinctions

Sir Kenneth Mather Memorial prize — The Genetics Society
Best PhD thesis in quantitative and population genetics — 2011

Teaching and related administration

Over the course of my research career I have sought out opportunities to develop a variety of different teaching materials, starting during my PhD when I wrote and delivered a 2-day course on R programming for researchers at the Roslin Institute. I have a passion for teaching and since then I have taught on topics relating to genetics, statistics and programming to undergraduates across disciplines in medicine, biology and psychology. There has been a major shift towards using genetic data in epidemiological research over the last several years and my academic background in statistical genetics has created opportunities and substantial demand for teaching in this field at the post-graduate and life-long learning levels. I am particularly proud of my record of writing and delivering courses on genetic epidemiology to participants who are under-represented in genomic data science. In line with this I am also leading the Medical School's Anti-Racism special interest group in Decolonising the Curriculum.

(i) Undergraduate and taught postgraduate (past 3 years)

Evidence Based Medicine unit in MB ChB programme — University of Bristol
Tutor — 2020

- 15 1st year undergraduate medical students
- 5 hours prep, 10 hours contact time

Genes and behaviour (PSYC30018) — University of Bristol
Lecturer — 2018, 2019

- 100 3rd year undergraduate students
- Wrote and delivered three lectures
- 6 hours prep, 4 hours contact time

Genomic Medicine iBSc — University of Bristol
Pastoral tutoring — 2017, 2018

- Two students on the Genomic Medicine iBSc
- 8 hours contact time

Genomic Data Science unit in Genomic Medicine iBSc — University of Bristol
Course lead — 2016-2018, 2020

- Up to 13 intercalating 3rd year medical dentistry and veterinary students
- Led, co-wrote and delivered 4-week module (20 credits)
- Organised material for several lecturers on programming, statistics, genetic analysis
- Set formative and summative coursework assessments
- Designed and wrote exams
- Delivered lectures, tutorials, practicals, assessed debates
- Transitioned to flipped classroom format
- 50 hours prep, 20 hours contact time
- Several students publish papers each year on the subject.

(ii) Major teaching responsibilities in previous years

Statistical genetics unit in Biomedical Capstone Course — University of Queensland
Course lead — 2012, 2013

- 250 3rd year undergraduate biomedical students
- Wrote four lectures and two workshops on GWAS written
- Set and marked coursework assessments
- Delivered lectures and practicals, and coordinated lecturers and tutors
- 25 hours prep, 10 hours contact time

(iii) Innovatory units and teaching methods

Flipped classroom: In the second year of the iBSc Genomic Medicine course I redesigned the R programming training section to be a flipped classroom, using online games for students to learn the basic principles of programming at home and the tutor led sessions as an opportunity to synthesise those skills into applied examples. The students were more adept at programming than in the previous year, and also more enthused about the subject. I transitioned the rest of the module to a flipped classroom in 2020 during the Covid-19 pandemic.

(iv) Contribution to Life-long learning and continuing professional development courses

Genetic Analysis of Population-based Association Studies short course, Wellcome Genome Campus
Course co-lead 2018-2020

- 40+ international post-graduate students and researchers
- Invited to co-lead a course that has been running for 12 years, approx
- Contribute to course design, and recent redesign for online teaching
- Delivering lectures and practicals
- 24 hours prep, 10 hours contact time

MR-Base workshop, MR conference Bristol
Course lead 2017, 2019

- 60 International post-graduate students and researchers
- Wrote and delivered lectures and tutorials on how to use the MR-Base database and R packages
- 12 hours prep, 6 hours contact time

Statistical methods for mediation short course University of Bristol
Lecturer 2017, 2018

- 40 post-graduate students and researchers
- Wrote and delivered one lecture + practical
- 8 hours prep, 3 hours contact time

Genomic medicine iBSc University of Bristol
Lecturer 2016-2018, 2020

- Up to 13 intercalating 3rd year medical, dentistry and veterinary students
- Wrote and delivered three lectures to other units
- 8 hours prep, 3 hours contact time

Genetic Epidemiology short course at EEPE Florence, Italy
Course co-lead 2016-2018

- 20 international post-graduate students and researchers
- Co-wrote and delivered (with Prof Dave Evans) 5-day course
- Prepared and delivered lectures + practicals
- 40 hours prep, 12 hours contact time

Statistical genetics short course University of Bristol
Lecturer 2015-2019

- 40 post-graduate students and researchers
- Wrote and delivered two lectures + practicals
- 16 hours prep, 6 hours contact time

UNIX and Genetic epidemiology Pelotas, Brazil

Course co-lead 2015

- 30 researchers and post-graduate students
- Co-wrote and delivered 7-day course
- Lectures and practicals
- Setup cloud-based computing system for practical sessions
- 40 hours prep, 15 hours contact time

Mendelian randomisation short course University of Bristol

Lecturer 2014-2020

- 40 post-graduate students and researchers
- Wrote and delivered two lectures + practicals
- 16 hours prep, 6 hours contact time

Genetic epidemiology, H3Africa project Johannesberg, South Africa

Course co-lead 2014

- 30 researchers and post-graduate students
- Co-wrote and delivered (with Dr Nic Timpson) 5-day course
- Delivered lectures and practicals
- 50 hours prep, 20 hours contact time

18th Summer Institute in Statistical Genetics Seattle USA

Teaching assistant 2013

- 40+ international post-graduate students and researchers
- Teaching assistant for: “Human Complex Traits” and “Animal Genetic Data Analysis”
- 20 hours prep, 10 hours contact time

Introduction to git and programming workflows University of Queensland

Workshop lead 2013

- 15 researchers
- Wrote and delivered one-day workshop
- 12 hours prep, 6 hours contact time

Introduction to Statistics University of Queensland

Lecturer 2012

- 40+ post-graduate students and researchers
- Wrote and delivered one lecture for a professional Development Course
- 3 hours prep, 1 hour contact time

Introduction to R University of Edinburgh

Course co-lead 2009

- 30 researchers and post-graduate students
- Co-wrote and delivered (with Joseph Powell) 2-day short course on R programming
- 15 hours prep, 6 hours contact time

(v) Collaborative teaching projects

Developed external speaker programmes for iBSc medical students, involving Jeff Barrett from OpenTargets and the Sanger Institute; Rob Scott from GlaxoSmithKlein; and Jonathan Ives from the Centre for Ethics in Medicine.

(iv) Postgraduate advising

PhD Supervision

Lily Andrews
CRUK 2020-2024

- Secondary supervisor

Amanda Forde
Science Foundation Ireland 2020-2024

- Secondary supervisor
- Co-supervising with international colleague (John Ferguson), Republic of Ireland

Giulio Centorame
NHMRC 2020-2024

- Secondary supervisor
- Co-supervising with international colleague (Dave Evans), Australia

Chris Moreno-Stokoe
BBSRC 2018-2022

- Secondary supervisor

Hannah Wilson
BBSRC and GSK 2017-2021

- Primary supervisor

Thomas Battram
Wellcome Trust 2016-2020

- Primary supervisor
- Passed viva with distinction

Laurence Howe
Wellcome Trust 2014-2018

- Secondary supervisor
- Passed viva with distinction
- Awarded best doctoral research prize 2018/2019 in Faculty of Health Sciences

(vii) Major achievements in teaching administration

International short courses on statistical genetics: I have organised, developed and delivered short courses (typically week-long, in collaboration with internal and/or external colleagues) to researchers at institutes in South Africa, Italy, Brazil and at the Wellcome Genome Centre in Cambridge.

Genomic Data Science module: I designed, organised and delivered a 20 credit module for medical students that involves teaching them programming and genomic data analysis from scratch. This has run for 4 years, and consistently gets very positive feedback, and the students achieve a high standard.

Anti-racism: I lead the Special Interest Group (11 members, fortnightly meetings) on decolonising the curriculum in the Bristol Medical School. We are developing an accreditation system that schedules a transparent process for course leads to work against. I also liaise with numerous related groups across the university.

MR-Base software course: I developed a course to teach researchers how to use my own software (MR-Base), with the course being heavily over-subscribed and gaining very positive feedback and increased usage of the software.

Research and related administration

During my PhD I developed statistical and computational methods to detect genetic interactions influencing complex traits. This work led to my first post-doctoral position that expanded my work into determining genetic architectures of complex disease, and developing expertise in handling high dimensional genomic data. My current research lies at the interface between statistical genetics, causal inference and high performance computing. A major output that integrates these fields has been the development of the MR-Base causal inference analytical platform. This software is used by researchers globally, has led directly to funding for five post-doctoral positions from three private companies (GSK, Biogen, CHDI). I currently lead or co-lead two major international collaborations: The GoDMC consortium comprising 56 cohorts analysing the genetics of DNA methylation; the OpenGWAS consortium that combines the resources of over 100 groups and consortia. I synthesise these activities under my current fellowship which aims to construct a map of causal relationships across genotypes and phenotypes, and to use that map in diverse ways, for example to improve our understanding of human evolutionary processes, and to improve the prediction of viable drug targets for disease.

(i) Publications

Selected publications organised by category are listed below. Full publication list on google scholar.

- H-index: **47**
- i10-index: **89**
- Number of publications: **150**

Academic journal papers (refereed)

1. Lyon, M., Andrews, S., Elsworth, B., Gaunt, T., Hemani, G. & Marcora, E. The variant call format provides efficient and robust storage of gwas summary statistics. *Genome Biology* (2021) Contrib.: 30% (senior author). International co-authors. Position: 5/6 (joint last).

2. Griffith, G., Morris, T., Tudball, M., Herbert, A. & Mancano, G. *et al.* Collider bias undermines our understanding of covid-19 disease risk and severity. *Nature communications* (2020) Contrib.: 40% (senior author). Position: 14/14.

3. Brumpton, B., Sanderson, E., Heilbron, K., Hartwig, F. & Harrison, S. *et al.* Avoiding dynastic, assortative mating, and population stratification biases in mendelian randomization through within-family analyses. *Nature communications* (2020) Contrib.: 30% (senior author). Position: 32/33 (joint last).

4. Lawson, D., Davies, N., Haworth, S., Ashraf, B. & Howe, L. *et al.* Is population structure in the genetic biobank era irrelevant, a challenge, or an opportunity? *Human Genetics* (2020) Contrib.: 10% (senior author). Position: 7/9.

5. Zheng, J., Haberland, V., Baird, D., Walker, V. & Haycock, P. *et al.* Phenome-wide mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nature Genetics* (2020) Contrib.: 20% (senior author). International co-authors. Non-academic co-authors. Position: 32/34 (joint last).

6. Morris, T., Davies, N., Hemani, G. & Smith, G. Population phenomena inflate genetic associations of complex social traits. *Science Advances* (2020) Contrib.: 20% (senior author).

7. Anderson, E., Howe, L., Wade, K., Ben-Shlomo, Y. & Hill, W. *et al.* Education, intelligence and alzheimer's disease: Evidence from a multivariable two-sample mendelian randomization study. *International journal of epidemiology* (2020) Contrib.: 30% (senior author). Position: 13/13.

8. Richardson, T., Hemani, G., Gaunt, T., Relton, C. & Smith, G. A transcriptome-wide mendelian randomization study to uncover tissue-dependent regulatory mechanisms across the human phenome. *Nature communications* (2020) Contrib.: 20% (main analyst).

9. Cho, Y., Haycock, P., Sanderson, E., Gaunt, T. & Zheng, J. *et al.* Exploiting horizontal pleiotropy to search for causal pathways within a mendelian randomization framework. *Nature communications* (2020) Contrib.: 40% (senior author). Position: 8/8.

10. Russell, A., Ford, T., Gunnell, D., Heron, J. & Joinson, C. *et al.* Investigating evidence for a causal association between inflammation and self-harm: A multivariable mendelian randomisation study. *Brain, behavior, and immunity* (2020) Contrib.: 20% (senior author). Position: 9/10 (joint last).

11. Richardson, T., Harrison, S., Hemani, G. & Smith, G. An atlas of polygenic risk score associations to highlight putative causal relationships across the human phenome. *Elife* (2019) Contrib.: 20% (main analyst).

12. Morris, A., Le, T., Wu, H., Akbarov, A. & Most, P. *et al.* Trans-ethnic kidney function association study reveals putative causal genes and effects on kidney-specific disease aetiologies. *Nature communications* (2019) Contrib.: 5% (main analyst). International co-authors. Position: 6/72.

13. Taylor, D., Jackson, A., Narisu, Hemani, G. & Erdos, M. *et al.* Integrative analysis of gene expression, dna methylation, physiological traits, and genetic variation in human skeletal muscle. *Proceedings of the National Academy of Sciences* (2019) Contrib.: 15% (main analyst). International co-authors. Position: 4/22.

14. Howe, L., Lawson, D., Davies, N., Pourcain, B., Lewis, S., Smith, G. & Hemani, G. Genetic evidence for assortative mating on alcohol consumption in the uk biobank. *Nature communications* (2019) Contrib.: 30% (senior author). Position: 7/7.

15. Howe, L., Richardson, T., Arathimos, R., Alvizi, L. & Passos-Bueno, M. *et al.* Evidence for dna methylation mediating genetic liability to non-syndromic cleft lip/palate. *Epigenomics* (2019) Contrib.: 20% (senior author). Position: 17/18.

16. Hemani, G., Zheng, J., Elsworth, B., Wade, K. & Haberland, V. *et al.* The mr-base platform supports systematic causal inference across the human phenome. *elife* (2018) Contrib.: 30% (main author). International co-authors. Position: 1/20.

17. Min, J., Hemani, G., Smith, G., Relton, C. & Suderman, M. Meffil: Efficient normalization and analysis of very large dna methylation datasets. *Bioinformatics* (2018) Contrib.: 20% (main analyst).

18. Haas, M., Aragam, K., Emdin, C., Bick, A. & Hemani, G. *et al.* Genetic association of albuminuria with cardiometabolic disease and blood pressure. *The American Journal of Human Genetics* (2018) Contrib.: 10% (main analyst). International co-authors. Position: 5/8.

19. Richardson, T., Haycock, P., Zheng, J., Timpson, N. & Gaunt, T. *et al.* Systematic mendelian randomization framework elucidates hundreds of cpg sites which may mediate the influence of genetic variants on disease. *Human molecular genetics* (2018) Contrib.: 30% (senior author). Position: 8/8.

20. Howe, L., Lee, M., Sharp, G., Smith, G. & Pourcain, B. *et al.* Investigating the shared genetics of non-syndromic cleft lip/palate and facial morphology. *PLoS genetics* (2018) Contrib.: 20% (senior author). Position: 16/17.

21. Ye, J., Richardson, T., McArdle, W., Relton, C., Gillespie, K., Suderman, M. & Hemani, G. Identification of loci where dna methylation potentially mediates genetic risk of type 1 diabetes. *Journal of autoimmunity* (2018) Contrib.: 30% (senior author). Position: 7/7.

22. Hemani, G., Tilling, K. & Smith, G. Orienting the causal relationship between imprecisely measured traits using gwas summary data. *PLoS genetics* (2017) Contrib.: 70% (main author).

23. Noyce, A., Kia, D., Hemani, G., Nicolas, A. & Price, T. *et al.* Estimating the causal influence of body mass index on risk of parkinson disease: A mendelian randomisation study. *PLoS medicine* (2017) Contrib.: 10% (main analyst).

Position: 3/19.

24. Richardson, T., Zheng, J., Smith, G., Timpson, N., Gaunt, T., Relton, C. & Hemani, G. Mendelian randomization analysis identifies cpg sites as putative mediators for genetic influences on cardiovascular disease risk. *The American Journal of Human Genetics* (2017) Contrib.: 30% (senior author). Position: 7/7.

25. Gaunt, T., Shihab, H., Hemani, G., Min, J. & Woodward, G. *et al.* Systematic identification of genetic influences on methylation across the human life course. *Genome biology* (2016) Contrib.: 20% (main author). Position: 3/14 (joint first).

26. White, J., Sofat, R., Hemani, G., Shah, T. & Engmann, J. *et al.* Plasma urate concentration and risk of coronary heart disease: A mendelian randomisation analysis. *The lancet Diabetes & endocrinology* (2016) Contrib.: 10% (main analyst). Position: 3/52.

27. Yang, J., Bakshi, A., Zhu, Z., Hemani, G. & Vinkhuyzen, A. *et al.* Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nature genetics* (2015) Contrib.: 5% (main analyst). International co-authors. Position: 4/25.

28. Robinson, M., Hemani, G., Medina-Gomez, C., Mezzavilla, M. & Esko, T. *et al.* Population genetic differentiation of height and body mass index across europe. *Nature genetics* (2015) Contrib.: 15% (main analyst). International co-authors. Position: 2/43.

29. Visscher, P., Hemani, G., Vinkhuyzen, A., Chen, G.-B. & Lee, S. *et al.* Statistical power to detect genetic (co) variance of complex traits using snp data in unrelated samples. *PLoS Genet* (2014) Contrib.: 10% (main analyst). International co-authors. Position: 2/8.

30. Hemani, G., Shakhbazov, K., Westra, H.-J., Esko, T. & Henders, A. *et al.* Detection and replication of epistasis influencing transcription in humans. *Nature* (2014) Contrib.: 50% (main author). International co-authors. Position: 1/14.

31. McRae, A., Powell, J., Henders, A., Bowdler, L. & Hemani, G. *et al.* Contribution of genetic variation to transgenerational inheritance of dna methylation. *Genome biology* (2014) Contrib.: 10% (main analyst). International co-authors. Position: 5/10.

32. Hemani, G., Knott, S. & Haley, C. An evolutionary perspective on epistasis and the missing heritability. *PLoS Genet* (2013) Contrib.: 80% (main author).

33. Hemani, G., Yang, J., Vinkhuyzen, A., Powell, J. & Willemsen, G. *et al.* Inference of the genetic architecture underlying bmi and height with the use of 20,240 sibling pairs. *The American Journal of Human Genetics* (2013) Contrib.: 40% (main author). International co-authors. Position: 1/23.

34. Speed, D., Hemani, G., Johnson, M. & Balding, D. Improved heritability estimation from genome-wide snps. *The American Journal of Human Genetics* (2012) Contrib.: 20% (main analyst). International co-authors.

35. Hemani, G., Theocharidis, A., Wei, W. & Haley, C. EpiGPU: Exhaustive pairwise epistasis scans parallelized on consumer level graphics cards. *Bioinformatics* (2011) Contrib.: 80% (main author).

Review articles

1. Hemani, G., Bowden, J. & Smith, G. Evaluating the potential role of pleiotropy in mendelian randomization studies. *Human molecular genetics* (2018) Contrib.: 60% (main author).

2. Richmond, R., Hemani, G., Tilling, K., Smith, G. & Relton, C. Challenges and novel approaches for investigating molecular mediation. *Human molecular genetics* (2016) Contrib.: 20% (main author).

3. Smith, G. & Hemani, G. Mendelian randomization: Genetic anchors for causal inference in epidemiological studies. *Human molecular genetics* (2014) Contrib.: 50% (main author).

4. Wei, W.-H., Hemani, G. & Haley, C. Detecting epistasis in human complex traits. *Nature Reviews Genetics* (2014) Contrib.: 40% (main author). International co-authors.

Selected published open source software

AscRtain 2020

- Sensitivity analysis for collider bias in observational data
- https://github.com/explodecomputer/epigpu

GoDMC 2019

- Website and API for querying genetic assocations with DNA methylation
- http://mqtldb.godmc.org.uk/

MR-TRYX 2019

- Exploiting horizontal pleiotropy in Mendelian randomization
- https://explodecomputer.github.io/tryx/

OpenGWAS 2019

- The OpenGWAS data infrastructure
- https://gwas.mrcieu.ac.uk/

USS pension model 2018

- Web-app for projected pensions across different valuations
- http://www.uss-pension-model.com/

ALSPAC data dictionary 2017

- R package and web-app for searching for ALSPAC variables
- http://variables.alspac.bris.ac.uk/

MR-Base 2016

- Automated Mendelian randomization
- https://www.mrbase.org/

simulateGP 2016

- Simulation methods for genotype-phenotype associations
- https://explodecomputer.github.io/simulateGP/

gctaPower 2013

- Power calculations for genomic REML analysis
- https://shiny.cnsgenomics.com/gctaPower/

epiGPU 2012

- Exhaustive searches for genetic interactions parallelised across graphics cards
- https://github.com/explodecomputer/epigpu

(ii) Forthcoming publications

1. Min, J., Hemani, G., Hannon, E., Dekkers, K. & Castillo-Fernandez, J. *et al.* Genomic and phenomic insights from an atlas of genetic effects on dna methylation. *medRxiv* (2020) Contrib.: 30% (main author). Accepted in Nature Genetics. Position: 2/150 (joint first).

2. Sanderson, E., Richardson, T., Hemani, G. & Smith, G. The use of negative control outcomes in mendelian randomisation to detect potential population stratification or selection bias. *BioRxiv* (2020) Contrib.: 20% (senior author). Accepted in International Journal of Epidemiology.

(iii) Research grants

Total income as PI, from 5 grant(s): **2,176,539 GBP**

Total income as CI, from 5 grant(s): **1,480,672 GBP**

Current income as PI, from 3 grant(s): **2,055,459 GBP**

Current income as CI, from 1 grant(s): **284,524.6 GBP**

Genetic architecture of Huntington's disease progression (Contracts pending)
Cure Huntington's Disease Initiative 2020

- Amount: 598,881 GBP ; Role: PI ; Dates: 2021-01-01 to 2022-12-31 ; Proportion: 10%

Aetiological Epidemiology
Biogen 2020

- Amount: 284,525 GBP ; Role: Co-I ; Dates: 2020-09-01 to 2022-08-31 ; Proportion: 5%

The causal map of the human phenome
Wellcome Trust and Royal Society, Sir Henry Dale Fellowship 2017

- Amount: 1,356,578 GBP ; Role: PI ; Dates: 2018-01-04 to 2023-06-30 ; Proportion: 100%

Classifying mechanisms of pleiotropy to improve causal modelling
BBSRC and GlaxoSmithKline, CASE studentship 2017

- Amount: 100,000 GBP ; Role: PI ; Dates: 2017-10-01 to 2021-09-30 ; Proportion: 5%

Pathways to self-harm: Biological mechanisms and genetic contribution
Medical Research Councal and Medical Research Foundation 2017

- Amount: 372,334 GBP ; Role: Co-I ; Dates: 2017-10-01 to 2019-10-01 ; Proportion: 5%

Identification of Traits and Biomarkers for Prediction of Huntington's Disease Phenotypes using Novel causal analysis Methodologies
Cure Huntington's Disease Initiative 2017

- Amount: 117,059 GBP ; Role: Co-I ; Dates: 2017-04-01 to 2019-03-31 ; Proportion: 10%

Translation of MR for drug target identification; De- tails
GlaxoSmithKline 2017

- Amount: 349,099 GBP ; Role: Co-I ; Dates: 2017-01-01 to 2020-01-01 ; Proportion: 5%

Translation of MR for drug target identification
Biogen 2017

- Amount: 436,165 USD ; Role: Co-I ; Dates: 2017-01-01 to 2020-01-01 ; Proportion: 5%

Dissecting genetic interactions in gene expression
University of Queensland, Early Career Research Grant 2013

- Amount: 34,000 AUD ; Role: PI ; Dates: 2013-01-01 to 2013-12-31 ; Proportion: 2%

Dissecting genetic interactions in complex traits
CASE Studentship, BBSRC and Monsanto 2007

- Amount: 100,000 GBP ; Role: PI ; Dates: 2007-09-01 to 2011-08-30 ; Proportion: 100%

(iv) Indications of external recognition

Editorships

PLoS Computational Biology
Invited guest editor 2017

Appointment to national or international bodies

Early Disease Detection Research Project
Member of chip design committee 2019

- Invitation to a committee for designing the genotyping array for the Early Disease Detection Research Project, which will genotype 5 million UK participants by 2024

Invitations for degree examinations

Jisu Shin University of South Australia
Dissertation of Master of precision medicine 2020

Edward Steere Witswatersrand University, SA
Dissertation of Master of Science in Engineering 2016

Invited lectures (last 3 years)

The Trøndelag Health Study, Norway Invited workshop lead
Multi-omics in longitudinal cohorts 2020

International Agency for Research on Cancer, France Invited talk
Collider bias in Covid-19 research 2020

Elizabeth Blackwell Institute data week keynote Invited talk
New data on Covid-19 is undermined by old statistical problems 2020

GRC Quantitative Genetics conference, Italy Invited session chair
Genetic architecture of complex traits 2019

Mendelian Randomization conference, Bristol — Conference plenary
Genetics of DNA methylation — 2019

SEGEG, University of Oxford — Invited talk
Automating Mendelian randomization — 2018

Edinburgh Alliance in Quantitative Genetics — Invited talk
Machine learning in Mendelian randomization — 2018

University College London — Invited talk
Causal graph of the human phenome] — 2018

(vi) Related administration

Leading group on Covid-19 epidemiology
MRC IEU — 2020 to present

- Initiated project, and recruited group of 12 researchers
- Provided analysis of ZOE symptom tracker app for external collaborators
- Culminated in four publications, numerous presentations including to SAGE and HDRUK

Initiating the OpenGWAS consortium
MRC IEU — 2020 to present

- The OpenGWAS data infrastructure receives 2 million queries per week
- The consortium brings together researchers who develop software for GWAS summary data
- Plans to expand the invitation to international collaborators

Member of the UoB Covid modelling subgroup
University of Bristol — 2020 to present

SEGEG conference organiser
South of England Genetic Epidemiology Group — 2019

- Organised the long-running SEGEG conference to be held in Bristol for the first time

Leading work package on MR method development (programme 1)
MRC IEU — 2018 to present

- Line managing one post doctoral scientist

Leading statistics and informatics theme in epigenetics programme 4
MRC IEU — 2018 to present

Leading Hemani research group
MRC IEU — 2018 to present

- Four post-docs and four PhD students
- Weekly group meetings including pastoral and career support

ALSPAC Board of Directors
ALSPAC — 2017 to present

Leading genetics work package

ALSPAC 2017 to present

- Line managing one post doctoral scientist

Academic leadership and citizenship

(i) Academic leadership in the discipline

I orchestrate the collection and harmonisation of summary data from genome-wide association studies (GWAS) across hundreds of international research groups through the OpenGWAS project. GWAS summary data is vitally important to a wide range of genetic and epidemiological research activities, and the OpenGWAS infrastructure is the largest of its kind comprising approximately 125 billion genetic associations. It has more than 1000 unique users a month and receives more than 2 million queries per week from all over the world.

I design analytical pipelines by unifying and developing statistical methods with researchers from multiple international groups. For example I developed the MR-Base software platform, which has now been cited nearly 1000 times since being published in 2018. I also developed the Genetics of DNA Methylation (GoDMC) analytical pipeline, which was recently diseminated to 36 groups around the world and resulted in a paper recently accepted in Nature Genetics.

I form part of a group coordinating the world's largest genetic data collection effort. The Our Future Health Project will genotype 5 million individuals in the UK within the next 4 years and I am part of the genotype chip design committee which meets on a monthly basis.

(ii) Academic leadership in the University

Medical Anti-Racism Taskforce University of Bristol

Group member 2020 to present

- Leading Special interest group on decolonising the curriculum

UoB Covid-19 Situation Report University of Bristol

Software developer 2020 to present

- Web app developed for internal use that provides daily updates and case mapping
- Used daily by the Incident Management Team in targeting Covid-19 actions

BRMS Equality Diversity Inclusion group University of Bristol

Group member 2018 to present

- Working in career progression subgroup

USS pension calculator University of Bristol

Software developer 2018 to present

- Developed a web-app that allows members of the USS pension scheme to calculate changes to the pension under proposed valuation changes
- Used widely by universities across the country
- http://www.uss-pension-model.com/

Teaching peer review University of Bristol

Reviewer 2018

- Causal inference short course, University of Bristol

(iii) Professional activities outside the University

2019 -	Contributing to the design for the Early Disease Detection Research Programme that will be used to genotype up to 5 million people in the UK
2017 -	Regular peer review for various grant bodies including the MRC, Wellcome Trust, Cancer Research UK
2012 -	Regular contributor to open source software projects (e.g. see https://github.com/explodecomputer/random-metal and https://github.com/explodecomputer/ldsc/)
2011 -	Regular peer review for 20+ academic journals

(iv) Contributions to society

2017 - 2019	Regularly provide private accommodation for asylum seekers and refugees through the *Refugees at Home* charity
2015 -	Contributing member to the charity *Statisticians Without Borders*
2014 -	Provide web and software support to local vegan organisations

(v) Entrepreneurship, enterprise and partnerships

2015 -	I have formal partnerships with GlaxoSmithKlein, Biogen, Pfizer and CHDI that arose through developing the MR-Base platform. This has led to five post-doctoral positions and one PhD position being funded by these organisations, and the development of a standardised contract system to enable future such collaborations to occur.

(vi) Good citizenship

2020	Developed software for the University's Incident Management Team which maps new Covid-19 cases amongst students across the region in real time. This software is used in daily team management meetings
2020	Leading a Special Interest Group on Decolonising the Curriculum, which will bring in a framework for all courses across the medical school for dealing with implicit bias within their learning materials
2020	Member of the Bristol Clear mentoring scheme (currently mentoring two early career researchers)
2016 -	Participated in numerous mock interview panels for research fellowships
2015 -	Extensive software development for the MRC IEU research community, most notably the MR-Base software platform which I created and maintain for others to use. I am pleased that this has grown to be a platform for numerous research papers, and contributed substantially to research grants, many of which I am not named on
2014 -	Voluntary curation and documentation of shared data resources
2014 -	Web and software development for the ALSPAC project (e.g. see http://variables.alspac.bris.ac.uk/ and https://github.com/explodecomputer/alspac)
2014 -	Throughout my time in Bristol I have made it a priority to provide informal training to early career researchers, particularly in genetics, software development and reproducible research.

Future plans

My fellowship is focused on implementing causal inference on a phenome-wide scale, creating a graph of the causal estimates of every phenotype against every other phenotype. I will create collaborations within the University and externally with experts in artificial intelligence, to explore new ways to exploit this graph for biological understanding and medical applications. I will use my causal graph to develop new ways to engage with the public, exploring how perceived ideas of medical interventions would shape future trajectories of population disease burden.

I plan to develop a new professional development course that guides junior researchers and post-graduate students through best practices in code and data management and reproducible digital research.

Through leading the special interest group on decolonising the curriculum, my goal is to develop a method of accreditation for courses across the medical school to introspectively examine potential biases in their teaching materials, identify ways of addressing them, and seek external review of proposed changes. This is a project of crucial importance for equiping the graduates and trainees in redressing biases that continue to incur health inequalities between ethnic groups.