Ethical considerations in AI Adoption in Human Genetics research

Dr Raquel Iniesta
Fair Machine Learning lab
Biostatistics and Health Informatics Department
IoPPN, King's College London

7th February 2015

From *master-servant* to *symbiosis*

- Clinical practitioners and machines have had a master–servant relationship for years: physicians understood the machine's functioning, decided what the machine would do and when.
  - The machine produced outputs that needed further human translation and interpretation.

- The emergence of AI-based medical tools to assist clinical decision-making is leading to a completely new paradigm which resembles a more symbiotic relationship, in which human doctors and computers become teammates aiming to solve a common goal.

  - ❖ diagnosing a patient
  - ❖ anticipating their disease progression
  - ❖ treating them with the most efficient drug

- A computer equipped with an AI system is capable to suggest answers for clinical questions, **even without being operated by a human**.

Relevant reading:
- Trujillo, A.C., Gregory, I.M., Ackerman, K.A.: Evolving Relationship between Humans and Machines, in IFAC – PapersOnLine, Elsevier B.V., 2019, pp. 366–371. https://doi.org/10.1016/j.ifacol.2019.01.015.
- Endsley, M.R.: Toward a Theory of Situation Awareness in Dynamic Systems. Hum. Factors **37**(1), 32–64 (1995). https://doi.org/10.1518/001872095779049543

Why we care about ethics in this context?

- We've been developing models to predict health outcomes for years.

- Why we suddenly care about "ethics"?

- Machine learning/AI used to *analyse data* for academic purposes

Relevant reading:
Iniesta R, Stahl D, McGuffin P. Machine learning, statistical learning and the future of biological research in psychiatry. Psychol Med. 2016 Sep;46(12):2455-65. doi: 10.1017/S0033291716001367

Why we care about ethics in this context?

Jump from *data analysis to **implementation:*** tools to be used in the clinical setting

Why we care about ethics in this context?

- **Jump** from *data analysis to **implementation:*** tools to be used in the clinical setting

- ***Fears:***
  - *Doctors**: will I be losing my job? Do these systems work?***
  - *Patients**: can a machine become a doctor? Are these methods working well?***
  - *Developers**: Is my algorithm safe?***

Fears translate into ethical dilemmas:

- *Fears from Doctors: **will I be losing my job? Do these systems work?***
  - **Human replacement**
  - **Bias, transparency, discrimination, security, privacy**

- *Fears from Patients: **can a machine become a doctor? Are these methods working well?***
  - **Patient rights to be treated as a person**
  - **Bias, transparency, discrimination, security, privacy**

- *Fears from Developers: **Is my algorithm safe?***
  - **Bias, transparency, discrimination, security, privacy**

Ethical dilemmas

- These are considered the reason for the **implementation gap**: less AI systems are implemented in healthcare than hoped

- Together with human replacement, diverse political entities and literature identified numerous ethical dilemmas

Relevant reading:
- Aristidou, A., Jena, R., Topol, E.J.: Bridging the chasm between AI and clinical implementation. Lancet **399**(10325), 620 (2022). https://doi.org/10.1016/S0140-6736(22)00235

Policy bodies regulation: EU AI act

EU AI act: the world's first comprehensive AI law developed by the European Parliament.

- Parliament's priority is to make sure that AI systems used in the EU are *safe*, *transparent*, *traceable*, *non-discriminatory* and *environmentally* friendly.

- The new rules establish obligations for providers and users depending on the level of risk:

Unacceptable risk	High risk	Limited risk
Ai systems considered a **threat to people** and **will be banned**. They include: • **Cognitive behavioural manipulation** of people or specific vulnerable groups • **Social scoring** of people based on behaviour, socio-economic status or personal characteristics • **Biometric identification** and categorisation of people	AI systems that **negatively affect safety or fundamental rights**. **Will be assessed** before on the market and throughout their lifecycle: • AI products under the EU's product safety legislation. Toys, aviation, cars, medical devices and lifts. • AI systems into specific areas that will have to be registered in an EU database	• Comply with **minimal transparency requirements** that would allow users to make informed decisions. • Users should be **made aware when they are interacting with AI**. Including AI systems that generate or manipulate image, audio or video content, for example deepfakes.

Policy bodies regulation: EU AI act

Generative AI, like ChatGPT, would have to comply with transparency requirements:

- Disclosing that the content was generated by AI

- Designing the model to prevent it from generating illegal content

- Publishing summaries of copyrighted data used for training

High-impact general-purpose AI models that might pose systemic risk, such as the more advanced **AI model GPT-4**, would have to undergo **thorough evaluations** and any serious incidents would have to be reported to the European Commission.

What are the ethical challenges of AI in healthcare?

❖ Principles of Biomedical Ethics by Beauchamp and Childress (1979)

❖ Systematic review (Li et al. 2023) of 45 papers on ethical principles and issues when applying AI in healthcare.

❖ The World Health Organisation 2021 guidance on Ethics & Governance of Artificial Intelligence for Health.

We can look for meaningful convergence between them, assuming **overlapping consensus**

Relevant reading:
- Beauchamp, T.L., Childress, J.F.: Principies of Biomedical Ethics. Oxford Univeristy Press, Oxford (1979)
- Li, F., Ruijs, N., Lu, Y.: "Ethics & AI: a systematic review on ethical concerns and related strategies for designing with AI in healthcare. AI **4**(1), 28−53 (2023). https://doi.org/10.3390/ai4010003
- World Health Organisation: *Ethics and governance of Artificial Intelligence for health*. 2021. [Online]. http://apps.who.int/bookorders.
- Iniesta, R. The human role to guarantee an ethical AI in healthcare: a five-facts approach. *AI Ethics* (2023). https://doi.org/10.1007/s43681-023-00353-x

(1) *Four ethical pillars*	(2) *Li et al. 2023*	(3) *WHO guidance*
Justice	**Justice** and **fairness**	Ensure inclusiveness and **equity**
Respect for **autonomy**	Freedom and **autonomy**	Protect **autonomy**
	Privacy	
	Transparency	Ensure **transparency**, explainability, and intelligibility
Non–maleficence	**Patient safety** and cybersecurity	Promote human well–being, human **safety,** and the public interest
	Trust	
Beneficence	**Beneficence**	
	Responsibility	Foster **responsibility** and accountability
	Solidarity	
	Sustainability	Promote artificial intelligence that is responsive and **sustainable**
	Dignity	
	Conflicts	

Principles of Biomedical Ethics

The four ethical pillars have been classically used to analyse ethical situations in medicine.

JUSTICE

MEDICAL BENEFITS SHOULD BE DISTRIBUTED FAIRLY

RESPECT FOR AUTONOMY

PATIENT AUTONOMY SHOULD BE MAXIMISED IN INFORMED MEDICAL DECISIONS.

NON-MALEFICENCE

THERE IS AN OBLIGATION NOT TO INFLICT HARM ON OTHERS

BENEFICENCE

AGENTS INVOLVED ON PATIENTS' HEALTH CARE SHOULD ACT IN A PATIENT'S BEST INTERESTS

Relevant reading:
- Beauchamp, T.L., Childress, J.F.: Principles of Biomedical Ethics. Oxford University Press, Oxford (1979)

Principles of Biomedical Ethics: naturally apply **for AI in healthcare**

JUSTICE

MEDICAL BENEFITS SHOULD BE DISTRIBUTED FAIRLY:

ALGORITHMIC FAIRNESS

RESPECT FOR AUTONOMY

PATIENT AUTONOMY SHOULD BE MAXIMISED IN INFORMED MEDICAL DECISIONS.

EMPOWERMENT OF PATIENTS ON AI-MEDICAL DECISIONS

NON-MALEFICENCE

THERE IS AN OBLIGATION NOT TO INFLICT HARM ON OTHERS

PATIENTS' SAFETY

BENEFICENCE

AGENTS INVOLVED ON PATIENTS' HEALTH CARE SHOULD ACT IN A PATIENT'S BEST INTERESTS

ALGORITHMIC BIAS AND TRANSPARENCY

Algorithmic Fairness

A **fair Machine Learning algorithm** should make equal mistakes for individuals with different realisations of a **protected attribute**.

The nine **protected characteristics** in the UK are:

- age
- sex
- gender reassignment
- marriage and civil partnership
- pregnancy and maternity
- Disability
- Race
- religion or belief
- sexual orientation

In the UK you are protected under the **Equality Act 2010** from these types of discrimination.

Algorithmic Fairness

Aim: Avoid AI models that disadvantage individuals based on protected attributes. *Maximise model fairness.*

There are no clear policies or guidelines to explain the use of protected attributes in building fair models

➢ **General rule**: A **protected attribute** is a feature that may not be used as the basis for decisions.

Dilemma between *privacy* and *fairness*:

Protected attributes are mostly private information and, therefore, unavailable, which challenges the use of protected attributes for fair purposes.

Relevant Reading:
- Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. A survey on bias and fairness in machine learning. ACM Computing Surveys (CSUR), 54(6):1–35, 2021.
- Faisal Kamiran and Toon Calders. Data preprocessing techniques for classification without discrimination. Knowledge and information systems, 33(1):1–33, 2012.

Algorithmic Bias and discrimination

- It relates to outcomes which are systematically **less favorable** to individuals within a particular group and where there is no relevant difference between groups that justifies such harms.

- AI that makes decisions that are systematically **unfair** to certain groups of people have real potential to cause harm.

  - *A widely used ML algorithm assigned the same level of disease risk to Black and White patients, even if Black patients were sicker than White patients. Consequently, the system was actively causing harm to Black patients by leading to allocation of fewer resources to them (breaching the non-maleficence principle)*

Avoiding/identifying bias

✓ Enough sample size

✓ Appropriate algorithm selection (ex. trees higher risk of bias than random Forests)

✓ **Diversiy**:

- The important variables for our investigation are in
- No groups are underrepresented
- **Protected features** are identified and analysed accordingly
- Enough observations for the relevant variable categories / ranges are included: *ethnicities, gender*

✓ Ideally, **two independent datasets**: Training and Testing

✓ Care in **data-preprocessing**: missing data imputation can cause bias

✓ Measuring models' accuracy separately for different protected categories to check whether the model is biased against some demographic categories.

✓ Encouraging ML techniques towards unbiased results

Algorithms replicate *what they saw in data*. Data should include the relevant information!

Transparency and explainability

- Opposite to non-transparent or "black-box", in **transparent** algorithms the patterns the algorithm follows to derive an output are **not opaque** to the developers, clinicians and patients. The rationale to derive an outcome can be understood.

- Non-transparent AI is in danger of breaching the Respect for Autonomy: risk of disempowerment of both patients and clinicians.

- Developers should consider **Explainable AI:** allows humans understand the reasoning behind AI decisions even if it's a black-box, increasing intelligibility.

- Using techniques to explain models: ex. **Shapley plots**

Freedom and autonomy: disempowerment of clinicians

i. Clinicians' should be *free to develop their clinical judgement*.
   - Novice clinicians too dependent on AI recommendations, at risk of not growing their own clinical judgement: **disempowerment of clinicians**.

ii. The ethical principle says that doctors *should make use of all their available knowledge and skills to make a clinical decision*.

   - **Explicit knowledge**: Knowledge that can be codified and written. Skills that can be transferred. AI knowledge is included.
   - **Tacit Knowledge**: Kind of knowledge that we *have* but that we do not know *how to explain.* It is a form of unconscious expertise enhanced by practice. Informed medical intuition <u>very often used by expert clinicians</u>.

Clinician's knowledge should **not be replaced but complemented** by automatically deployed AI recommendations.

Freedom and autonomy: Empowerment of patients

- The Respect for autonomy of patients should be warranted.

- A patient **subjective experience** should be incorporated in any clinical decision if this is to serve the patient, regardless the AI. Space to elaborate suffering. AI can **hardly incorporate** subjectivity.

- Patients Public Involvement (PPI) Developers and Clinicians should generate activities, patients should enrole: education in AI, **empowerment patients** to take informed decisions

- Patients should be informed that they have the **right of not being subject to an AI decision** (PGDR article 22)

Safety and privacy

- University Departments, Healthtech companies, should implement and use **protocols that** should guarantee users' **identity,** *data security and privacy,* and the build of <u>transparent</u> model development that can produce <u>fair</u> and <u>non-discriminant</u> outputs in search of the *Beneficence and Justice* ethical principles.

- EU law: "**patients' perspective** on data sharing, consent and data privacy should be taken into account in healthcare & research

(1) *Four ethical pillars*	(2) *Li et al. 2023*	(3) *WHO guidance*
Justice	**Justice** and **fairness**	Ensure inclusiveness and **equity**
Respect for **autonomy**	Freedom and **autonomy**	Protect **autonomy**
	Privacy	
	Transparency	Ensure **transparency**, explainability, and intelligibility
Non–maleficence	**Patient safety** and cybersecurity	Promote human well–being, human **safety,** and the public interest
	Trust	
Beneficence	**Beneficence**	
	Responsibility	Foster **responsibility** and accountability
	Solidarity	
	Sustainability	Promote artificial intelligence that is responsive and **sustainable**
	Dignity	
	Conflicts	

Under the collaborative paradigm…

Responsibility on AI-assisted clinical decisions is **shared and distributed** amongst humans:

✓Clinicians: responsible of clinical decisions

✓Patients: respect for autonomy and shared decisions

✓Developers: reduce bias, increase fairness, intelligibility

Dignity: Ethical AI should avoid dehumanisation of patients

AI follow the **biomedical** model: patients are "mechanical systems made up of interacting parts".

Risks *objectification & mechanization*, risk of **dehumanisation in medicine**

The **subjective patient experience** is key to achieve good treatment response. AI can **hardly** incorporate subjectivity.

Chatbots or chatGPT should not be constantly **replacing humans**. Only humans have *consciousness*, *empathy*, know the *context environment*.

Global moral imperative of the medical profession "*each patient must be treated as a person*"

The **human warranty** is crucial. "human in the loop" , which implies evaluation by patients and clinicians in the development and deployment of AI technologies.

Five facts for the human role that guarantees an ethical AI in healthcare

1   The four classical ethical pillars of the medical profession are valid for assessing AI ethical risks in healthcare.

2   AI technologies are a complement and not a replacement of clinician's knowledge.

3   Clinicians are accountable for their clinical decisions and their decisions are to be respected, regardless the assistance of an AI system.

4   The empowerment and education of patients is necessary for an ethical AI in healthcare.

5   Developers are accountable for the automated decisions provided by the tools they develop. Their awareness and education on the ethical concerns can ensure a better alignment between algorithms and values.

Relevant Reading:

Iniesta, R. The human role to guarantee an ethical AI in healthcare: a five-facts approach. *AI Ethics* (2023). https://doi.org/10.1007/s43681-023-00353-x

Facilitating collaboration:
the Participatory Modelling (PM)

"... an approach for including a broad group of stakeholders in the process of formal decision analysis." (Stave, 2010)

PM is a purposeful learning process for action that engages the **implicit and explicit knowledge of stakeholders** to create formalized and shared representations of reality.
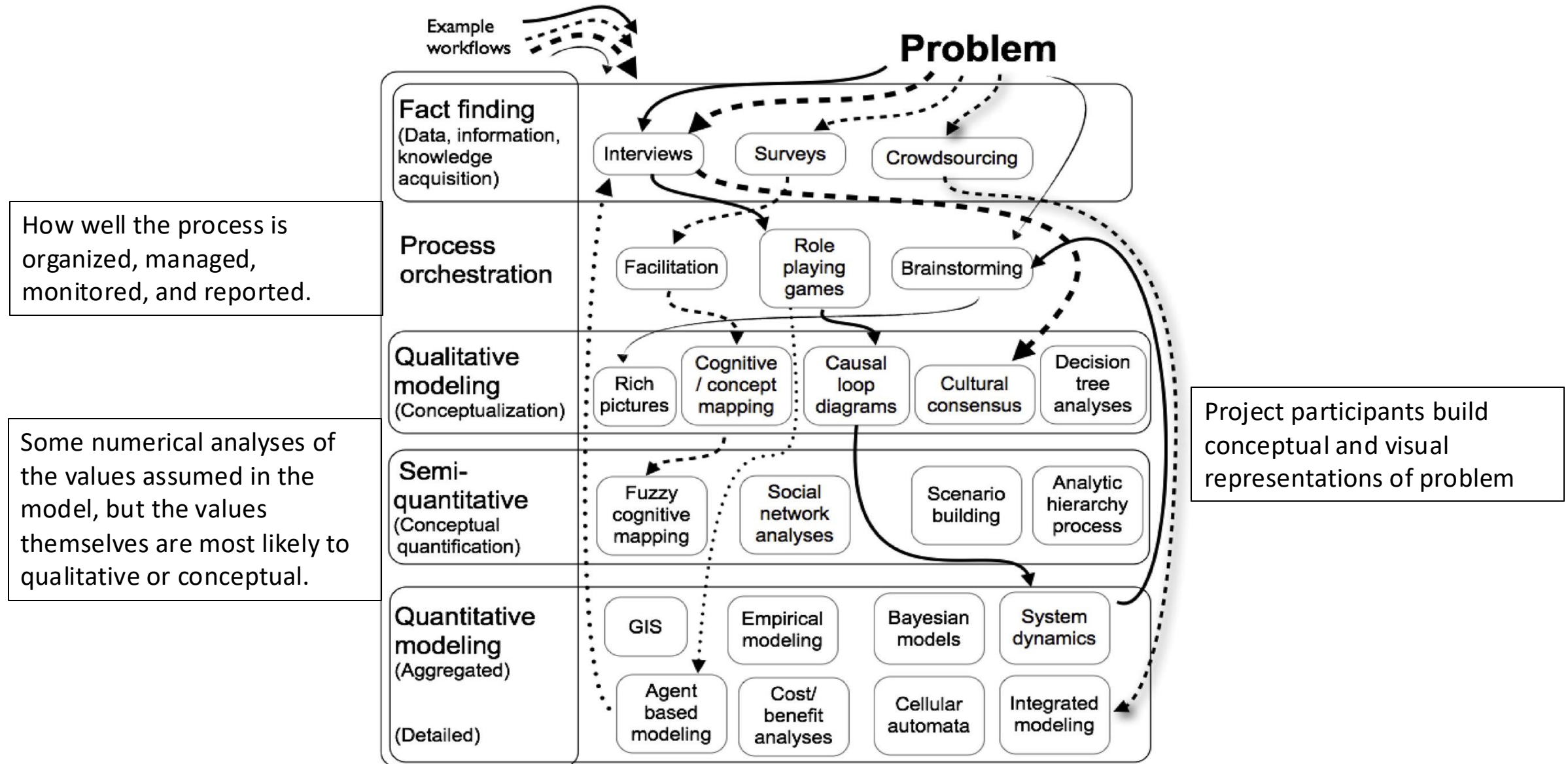
"engaging non-scientists in the scientific process"

Stakeholders: modelers, researchers, patients

Participatory modelling:

How well the process is organized, managed, monitored, and reported.

Some numerical analyses of the values assumed in the model, but the values themselves are most likely to qualitative or conceptual.

Project participants build conceptual and visual representations of problem

Example workflows

Problem

Fact finding
(Data, information, knowledge acquisition)

Interviews   Surveys   Crowdsourcing

Process orchestration

Facilitation   Role playing games   Brainstorming

Qualitative modeling
(Conceptualization)

Rich pictures   Cognitive / concept mapping   Causal loop diagrams   Cultural consensus   Decision tree analyses

Semi-quantitative
(Conceptual quantification)

Fuzzy cognitive mapping   Social network analyses   Scenario building   Analytic hierarchy process

Quantitative modeling
(Aggregated)

GIS   Empirical modeling   Bayesian models   System dynamics

(Detailed)

Agent based modeling   Cost/benefit analyses   Cellular automata   Integrated modeling

Thank you!

Raquel Iniesta

raquel.Iniesta@kcl.ac.uk

Fair Machine Learning Lab

Office C1-43 , SGDP Center

Biostatistics and Health Informatics Department

Institute of Psychiatry, Psychology and Neurosciences

Denmark Hill, London