Data science in the medical curriculum:

Equipping medical students for the digital age

HDRUK
Health Data Research UK

msc Medical Schools Council
For the future of medicine

Contents

Introduction	03
Proposed learning competencies	06
Resources	09
Glossary	10
Appendix 1 - MSC & HDR UK survey results	13

Introduction

As the healthcare landscape continues to evolve, integrating data science and technology into medical research and practice is becoming increasingly essential. Advances in data analytics, machine learning, and artificial intelligence have the potential to revolutionise patient care, biomedical research, and healthcare systems management. As such, it is imperative that medical education evolves to equip future doctors with the knowledge and skills necessary to navigate and harness the power of data science effectively.

The exponential growth in healthcare data—ranging from electronic health records and genomic data to wearable technology outputs and large-scale epidemiological datasets—demands a new set of skills. Medical professionals must become proficient in critical interpretation of and leveraging insights from complex data to make informed decisions. However, a 2023 survey of medical school staff and students conducted by the Medical Schools Council (MSC) and Health Data Research UK (HDR UK) revealed widespread recognition of the importance of data science in medicine and the need for improved teaching provisions in this area.

To address this issue, MSC convened a working group with expertise in data science and medical education, to ascertain what data science competencies medical schools should strive to support their medical students to achieve on their curricula. The focus of the group was on deciding what core knowledge all medical graduates should possess, and all medical schools could be capable of delivering, therefore technical skills such as programming were not considered for inclusion. These skills could be acquired by students during special Study Modules or through widely accessible specialist training providers.

Methods

MSC convened a working group recruited from members of Medical Schools Council with expertise in curriculum development, data science and digital health. Additional expertise was incorporated through involvement of educational leads from Health Data Research UK. Representatives from two medical schools that have successfully introduced Data Science modules in their curricula were invited in addition to a leader of a global consensus initiative that has defined recommended digital health competencies for medical students. The group held regular online meetings to agree the scope of the competencies and their feasibility within curriculum constraints. Following discussions of remit, the group participated in a formal consensus voting exercise whereby a threshold of 60% voting for a core competency would result in its inclusion.

This paper outlines the essential learning outcomes and competencies that the group agreed met these criteria. The recommendations are adapted from a broader set of digital health competencies defined through a global consensus exercise and refined to meet the needs of all UK medical graduates[1].

Conclusion

The recommended learning outcomes and competencies provide a structured framework for supporting the integration of data science into medical education. The proposed competencies span several key areas, including health informatics, artificial intelligence, precision medicine, data governance and the ethical, professionalism, legal, and regulatory considerations in digital health. Some of the competencies will already be covered by modules and programmes taught at medical schools, such as those relating to professionalism and ethical standards – these could be expanded to include digital health. This paper also contains a glossary and links to existing freely available learning materials and resources to accompany the recommendations to support schools in its implementation.

During discussions, it was recognised that the emergence of new technologies, such as large language models and AI co-pilots, would most probably radically alter the current landscape of data science training. It was determined that the current recommendations should remain agnostic to these trends until as such time that some form of stability emerges and scalable learning recommendations, particularly to technical skills in analysis and programming, could be developed.

By adopting these recommended learning outcomes and competencies, medical schools can ensure that their graduates are well-versed in the innovative technologies that are shaping the future of healthcare. By embedding data science into medical training, a new generation of doctors who are adept at using data to enhance patient outcomes, improve healthcare delivery, and contribute to the advancement of medical knowledge and more efficient healthcare systems will be fostered.

1 Car J, Ong QC, Erlikh Fox T, et al. Digital health competencies in medical education (DECODE): An international consensus statement based on a Delphi study. JAMA Netw Open. 2025.

Table of group membership

Professor Amir Sam	Imperial College London
Professor Christopher Yau	Health Data Research UK / University of Oxford
Clare Owen	Medical Schools Council
Dr Dimitrios Doudesis	University of Edinburgh
Professor Eleanor Hothersall-Davies	University of Dundee
Professor Jo Knight	Lancaster University
Dr Katie Petty-Saphon	Medical Schools Council
Professor Amitava Banerjee	University College London
Dr Nicola Rennie	University of Lancaster
Professor Sinéad Langan (Chair)	London School of Hygiene and Tropical Medicine
Dr Tom Levett	Brighton & Sussex Medical School
Dr Venkat Thiagarajan	Queen Mary University of London
Professor Josip Car	King's College London

This report has been authored by Dr Chris Fassnidge, Senior Policy Officer at the Medical Schools Council.

Proposed learning competencies

The Medical Schools Council recommends that at the point of graduating from medical schools, doctors should be able to adequately perform the competencies listed below. Medical schools may wish to adapt this list to suit their own requirements.

Data governance and data management

- Describe what constitutes health data.
- Explain the importance of data governance policies in healthcare.
- Demonstrate an understanding of data required for specific functions of electronic health records (e.g., laboratory results, clinical decision support), its location, accessibility, and record manipulation tracking.

Health informatics

- Explain the concept of health informatics.
- Describe the role of health data and informatics in supporting health through enhanced data collection, management, sharing, and application in the context of both infectious and chronic diseases.
- Interpret and explain findings generated from health data to patients and the public.

Professionalism, ethical, legal, and regulatory considerations in digital health

- Apply best practices for managing digital patient data.
- Describe protected health information.
- Apply prevailing privacy and security rules when handling protected health information.
- Demonstrate compliance with ethical conduct and codes of practice when processing digital patient data.
- Explain how professional, clinical, and research ethics are applied and practiced within digital health space.

Digital identity, safety, and security

- Describe common behaviours of doctors which may compromise data security.
- Demonstrate awareness of prevailing rules or regulations on sharing protected health information via cross-platform instant messaging services (e.g., WhatsApp, Telegram).

- Explain the concepts of medical professional, digital identity and digital intelligence.

Artificial intelligence in healthcare

- Describe basic principles of artificial intelligence (AI), natural language processing (NLP), speech recognition, machine learning (ML), automated image interpretation and predictive analytics.
- Critique limitations of and barriers to using AI in healthcare.
- Explain the role of big data in healthcare data science and delivery.
- Explain the importance of rigorous real-world clinical validation of AI-based technologies before implementation in patient care.
- Maintain vigilance and validate advice given by machines to avoid automation bias.
- Describe application of automated image interpretation and pattern recognition in radiology, radiotherapy, and retinal photography.
- Discuss how integrating big data from different sources (e.g., data in electronic health records, hospital records, medical records of patients, results of medical examinations; data from devices that are a part of internet of medical things and wearables, data from biomedical research) can yield useful insights into healthcare provision, utilization, optimization, and improvement.

Clinical Academic Research

- Describe how advanced data analysis can help identify patterns, trends, risk factors, and correlations within health data, which can lead to new research questions and hypotheses.
- Discuss how data from everyday clinical practice can be used to assess treatment effectiveness, monitor drug safety, and drive advancements in healthcare.
- Discuss how data science can be used to develop models that predict outcomes, such as disease progression, treatment effectiveness, or patient risk.

Digital diagnostics algorithms

- Describe how digital diagnostic algorithms can enhance, guide, and record aspects of clinical examination.
- Recognize the limitations of fully automated AI-led diagnostic platforms and traditional methods.

Precision medicine

- Describe the concept of precision medicine and its implications in healthcare.
- Discuss how data collected through digital health technologies (e.g., wearable sensors, mobile health apps) can support precision medicine.

Digital health literacy

- Describe digital health literacy and its impact on health outcomes and access to care.
- Describe how higher health literacy leads to reduced healthcare costs, hospitalisation rates, readmissions and emergency department visits.

Health inequalities

- Explain the need for datasets to be diverse and representative of patient populations, including those from marginalised groups.
- Identify and mitigate potential biases in data and algorithms that can perpetuate existing health disparities.
- Understand the digital divide and its impact on access to healthcare services and information.
- Recognise the heightened risk of data breaches and misuse for vulnerable populations.

Personal health records

- Explain both the positive and negative impacts of personal health records on patient-centred healthcare (e.g., patient-provider communication, education and lifestyle changes, health self-management)
- Explain what constitute appropriate use of and access to patient health records.

Foundation and principles of health information systems

- Discuss both positive and negative impacts of electronic health records on patient care.

More details on the context and application of these core competencies are available in the original detailed manuscript.

Resources

Introduction to Data Science for Healthcare Professionals

A six-week course from King's College London, exploring data science concepts, focusing on how data is collected, analysed, and used to inform professional practice in the healthcare sector. (Free course from King's College London)

www.futurelearn.com/courses/introduction-to-data-science-for-healthcare-professionals

Digital Health - Design, Development, Evaluation

Introduce learners to the role and application of digital health technologies in public health and healthcare globally and the design, development, implementation and evaluation of digital health interventions. (Free course from Imperial College London)

www.coursera.org/specializations/digitalhealth

The Power of Data in Health and Social Care

On this course, you'll discover the principles of healthcare data analytics. Discover best practices around data governance, tools to help you manage and analyse data, and how to work collaboratively with data scientists and digital teams. (Free course from The University of Strathclyde)

www.futurelearn.com/courses/data-for-health-and-care

Protecting Health Data in the Modern Age: Getting to Grips with the GDPR

This course will explore the protection of health data in light of the GDPR. You will learn about rights, obligations, risks, safeguards and many other related aspects. (Free course from The University of Groningen)

www.futurelearn.com/courses/protecting-health-data

Introduction to Artificial Intelligence for Clinical Researchers

The NIHR CRN recognises the increasing importance of AI-focused clinical research and would like to build awareness and basic knowledge of AI among affiliated researchers. This self-paced training course is aimed towards colleagues involved in different stages of clinical research that can benefit from AI methods. Specific groups include research nurses, research managers, full-time researchers, and clinicians. (Free from NIHR Learn)

learn.nihr.ac.uk/

Glossary

Artificial Intelligence (AI)

The capacity of computers or other machines to exhibit or simulate intelligent behaviour; the field of study concerned with this. Including software used to perform tasks or produce output previously thought to require human intelligence, esp. by using machine learning to extrapolate from large collections of data.

Automated Image Interpretation

In the context of medicine, automated image interpretation involves the use of computer algorithms and artificial intelligence techniques to analyse medical images, such as X-rays, CT scans, MRIs, and ultrasounds. These algorithms are trained to learn patterns and features that are indicative of different medical conditions.

Automation Bias

The tendency to use automated cues (such as Clinical Decision Support alerts) as a heuristic replacement for vigilant information seeking and processing.

Big Data

Data of a very large size, typically to the extent that its manipulation and management present significant logistical challenges; (also) the branch of computing involving such data.

Data Science

An interdisciplinary field that integrates statistical, computational, and artificial intelligence techniques to analyse data, uncover patterns, and extract meaningful insights for informed decision-making.

Digital Diagnostics

The use of digital technologies, including artificial intelligence, wearable sensors, and advanced data analysis, to diagnose, monitor, and manage health conditions.

Digital Health Literacy

The ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem.

Digital Identity

The electronic representation of an individual, organisation, application, or device, enabling secure and authenticated access to digital services. It includes credentials like personal details, biometric data, and secure login methods, verifying identity in online and offline interactions. Acting as a unique digital fingerprint, it ensures privacy, security, and integrity across various domains, from healthcare to finance.

Digital Intelligence

Digital Intelligence in medicine refers to the use of advanced technologies to analyse and interpret health-related data. This intelligence supports informed decision-making, predicts outcomes, and optimises treatments by uncovering patterns and insights not readily apparent. It can be utilised by individuals, clinicians, organisations, or applications to enhance healthcare delivery.

Health Informatics

The practice of gathering, studying, and utilising data to improve healthcare. Hospital informatics merges technology and medicine by implementing digital information technology and artificial intelligence (AI) to enhance the quality of healthcare services.

Internet Of Things

A proposed development of the internet in which many everyday objects are embedded with microchips giving them network connectivity, allowing them to send and receive data.

Machine Learning

The capacity of computers to learn and adapt without following explicit instructions, by using algorithms and statistical models to analyse and infer from patterns in data; the field of artificial intelligence concerned with this.

Natural Language Processing

A form of computational linguistics in which natural-language texts are processed by computer (for automatic machine translation, literary text analysis, etc.)

Pattern Recognition

The process by which a computer, the brain, etc., detects and identifies ordered structures in data or in visual images or other sensory stimuli.

Personal Health Records

Information about a patients health and care that is stored online, is available to and managed by the patient, with access available by health professionals.

Precision Medicine

A form of medicine that considers variability in genes, environment, and lifestyle to decide disease treatment and prevention, more accurately predicting which treatment and prevention strategies will work best for a particular patient.

Predictive Analytics

The processing and evaluating of historic and real-time information to create forecasts, predictions and recommendations about individual patient care or wider public health.

Appendix 1: MSC & HDR UK survey results

The Medical Schools Council would like to acknowledge the work of Adenrele Adekoya, who conducted and analysed this survey on behalf of HDR UK and MSC.

A total of 175 students responded to the survey across 19 medical schools in the United Kingdom from which 173 responses were used in analysis **(Figure 1a)**. Students were drawn from across various stages of academic progression **(Figure 1b)**. The survey examined students existing awareness of data science and artificial intelligence including their ability to recall any data science resources **(Figure 1c)** and uses of artificial intelligence **(Figure 1d)**. The survey found that most students from across all years of study did not recognise any teaching as being specifically designated as "data science" **(Figure 2a)** but respondents recognised the importance of data science in medicine **(Figure 2b)**. Students did not feel that the provision of data science training was sufficiently adequate in their courses particularly by students in the later years of their studies **(Figure 2c)**.

A total of 51 medical school staff were also surveyed from 21 different schools. Responses correlated with the student-based perception that the medical courses contained little to no data science training **(Figure 3a)**. Of the eight medical schools which offered data science training, four did not go beyond the teaching of statistics and consider wider areas of data science **(Figure 3b)**. Less than half of respondents suggested that their courses would increase the level of data science training though the majority of staff respondents recognised the future importance of the topic **(Figure 3c)**.

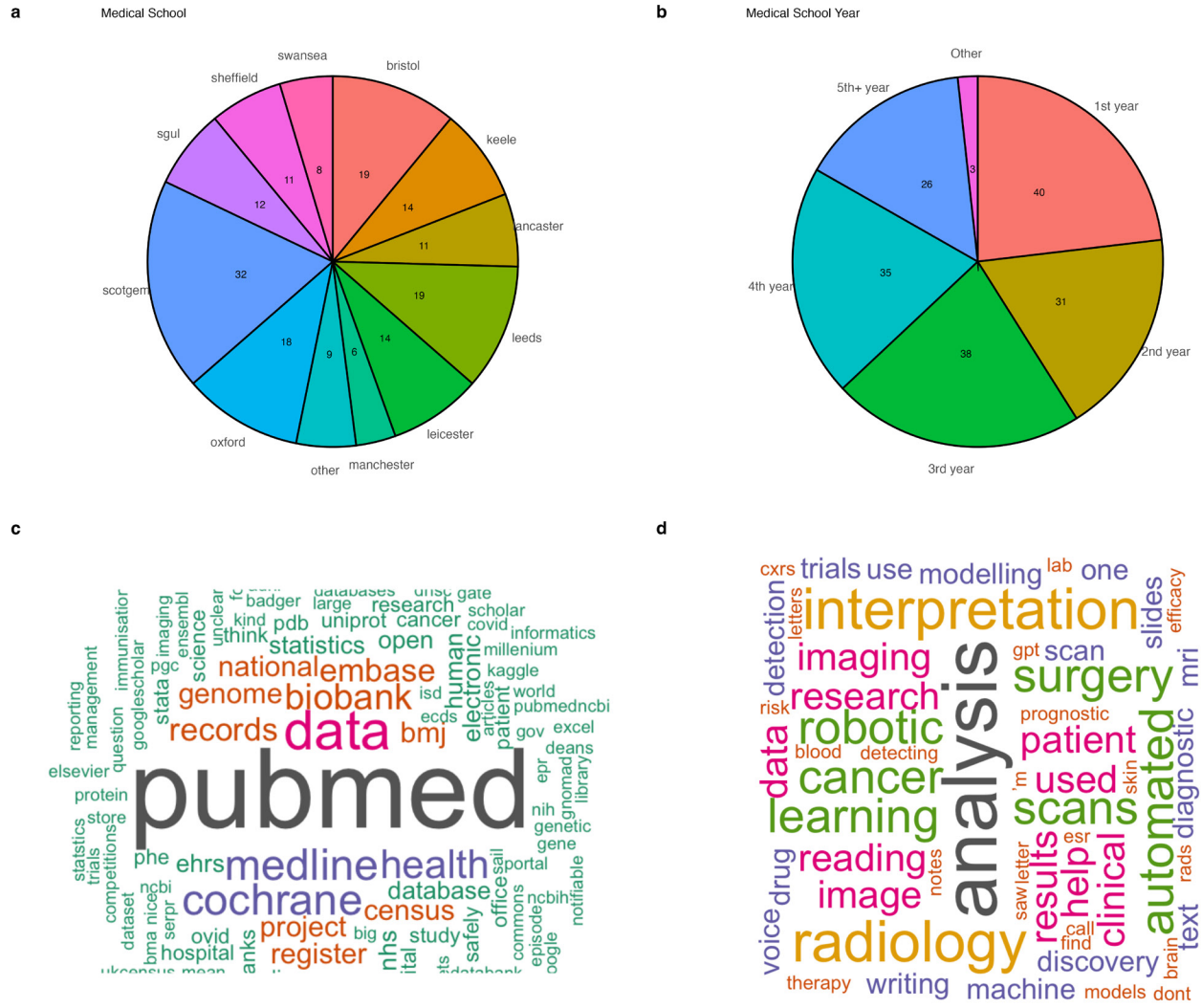

Figure 1. Overview of student survey respondent cohort.

The distribution of respondents across (a) UK medical schools and (b) year of study. Words that students associated with (c) data science resources and (d) uses of artificial intelligence in medicine.

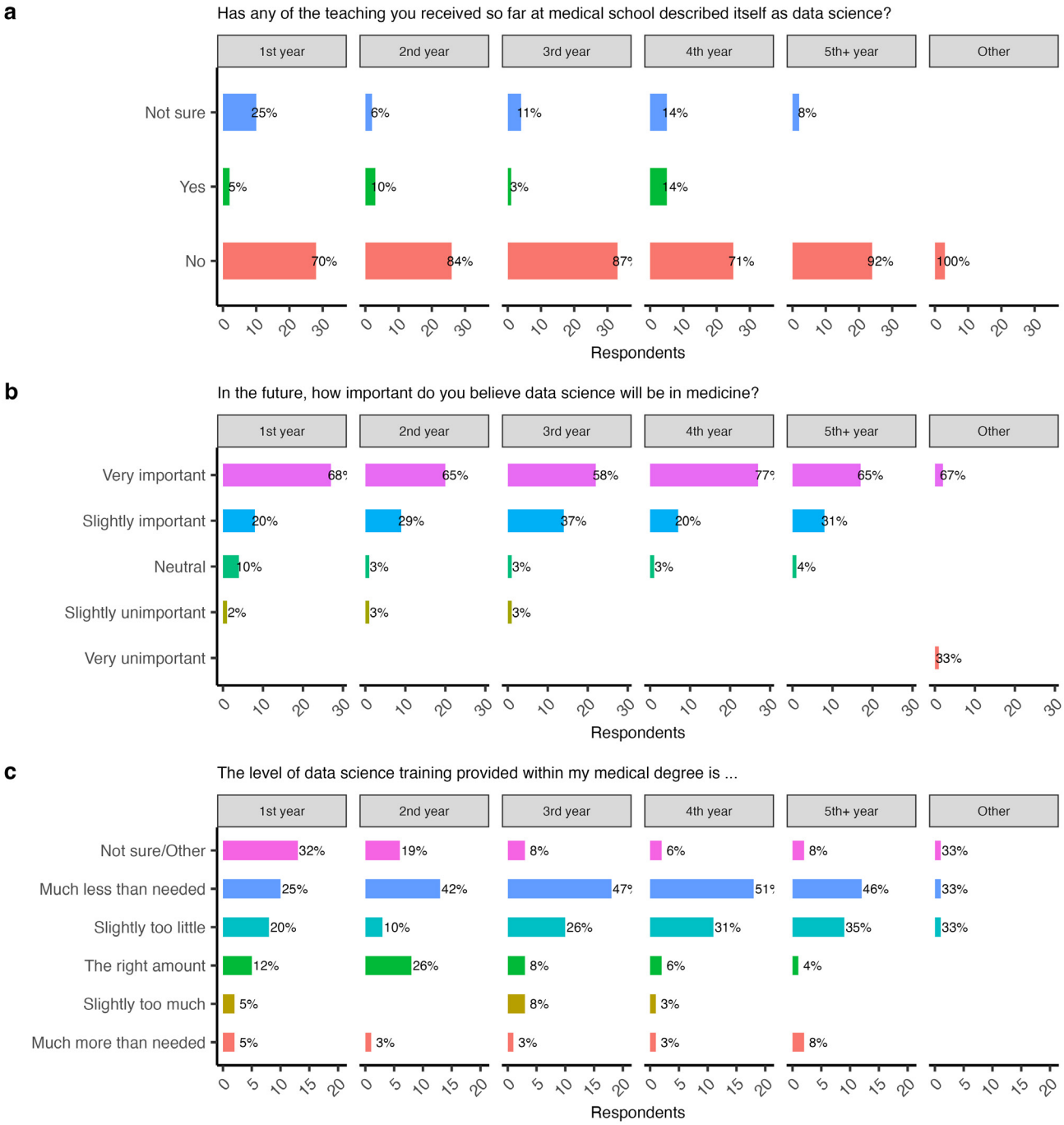

Figure 2. Overview of student survey findings.

Analysis was stratified by the year of study at the point of response and showed responses to questions on (a) the amount of teaching described as data science received during their studies, (b) the future importance of data science in medicine and (c) the required level of data science training for medics.

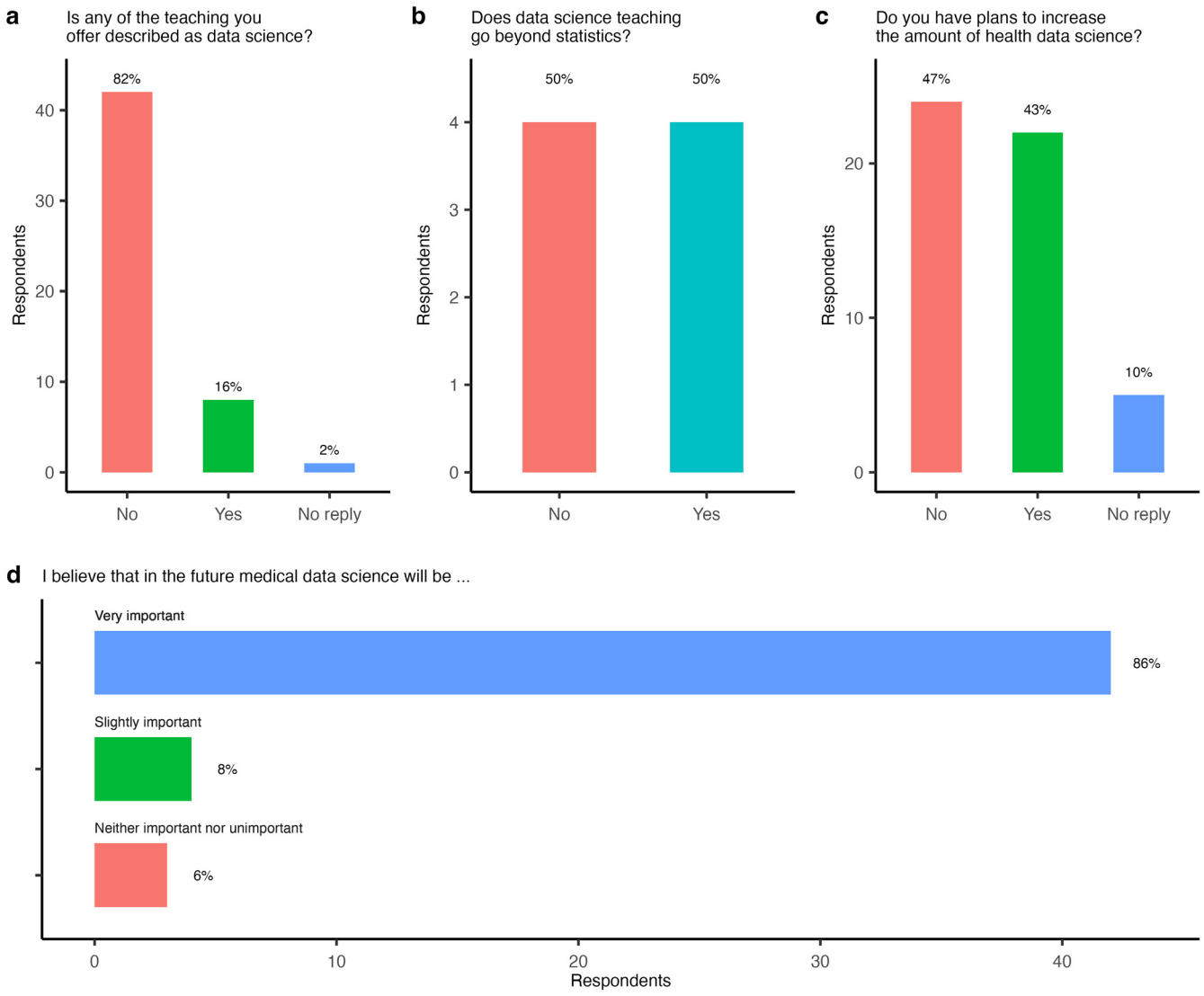

Figure 3. Overview of medical school staff survey

Staff respondents were asked to specify if (a) any teaching could be described as data science and (b) how much goes beyond statistics, (c) plans to increase the level of data science provision within courses and (d) the future importance of medical data science.

msc Medical Schools Council

For the future of medicine

The Medical Schools Council is the representative body for UK medical schools.

Find out more:
www.medschools.ac.uk