Data Science in Undergraduate Medicine: Course Overview and Student Perspectives

Dimitrios Doudesis[1,2] and Areti Manataki[3,*]

[1] *BHF Centre for Cardiovascular Science, University of Edinburgh, Edinburgh, UK*
[2] *Usher Institute, University of Edinburgh, Edinburgh, UK*
D.Doudesis@ed.ac.uk
[3] *School of Computer Science, University of St Andrews, St Andrews, UK*
A.Manataki@st-andrews.ac.uk
** Corresponding author*

Keywords: data science; health; medicine; education; training; health informatics

Structured Abstract:

Background:

Despite the growing interest in health data science education, it is not embedded in undergraduate medical curricula and little is known about best teaching practices. This paper presents a highly innovative course in a UK university that introduces undergraduate medical students to data science. It also discusses a study on student perspectives on the learning and teaching of health data science.

Methods:

The pedagogical design elements of the Data Science in Medicine course are discussed, along with its syllabus, assessment methodology and flipped classroom delivery. The course has been offered to approximately 630 students over three years. Student perspectives were investigated through three focus groups with the participation of 19 students across different study years in medicine. An experiment was conducted regarding instructor-led vs. video-based modalities of online programming labs, with the participation of 8 students.

Results:

The course has led to improved data competency among medical students and to a positive change in their opinions about data science. Motivating the course and showing relevance to clinical practice was one of the biggest challenges. Statistics was perceived by focus group participants as an essential data skill. Including data science in the medical curriculum was perceived as important by Year 1 students, while opinions varied between Year 4/5 participants. Video-based online labs were preferred over instructor-led online labs, and

This is the accepted manuscript of an article that has been published in the International Journal of Medical Informatics. Changes were made to this version by the publisher prior to publication.
The final published version of this paper is available online at https://doi.org/10.1016/j.ijmedinf.2021.104668
Citation for published version: Doudesis, D., & Manataki, A. (2022). Data Science in Undergraduate Medicine: Course Overview and Student Perspectives. International Journal of Medical Informatics, 159, 104668.

they were found to be more useful and enjoyable, without leading to any significant difference in academic performance.

Conclusions:
Teaching data science to undergraduate medicine students is highly desirable and feasible. We recommend including statistics in the curriculum and practical skill development through simple and clinically-relevant data science tasks, supported through video-based online labs. Further reporting on similar courses is needed, as well as larger-scale studies on student perspectives.

1 Introduction

The medical world is undergoing a data revolution. Data science is rapidly transforming how medicine is understood, how biomedical research is conducted and how healthcare is delivered [1][2]. At the same time, there is a shortage of data and digital skills in the healthcare sector [3], posing a threat to realising the full potential of data-intensive medicine.

There is an imperative need to train current and future medics in data science [3]-[5]. According to the Topol review, "within 20 years, 90% of all jobs in the NHS will require some element of digital skills" [6].

Despite the growing interest in health data science training, educational programmes around the world are, overall, lagging. There have been some recent developments, such as training funded by Health Data Research UK and the Big Data to Knowledge programme in the US, but the vast majority of these are targeted to graduate more advanced students. Data science training in the undergraduate medical curriculum is scarce. This is in stark contrast with experts' recommendations regarding developing a basic understanding of data science in undergraduates [6]-[8].

There is also hardly any published research on how to teach health data science. There are some recent publications on data science education for general audiences or

statistics/computer science students [9]-[11], but it is unknown to what extent the lessons learnt apply to undergraduate medical students. As Dunn and Bourne put it, “the [biomedical] community doesn't know how to do or review data science training” [12]. This poses a huge challenge to course instructors wishing to offer data science courses in medical curricula.

This paper addresses this gap in two ways. First, we present an undergraduate data science course for medical students, which is the first of its kind in the UK. We describe the course syllabus, assessment and flipped classroom delivery, and we discuss key pedagogical considerations. This can serve as inspiration, or even as a template, to colleagues that plan to develop similar courses.

Second, we investigate student perspectives on the learning and teaching of data science in undergraduate medicine. Following a mixed methods approach, we shed light into student perceptions, experiences and opinions, including general attitudes and data skills deemed important. We also present an experiment regarding different modalities of online programming labs, thus providing insight into hands-on data skills development.

2 Course Overview

Data Science in Medicine (DSM) is a 6-week compulsory course for Year 2 undergraduate medicine students at the University of Edinburgh. It has been offered since 2018 to approximately 210 students annually. The course aim is to equip students with the key foundations and data skills for the data-intensive medicine of the future. Being an introductory-level course, no prior data science or programming experience is assumed. Note that the course has been offered in combination with an older epidemiology course, but here we focus on the data science component.

There are three course **themes** (see Table 1). More emphasis is placed on the first theme, following literature recommendations regarding the prominent role of statistical foundations and data analytics [6], [7], [13], [14]. Relational databases and knowledge graphs are also included, given the great importance of data representation principles.

Course theme	Topics covered	Duration
Statistical analysis of biomedical data	• Summarising Data with Statistics: ○ Data scales ○ Summary statistics: mean, median, mode, range, variance, standard deviation ○ Populations and samples • Visualising Data: ○ Visualisations for quantitative data: histograms, box plots ○ Visualisations for qualitative data: bar charts, pie charts ○ Visualisations for bivariate data: scatter plots, line graphs • Hypothesis Testing: ○ Correlation between numerical variables: correlation coefficient and testing ○ Association between categorical variables: Chi-square test of independence ○ Comparing the mean of a sample to a population with a known mean: one sample t-test ○ Comparing the means of two samples that were independently drawn:	4 teaching units

	independent samples t-test	

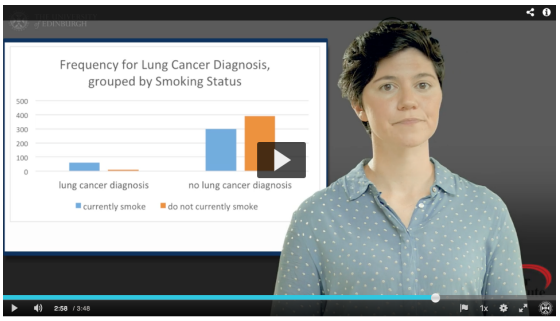
Relational databases for medicine and healthcare	• The Relational Model
	  ○ Integrity constraints in the relational model
	  ○ Creating and modifying tables with the Data Definition Language
	  ○ Declaring primary and foreign key constraints
	• Querying Relational Databases with SQL
	  ○ SQL query syntax and basic querying
	  ○ Set operations, nested queries and aggregate operations
		2 teaching units

Medical ontologies and graph data	• Graph Data & RDF
	  ○ RDF triple visualisation, unique identifiers and merging RDF data
	  ○ Expressing RDF data in the Turtle language
	• Ontologies in Medicine
	  ○ Benefits and ontology components
	  ○ Examples of medical ontologies: Gene Ontology, Disease Ontology and SNOMED-CT
		1 teaching unit

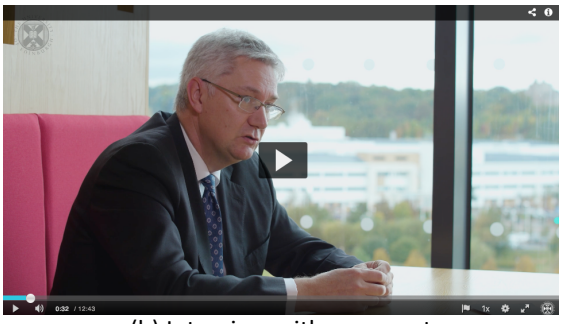
Table 1: Topics covered in the Data Science in Medicine course

The **pedagogical design** of the course is focussed on providing medical students with practical, hands-on experience of working with data. This is achieved through the use of synthetic but realistic clinical datasets, an approach that has received increasing interest [15]. This also enables the discussion of clinically-relevant cases, thus demonstrating relevance to the medical curriculum. Finally, we employ team-based learning strategies, which are associated with increased student engagement and academic performance [16].

A **blended learning** approach is adopted through a flipped classroom strategy, as it is known to lead to enhanced learning and better student motivation and engagement [17], [18]. Lectures are offered as pre-recorded online videos that the students can watch at their own time, accompanied by slides and readings. Following best practice in online learning [19], we use 5-minute high-quality production videos (see Figure 1 for examples of teaching materials and visit https://github.com/amanatak/data-science-in-medicine-2020 to freely access all course materials). Optional interview videos with experts are also included, so as to motivate the subject. The face-to-face component of the course consists of practical tutorials and computer labs. Tutorials involve data-driven clinical exercises that the students attempt in advance and discuss in groups, facilitated by a tutor (see Table 2 for a sample tutorial exercise). Inspired by problem-based learning [20], tutorials not only help students identify gaps in their knowledge, but they also contribute to the development of key skills, e.g. critical thinking and communication. The hands-on, skill-development orientation of DSM is further enabled through two programming labs on analysing data with the use of R and RStudio. R was chosen because it is powerful and versatile, and increasingly popular in the health sciences [14]. We have adopted a pair programming approach, which is an effective pedagogical tool, associated with higher student satisfaction and code quality [21].



(a) Lecture video



(b) Interview with an expert

Part 1: What's wrong with this picture?

The aim of good data graphics is to *display data accurately and clearly*. They also help us *tell a story*. However, it is not uncommon to see difficult to read, confusing or even misleading data visualisations used. In some cases, the choice of graph is not appropriate for the story that the researcher or company is trying to tell. In this part of the tutorial we'll try to "debug" data visualisations.

(1) What are the main issues in the pie chart provided below?

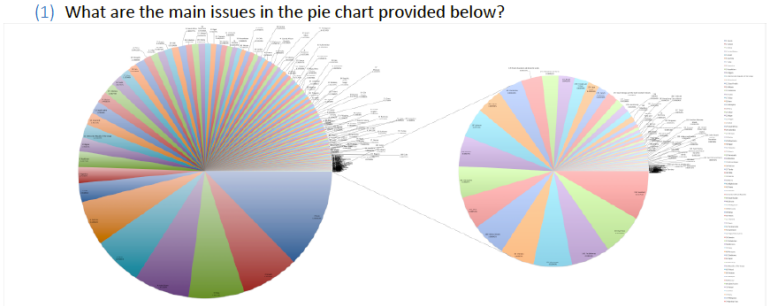
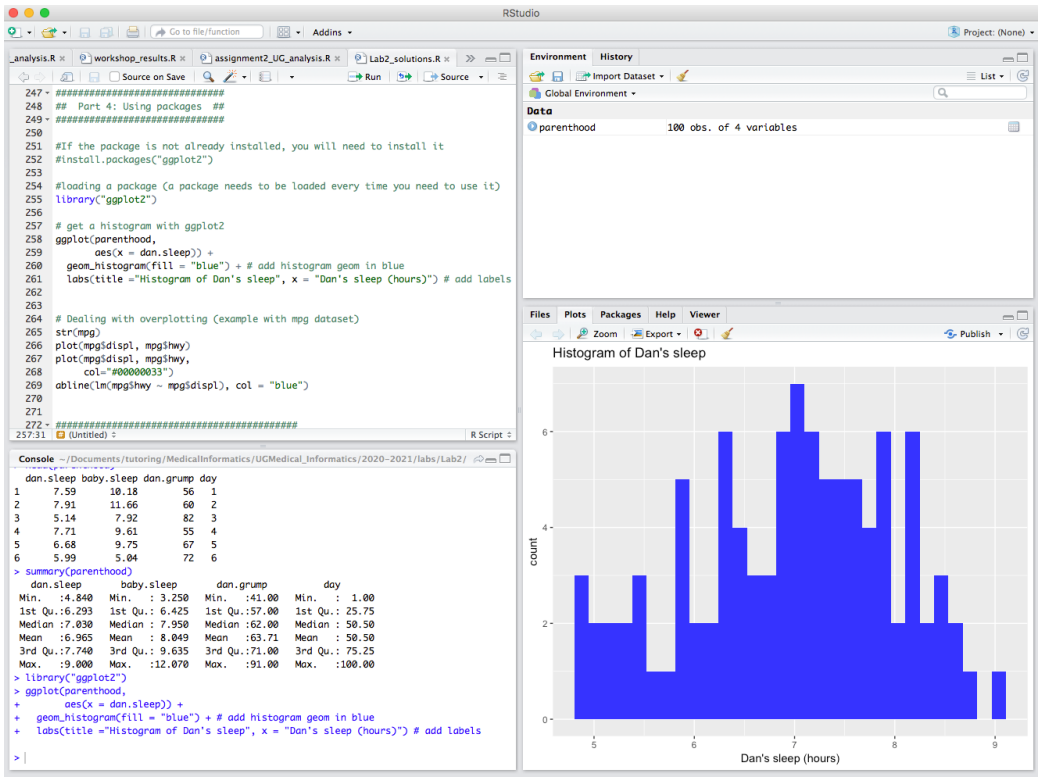


Figure 1: Pie chart of countries by area. (Attribution: Vartak.sourabh1985 at English Wikipedia)

(c) Tutorial exercise



(d) Lab worksheet in R

Figure 1: Examples of teaching materials for different learning activities. (a) Lecture on chi-square testing, where the motivating clinical question is whether there is an association between smoking status and lung cancer diagnosis. (b) Interview with an expert, discussing the opportunities that data science brings to biomedicine and healthcare. (c) Tutorial exercise, covering principles of effective data visualisation and common pitfalls. (d) Programming lab in R, focussing on statistical analysis of data.

Sample tutorial exercise: Arguing about correlation between two numerical health-related variables

Learning outcomes:

- interpret a scatterplot with regards to existence of correlation between two numerical variables
- interpret Pearson's correlation coefficient result with regards to existence of correlation between two numerical variables
- carry out hypothesis testing and draw conclusions about correlation between two numerical variables
- recognise that correlation does not imply causation

Dataset: Body Mass Index (BMI) and weekly hours of exercise for 12 study participants

Steps:

1. Draw a scatterplot for the two variables, including a line of best fit. Based on this graph, does there appear to be any correlation between BMI and weekly hours of exercise? If so, is it positive or negative?
2. Based on this sample, we estimate Pearson's correlation coefficient between BMI and weekly hours of exercise for the wider population. Its

value is -0.9829309. Is this an indication of a strong correlation? Is it positive or negative?

3. Use the statistic provided above to carry out hypothesis testing. What are the null and alternative hypotheses? What are the results of the test? What conclusions can you draw based on this test and your analysis for the previous two questions?
4. Upon presenting the results of your analysis to the local community health centre, someone from their team says: "So an increase in hours of exercise causes a decrease in BMI!" Would you agree or disagree with this statement, and why?

Table 2: Sample exercise in Tutorial 3, which is focussed on hypothesis testing.

Assessment is based on practical coursework and a final exam. There are two assignments on statistical analysis of synthetic health data (Table 3). The final exam covers all course themes and it is part of a wider knowledge test in Year 2 of the Medical School.

	Concepts	**Data**	**Examples of motivating healthcare questions**	**Data skills**
A1	Summarising and visualising data	Synthetic hospital admissions	-How are patient ages distributed for hospital admissions? -What patient ages do emergency hospital admissions involve, and how are they distributed? -What differences are there between hospitals, in terms	Data wrangling, statistical analysis, data visualisation, data storytelling.

			of patient length of stay? -What is the yearly trend of hospital admissions?	
A2	Hypothesis testing	Synthetic hospital admissions	-Is there a correlation between patient age and length of stay in hospital? -Is there an association between patient gender and type of hospital admission? -Is the average length of stay significantly different for male and female patients? -Are there any numerical variables in the dataset that are correlated with each other?	Data wrangling, exploratory data analysis, hypothesis testing, data storytelling.

Table 3: Assignments in Data Science in Medicine, and the main concepts and skills that they cover, along with motivating questions on a synthetic hospital admissions dataset.

3 Student perspectives

3.1 Research questions

The research questions investigated in this study are:

RQ1. What perceptions and opinions do medical students have around learning data science as part of their curriculum?

RQ2. What perceptions and experiences do medical students have after completing a data science course?

RQ3. What perceptions and experiences do medical students have in online instructor-led vs. video-based programming labs in data science?

3.2 Methods

The study was conducted during spring/summer 2020 in the Medical School of the University of Edinburgh, following a mixed methods approach. Focus groups were chosen to investigate RQ1 and RQ2, because they can provide insight into complex perceptions, making use of group dynamics to explore views, generate ideas and stimulate discussion [22]. A total number of 19 medical students across different study years were recruited, incentivised through a voucher, and they all provided informed consent. Three focus groups were conducted, divided by year of study (see Table 4). Given their background, all focus groups contributed to investigating RQ1, while only the Y2/3 group contributed to the study of RQ2. The focus groups had a duration of 60-75 minutes, they were conducted via a web conferencing tool due to COVID-19 restrictions, and they were moderated by the same facilitator (DD). They were audio recorded, transcribed and analysed following thematic analysis [23] with the use of the NVivo software.

Focus Group	**Number of participants**	**Year and background**	**Research Questions**
Group 1	6	Year 1, scheduled to do the DSM course in the following year	RQ1
Group 2	7	Year 2/3, had done the DSM course in the past	RQ1, RQ2
Group 3	6	Year 4/5, DSM was not part of their curriculum	RQ1

Table 4: Focus group composition and research questions investigated

RQ3 was investigated through an experiment for between-group comparison of two modalities of online programming labs: in the instructor-led lab, the facilitator demonstrated live the different steps for the students to follow on their computers; in the

video-based lab, the students watched and followed screen-capture videos on their computers at their own time. Both labs were 90 minutes long and involved the same teaching activities for analysing data with R. The facilitator was available during both labs to answer student questions. A total number of 8 participants were recruited in a similar fashion to RQ1/RQ2, but students from Years 2/3 were excluded, as they had already followed a similar lab in the DSM course. They were split in two groups of 4, one for each modality. Data was collected at the end of each lab through a short test in R, through an anonymous questionnaire consisting of five Likert-type scale questions (see Table 5) and through a brief focus group discussion. While the focus of the test was on academic performance, the survey and the focus group discussions were focussed on student perceptions and experiences with the corresponding modality (i.e. perceived usefulness, enjoyment, confidence, general feedback, etc.). The student tests were marked out of 10 points by the lab facilitator. The focus group discussions were analysed in a similar way to the ones for RQ1/RQ2.

Question ID	**Question**	**Possible answers**
Question 1	How would you rate the programming workshop?	1-10, where 1 corresponds to terrible and 10 to excellent
Question 2	How useful did you find the live demonstration/videos for step-by-step learning?	1-10, where 1 corresponds to not useful at all and 10 to very useful
Question 3	How supported did you feel by having an instructor in the 'room' for answering questions you may have had?	1-10, where 1 corresponds to not supported at all and 10 to very supported

Question 4	How confident do you feel about your R skills developed through the workshop?	1-10, where 1 corresponds to not confident at all and 10 to very confident
Question 5	How much did you enjoy the programming workshop?	1-10, where 1 corresponds to hated it and 10 to loved it

Table 5: Questionnaire used as part of the experiment for RQ3

3.3 Results

3.3.1 General student perceptions about learning data science as part of the medical curriculum

In order to shed light into how medical students **conceptualise** data science, participants were asked to state some words/phrases that come to mind related to data science in medicine. A range of terms was identified, some generic and some more medicine-focussed (see Figure 2). It is interesting to note that coding, R and Python were also mentioned by Year 1 students, who had not been exposed to DSM.

Figure 2: Word cloud of terms indicated by medical students as associated with data science in medicine. Font size corresponds to term frequency.

Regarding **general attitudes and opinions** about including a data science course in the medical curriculum, all Year 1 participants agreed that it's a good idea, since they believe

that future doctors should have a basic understanding of data science. Opinions varied between Year 4/5 participants. Four out of six said that data skills would be extremely useful, while the other two pointed out other skills that were deemed more important, e.g. patient communication.

When asked which **data skills they thought were important** in the medical curriculum and practice, the vast majority of participants identified statistics as essential, with some highlighting data analysis and visualisation. Opinions were split about programming skills, with some stating that only a basic understanding is needed, and others emphasising the importance of R programming. IT skills were identified as important by a small number of participants, with one student mentioning SQL database skills and another one data security.

Participants also discussed **data skills that they would like to learn** as part of their studies. Almost all Year 1 students agreed that a broad understanding of data science fundamentals would be useful, which can then be further specialised based on their career path. Almost all Year 4/5 students agreed that statistical analysis tools would be useful, while Year 2/3 students emphasised R programming.

Participants highlighted case studies and clinical examples as a good way to **motivate data science topics in the medical curriculum** and show relevance. Discussing data-driven dissertations, organising a hackathon and including group projects were also suggested by some participants.

3.3.2 Experiences and perceptions of medical students after completing a data science course

All participants indicated a positive change in their opinion about data science upon completing the DSM course. Two students stated that they realised that “programming is

something that you can learn", with one of them saying: "at the beginning when I saw code I would just get a panic attack […], but then the course has shown me that this is something that can be learnt […] and that R and data science is quite an interesting topic". One participant pointed out that "even though I didn't particularly enjoy the course, it opened my eyes to a side of medicine that I hadn't really considered before". Another participant said that they became very interested in data science, leading them to choose a different intercalation choice in Year 3 and perhaps even a different career in the future. Two other participants mentioned that only after conducting their research project in Year 3 did they realise how valuable the course was.

The majority of students identified the online video lectures as one of their favourite aspects of the course, as they allowed for self-paced learning and supported refreshing their knowledge in subsequent years. Regarding course weaknesses, all participants mentioned that the programming labs were overcrowded and some indicated that they would prefer a self-paced rather than an instructor-led lab. They also recommended splitting assignments in smaller tasks spread out during the semester.

3.3.3 Instructor-led vs. video-based programming labs in data science

Upon marking the end-of-lab tests, we found that the student scores were comparable (see Table 6).

Programming lab modality	**Mean of test score (out of 10)**	**Standard deviation of test score**
Instructor-led	8	2.16
Video-based	8	1.41

Table 6: Student test scores in each programming lab modality

The questionnaire results (Figure 3) indicate that the students in the video-based lab were on average more satisfied and found the lab more useful and enjoyable. They also felt more supported and confident about the skills developed. An independent samples t-test showed a statistically significant difference between the two groups for Question 3.

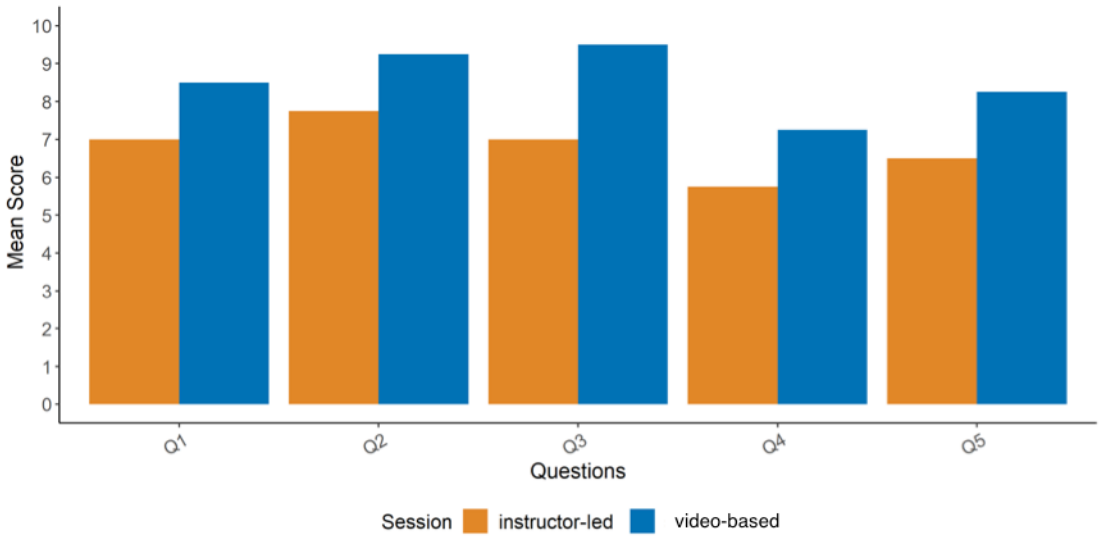


Figure 3: Questionnaire results averaged by group

The findings from the focus group discussions confirm these results and shed further light into the student experience. All participants in the video-based group agreed that they prefer this modality to the instructor-led one, as it gives them more flexibility. A student commented that “doing it in a big group together, there’s pressure about finishing, like who gets there quicker.” All participants found it a good idea to follow the videos at their own time at home and then join a live Q&A session with a demonstrator.

Regarding the instructor-led session, all participants said they enjoyed it, but they did not feel confident to apply what they learnt without additional practice. There were mixed opinions regarding lab modality, with some students indicating the flexibility brought by videos and others highlighting that a live demonstration can be more engaging and motivating.

4 Discussion

4.1 Lessons learnt from the Data Science in Medicine course

The DSM course is the first of its kind in the UK, and it has been characterised by the external examiner as "being at the cutting edge of medical education". The impact of the course has been significant. By offering it in Year 2, students have been better prepared for data-driven projects in Years 2, 3 and 5 of the medical curriculum (part of "student selected components" in the Edinburgh Medical School). Colleagues supervising student research projects have reported improved data competency compared to previous student cohorts that had not been offered the course. Improved data skills also mean that the students have been better equipped for certain intercalation options at Edinburgh and elsewhere that require a basic understanding of data science, such as Psychology, Epidemiology, Molecular Genetics, as well as Mathematics, Computers and Medicine, or even Medical Physics and Biomedical Engineering. Several students have decided to dedicate their intercalation year to data science/computational degrees, as a direct result of following the DSM course. It will be interesting to see how students use the course in the workplace, as they start graduating within the next few years.

Developing such an innovative course has been a rewarding and creative experience, but not without its challenges, especially given the lack of literature to guide this process. Our, rather bold, decision to focus on practical data skill development has been a successful one. Students performed very well in their practical coursework and quickly grasped new programming concepts.

Keeping data science tasks as simple as possible has been key in achieving this. This is somewhat different to "generic" and longer/advanced data science courses described in literature [9], [10]. Deciding on the level of difficulty and the course topics has been a non-

trivial task, and it has been continuously improved over course iterations. We consider the syllabus presented in Section 2 to be stable for future offerings.

We were faced with some logistical challenges when running programming labs, in particular large group sizes and impractical space layout for students to follow the instructor-led steps. Self-directed, video-based labs (on campus or online) that are supported by a demonstrator provide a good alternative. The experiment presented in Section 3.3.3 further supports this.

Motivating the course and showing relevance to clinical practice was one of the biggest challenges. For many students, learning data science was outside their interests and it was hard to convince them how these skills will be useful in their future studies and career. We anticipate this to be a challenge for other universities too. The use of clinically-relevant datasets and interview videos with experts has improved the course relevance. Another idea involves presenting the course around a set of diverse case studies [9], ideally from the viewpoint of different medical career paths.

The flipped classroom approach was deemed successful. The lecture videos were highly rated, and students valued the feedback received in the tutorials. A key success factor was team-based learning during the face-to-face component, as well guiding and supporting students during pre-class tasks through the virtual learning environment [24].

4.2 Discussion of findings from small-scale study

The study on student perspectives confirmed some of our previous experiences with DSM and highlighted some areas for future research. Regarding perceptions around including data science in the medical curriculum, it was interesting to see a difference between Year 1 (all positive) and Year 4/5 students (mixed opinions). It is worth investigating whether this difference can be replicated in a bigger study and how it can be explained. The perceived

importance of statistics by students is in line with literature recommendations [6], [7] and confirms our choice to place more emphasis on this theme in DSM.

The success of DSM can be best demonstrated through the change in student perceptions after taking the course. Broadening students' horizons and inspiring them to explore data-intensive career paths is, in our opinion, a testament to the power of data science in health. We believe that this will also be seen in similar courses developed by other universities in the future.

One of the most interesting findings was the fact that video-based online labs were, in general, preferred over instructor-led online labs, and they were found to be more useful and enjoyable, without leading to any significant difference in academic performance. This is particularly important in times of COVID-19 hybrid/online teaching.

It is worth noting that this is a preliminary study, carried out in a single UK-based university with the participation of 27 students, which is a small sample. General limitations of focus groups and surveys also apply here, including sampling bias (e.g. students that volunteered to participate in the focus groups may not be representative of the entire student population for DSM) and instrument bias (e.g. participants may have misunderstood some survey questions). Sampling bias could also affect the experiment results, for instance regarding learning capabilities. Given these limitations, further research is needed to confirm the generalisability of our findings, ideally with a larger sample of participants and in several different countries.

4.3 Concluding remarks

Teaching data science to undergraduate medicine students is a strategic priority worldwide, yet there is no published research describing existing courses or shedding light on effective pedagogy. Aiming to address this gap, we described in this paper a highly innovative course

in a UK university that introduces undergraduate medical students to data science. We presented key design decisions and lessons learnt, and by making the course materials openly licensed and freely available at https://github.com/amanatak/data-science-in-medicine-2020, we invite the rest of the community to adapt them for their own curricula. We consider this to be an important contribution, and a first step towards getting the conversation started in the health data science education community.

The preliminary study on student perspectives provides further insight into the learning and teaching of data science in undergraduate medicine, including data skills and topics that are perceived as important and different programming lab modalities. These findings can help even further shape undergraduate medicine curricula in data science.

Future studies should investigate student perspectives at a greater scale, across different countries and larger student populations. Further reporting on similar courses is needed, along with pedagogical evaluation, so as to help the community effectively respond to the widely-recognised need to train future medics in data skills.

Acknowledgements

The authors would like to thank the Usher Institute at the University of Edinburgh for funding this work through an Usher Education Development Grant. We would also like to thank the following students from the Medical School for helping us recruit participants for this study: Adam Tobias, Aya Riad, and Muchen Jiang.

References

[1] Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. The New England Journal of Medicine. 2016; 375(13):1216-1219.

[2] 2019a: Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine. 2019; 25(1):44-56.

[3] Sharma V, Moulton G, Ainsworth J, Augustine T. Training digitally competent clinicians. BMJ. 2021; 372(757).

[4] Aldridge RW. Research and training recommendations for public health data science. The Lancet Public Health. 2019; 4(8):e373.

[5] Kolachalama VB, Garg PS. Machine learning and medical education. NPJ digital medicine. 2018; 1(1):1-3.

[6] Topol, E. The Topol review: preparing the healthcare workforce to deliver the digital future. Health Education England; 2019.

[7] National Academies of Sciences, Engineering, and Medicine. Data science for undergraduates: Opportunities and options. National Academies Press; 2018.

[8] Attwood TK, Blackford S, Brazas MD, Davies A, Schneider MV. A global perspective on evolving bioinformatics and data science training needs. Briefings in Bioinformatics. 2019; 20(2):398-404.

[9] Hicks SC, Irizarry RA. A guide to teaching data science. The American Statistician. 2018; 72(4):382-91.

[10] Donoghue T, Voytek B, Ellis SE. Teaching Creative and Practical Data Science at Scale. Journal of Statistics and Data Science Education. 2021; 29(sup1):S27-39.

[11] Çetinkaya-Rundel M, Ellison V. A fresh look at introductory data science. Journal of Statistics and Data Science Education. 2021; 29(sup1):S16-26.

[12] Dunn MC, Bourne PE. Building the biomedical data science workforce. PLoS Biology. 2017; 15(7):e2003082.

[13] Davies A, Mueller J, Moulton G. Core competencies for clinical informaticians: a systematic review. International Journal of Medical Informatics. 2020; 141:104237.

[14] Meyer MA. Healthcare data scientist qualifications, skills, and job focus: a content analysis of job postings. Journal of the American Medical Informatics Association. 2019; 26(5):383-391.

[15] Laderas T, Vasilevsky N, Pederson B, Haendel M, McWeeney S, Dorr DA. Teaching data science fundamentals through realistic synthetic clinical cardiovascular data. bioRxiv. 2018; 1:232611.

[16] Michaelsen, LK, Sweet, M. Team-based learning. New Directions for Teaching and Learning. 2011; 2011(128):41-51.

[17] Wilson SG. The Flipped Class: A Method to Address the Challenges of an Undergraduate Statistics Course. Teaching of Psychology. 2013; 40(3):193-199.

[18] Chen F, Lui AM, Martinelli SM. A systematic review of the effectiveness of flipped classrooms in medical education. Medical Education. 2017; 51(6):585-597.

[19] Guo PJ, Kim J, Rubin R. How video production affects student engagement: An empirical study of MOOC videos. In Proceedings of the first ACM Conference on Learning@ Scale 2014 (pp. 41-50).

[20] Wood DF. Problem based learning. BMJ. 2003; 326(7384), 328-330.

[21] Salleh N, Mendes E, Grundy J. Empirical studies of pair programming for CS/SE teaching in higher education: A systematic literature review. IEEE Transactions on Software Engineering. 2010; 37(4):509-525.

[22] Bowling A. Research methods in health: investigating health and health services. McGraw-Hill Education (UK); 2014.

[23] Patton MQ. Qualitative Research and Evaluation Methods. Thousand Oaks, CA: SAGE; 2014.

[24] Akçayır G, Akçayır M. The flipped classroom: A review of its advantages and challenges. Computers & Education. 2018; 126:334-345.