



Ellis, D. E., McQueenie, R., Williamson, A. and Wilson, P. (2020) Missed appointments in healthcare systems: A national retrospective data linkage project. In: *SAGE Research Methods Cases*. SAGE. ISBN 9781529723014 (doi:[10.4135/9781529723014](https://doi.org/10.4135/9781529723014))

The material cannot be used for any other purpose without further permission of the publisher and is for private use only.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

<http://eprints.gla.ac.uk/211624/>

Deposited on 11 March 2020

Enlighten – Research publications by members of the University of
Glasgow

<http://eprints.gla.ac.uk>

Missed Appointments in Health Care Systems: A National Retrospective Data Linkage Project

David A. Ellis

Lancaster University, UK

Ross McQueenie

University of Glasgow, UK

Andrea E. Williamson

University of Glasgow, UK

Philip Wilson

University of Aberdeen, UK

Discipline

Public Health [D26]

Sub-discipline

Health Services Research [SD-PH-3]

Academic Level

Postgraduate

Contributor Biographies

David A. Ellis holds a 50th Anniversary Lectureship in Psychology at Lancaster University and an Honorary Research Fellowship at The University of Lincoln. Much of his research considers the impact of technology on people and society, which capitalizes on methodological developments across ubiquitous computer systems—sometimes referred to as digital traces. His research has appeared in journals associated with psychological science (e.g., *Computers in Human Behavior*), medicine (e.g., *BMJ*), public health (e.g., *The Lancet Public Health*), and computer science (e.g., *The International Journal of Neural Systems*). In addition to university-led research, he continues to work collaboratively with external partners to generate applied impact, and regularly appears in the media to discuss his work. A co-investigator as part of the *ESRC* administered Centre for Research and Evidence on Security Threats (CREST), his research has also received funding from the *Chief Scientist Office*, *DSTL*, *Unilever*, and the *EPSRC*.

Ross McQueenie is a general practice and primary care researcher at University of Glasgow studying rheumatoid arthritis with UK Biobank, a large data set containing lifestyle, demographic, and biomarker data from over 500,000 participants. This work involves modeling the interaction between rheumatoid arthritis and both multimorbidity and individual long-term conditions (including stroke, atrial fibrillation, chronic obstructive pulmonary disease, and chronic pain). He has also examined the causes and effects of missed general practice appointments on patients using a first of its kind data set involving over 800,000 patients in over 130 practices across Scotland. This data linked general practice records with death data from Scottish Morbidity Records and secondary care data sets including A&E, mental health care, inpatient, and outpatient data.

Andrea E. Williamson is a senior academic GP combining teaching, research, and clinical practice. She is deputy director of the GP teaching team at Glasgow Undergraduate Medical School. Her clinical work is in addictions, and general practice at Glasgow Homeless Health Services. She is a founding member of the steering group of The Scottish Deep End Project, deputy chair of the Health Inequalities Standing Group of the RCGP, and is a past Commissioner on the Poverty Truth Commission.

She is the principal investigator on the Epidemiology of Serial Missed Appointments in the NHS study and is involved in ongoing research projects in homelessness health, adverse childhood experiences, substance use, complex interventions evaluation, and medical education research. Her overarching research interests are in engagement in care, and the care of marginalized patient groups. As an applied researcher, this involves fostering the strengths of many disciplines and methodological approaches.

Phil Wilson studied physiology at Balliol College, Oxford, followed by a Wellcome Foundation doctoral studentship in the Department of Anatomy at Oxford, and clinical studies at Cambridge. After junior hospital jobs, he became a partner in a small practice in Glasgow. In 1997, this practice became one of the first five CSO-funded research practices in Scotland and he took up a part-time appointment in the Department of General Practice at the University of Glasgow. Between 1998 and May 2001, he was medical director of WestNet, the West of Scotland Primary Care Research Network. He was appointed a primary care career scientist by the Chief Scientist Office in October 2003, with a program of research involving the role of primary care in improving the mental health of pre-school children. In August 2012, he was appointed professor of primary care and rural health at the University of Aberdeen, and in September 2017, he

became a visiting professor of child health in general practice at the University of Copenhagen. He has over 140 peer-reviewed publications that span a wide variety of clinical topics.

Published Articles

Wilson, P., McQueenie, R., Ellis, D., & Williamson, A. (2019). Missed GP appointments linked to higher risk of death. *BMJ*, *364*, 1485.

McQueenie, R., Ellis, D. A., McConnachie, A., Wilson, P., & Williamson, A. E. (2019). Morbidity, mortality and missed appointments in health care: A national retrospective data linkage study. *BMC Medicine*, *17*, 2.

Ellis, D. A., McQueenie, R., McConnachie, A., Wilson, P., & Williamson, A. (2018). Non-attending patients in general practice. *The Lancet Public Health*, *3*(3), e113.

Ellis, D. A., McQueenie, R., McConnachie, A., Wilson, P., & Williamson, A. E. (2017). Demographic and practice factors predicting repeated non-attendance in primary care: A national retrospective cohort analysis. *The Lancet Public Health*, *2*(12), e551–e559.

Williamson, A. E., Ellis, D. A., Wilson, P., McQueenie, R., & McConnachie, A. (2017). Understanding repeated non-attendance in health services: A pilot analysis of administrative data and full study protocol for a national retrospective cohort. *BMJ Open*, *7*(2), e014120.

Abstract

Health care systems across the world generate large volumes of data about patients including information about their age, sex, and medical history. It also captures information on how patients interact across multiple points of care (e.g., hospitals, dentists, and general practice).

Advances in data availability and computational power now mean that much of this data can be leveraged for social good. This ranges from the use of behavioral analytics to better predict service demand through to understanding the impact of behavior change interventions. In this

project, we used patient data to explore the causes of low engagement in health care and the impact this has on patients and services. This also involved linking data sets from different organizations (e.g., health, death, and education). We observed that serially missing general practice appointments provided a risk marker for vulnerability and poorer health outcomes. While the project was administratively and methodologically challenging, the interdisciplinary background of the team ensured that the project was ultimately successful. This was particularly important when navigating a variety of different systems used to manage and distribute sensitive patient data. Our results have already started to inform debates concerning how best to reduce non-attendance and increase patient engagement within health care systems. Following a series of high-profile publications and associated impact events, non-academic beneficiaries have included governments, policymakers, and medical practitioners.

Learning Outcomes

By the end of this case, students should be able to

- Appreciate the value of data generated by health care systems to better understand patient behavior and long-term outcomes
- Understand how similar data sets and methods might be applied to related research questions
- Evaluate the ethical and security concerns associated with research designs that rely on sensitive medical data
- Compare conceptual and methodological differences between interdisciplinary and single discipline approaches

Case Study

Project Overview and Context

Tackling health inequalities is a global health priority (World Health Organization and Commission on Social Determinants of Health, 2008). However, in order for health providers to have an effective role, it is important to understand the reasons behind, risks associated with and needs of patients who do not engage effectively with health care provision (even if it is free at the point of access). This can then help tailor services better to meet those needs. There remains a lack of published work concerning repeated missed appointments with previous research typically focusing on single missed appointments and the financial costs associated with non-attendance. One estimate has placed the cost of missed U.K. general practice (GP; community-based family medicine) appointments at £150 million per year (George & Rubin, 2003). Scottish government data also suggests that each missed hospital outpatient appointment costs National Health Services (NHS) Scotland £120 (Campbell, Millard, McCartney, & McCullough, 2015). International data on costs to health care systems are sparse. However, in complex adaptive systems like health care, the financial costs are contestable because clinicians will “catch up” or get on with other care or administrative tasks. Beyond this, it is arguably more important to understand the costs of and opportunities missed when it comes to improving patients’ health.

Factors reported to be associated with missing a single appointment include age, sex, transport logistics, and clinic or practitioner factors such as day of the week, booking efficiency, and rapport between staff and patients (e.g., Ellis & Jenkins, 2012; Murdock, Rodgers, Lindsay, & Tham, 2002; Waller & Hodgkin, 2000). Whether these factors are also associated with patients who do not attend multiple appointments remains unclear (Williamson, Ellis, Wilson, McQueenie, & McConnachie, 2017). While information about patients who miss multiple appointments has previously been limited, clinicians themselves frequently report that patients

who *serially miss appointments* (SMA) are of particular concern because they may have very poor health, may be socially disadvantaged, or high users of unscheduled care compared with patients who occasionally or never miss appointments (Ellis, McQueenie, McConnachie, Wilson, & Williamson, 2017). At present, little agreement exists on what works in practice to reduce missed appointments (Ellis & Jenkins, 2012).

Therefore, our research aimed to examine the relationship between general practice appointment attendance, health care utilization, preventive health activity, health outcomes, social circumstances, and education outcomes over time. These epidemiological findings can, in turn, help develop targeted interventions and even new data-driven tools to help improve attendance, care, and long-term patient outcomes in the future.

Section Summary

- Health inequalities remain a challenge for global health.
- Existing research concerning missed appointments and their impacts has focused on small samples involving a single missed appointment.
- Patients who serially miss appointments are suspected to have poor health and be socially disadvantaged.
- The research aimed to better understand the prevalence and impact of patient attendance across a national health system.

Research Design

The overall aim of the research project was to determine the relationship between GP appointment attendance, service-based factors, health care usage, preventive health update, health outcomes, and social circumstances across the lifespan. One way of achieving this aim is to use extracted health service and other relevant administrative data. This includes patient

appointment attendance records, age, gender, and diagnostic information, which can be retrieved from NHS GP systems directly. Other linked data were provided by a variety of other government organizations. For example, all-cause premature mortality was extracted from a separate database using a predefined follow-up period.

While some countries use a national identification number to link data sets across different systems, this is slightly more involved in the United Kingdom. Every patient in the Scottish National Health Service (NHS) has a Community Health Index (CHI) number, which is a unique identifier used across all NHS systems. This formed the cohort for the study. All data provided supplied identifiers that were probability matched to the study cohort (based on the CHI number and using other patient identifiers probabilistically for the small number of records including those where the CHI number was missing). When combined, this generates a unique set of index numbers for those individuals successfully matched to the study cohort. Each data provider will receive a different set of unique index numbers and used these index numbers as the basis of their data extract. A third party then replaced the different index numbers with a common number across all files. This common number is the unique patient identifier that the research team worked with during the research project.

Data Acquisition

It was important in the first instance to ensure that the population we wished to study existed in large enough numbers to allow for a further analysis. Therefore, we undertook some preliminary research, which confirmed that a small core group of patients who miss multiple appointments did indeed exist. In addition, the odds of missing a subsequent appointment increase among patients who had missed at least one appointment in the previous 12 months (Waller & Hodgkin, 2000; Williamson et al., 2017). This was confirmed in later work (Figure 1),

which already has significant implications for patients, practitioners, and service managers. Findings from a focus group analysis of general practitioners also showed that clinicians make clear distinctions between patients who miss a few appointments and those who miss many (Williamson et al., 2017).

[insert Figure 1 here]

Figure 1.

Caption: Distribution of patients’ total number of missed appointments over 3 years. The smaller histogram shows the same data set for patients who miss two or more appointments (Ellis, McQueenie, McConnachie, Wilson, & Williamson, 2017).

<ALT-TXT An image with two vertical bar graphs shows the distribution of patients’ missed appointments in 3 years.>

<LONG-DESC A large graph is on the bottom left of the image. Its x axis is consecutively scaled from 0 to 15 and is labeled “Number of appointments missed.” Its y axis is scaled from 0 to 300,000 with a gap of 100,000 and is labeled “Frequency.” The data shown by the graph, with values, are tabulated as follows:

No. of appointments missed	Frequency
0	297,002
1	101,450
2	47,170
3	26,660
4	17,030
5	11,478
6	8,367
7	6,257
8	4,952
9	3,874
10	3,239
11	2,587
12	2,167
13	1,910
14	1,561
15	1,318

A smaller graph is on the top right of the image. Its x axis is consecutively scaled from 2 to 15 and is labeled “Number of appointments missed.” Its y axis is scaled from 0 to 50,000 with a gap of 10,000 and is labeled “Frequency.” The data shown by the graph, with values, are tabulated as follows:

No. of appointments missed	Frequency
2	47,170
3	26,660

4	17,030
5	11,478
6	8,367
7	6,257
8	4,952
9	3,874
10	3,239
11	2,587
12	2,167
13	1,910
14	1,561
15	1,318

>

NHS general practice has almost universal coverage of the U.K. population. Patients are registered with one general practice, meaning a targeted sample of general practices can achieve population representation. Moreover, unlike most other parts of the U.K. NHS, such as specialist hospital care, where general practitioners or other clinicians control access via referral, a patient can schedule an appointment with the general practice team at their discretion. General practice appointments therefore provide an ideal starting point when seeking to understanding serial non-attendance in the context of engagement in care (Williamson et al., 2017). Following preliminary work, we then extracted NHS general practice data that were routinely collected across Scotland over 3 years. We worked directly with a trusted third party (TTP) who provided the data set, which was retrieved from the Enhanced Services Contracting Reporting Options system (EScro). EScro was originally designed to assess performance against NHS service-level agreements (Figure 2). However, this system also integrates data held in disparate clinical systems, which can then be provided to researchers via a TTP. General practices were recruited to the study by the TTP via a written request to each practice detailing the project.

[insert Figure 2 here]

Figure 2.

Caption: Pathway for general practice data transfer (prior to any linkage or analysis), which was moved between local and cloud-based systems. Primary care practices use software systems (e.g., EScro) to record information about patients.

<ALT-TXT A flowchart explains the flow of data between local and cloud-based systems.>

<LONG-DESC The flowchart is in two parts. On the left is a dotted rectangle labeled “Local Systems” and on the right is a dotted square labeled “Cloud System.” The flow between them is as follows:

On the left of Local systems is a rectangle labeled “Primary care practices” from which a downward, zigzag arrow points to a diamond-shaped box labeled “EScro Systems,” from which five downward arrows point to a rectangle labeled “Trusted third party.” A bidirectional arrow, labeled “Recruitment,” connects Primary care practices and Trusted third party. A rightward arrow from Trusted third party points to the only rectangle inside Cloud system, which is labeled “Safehaven.”

On the right of the local systems is a rectangle labeled “Research Team.” Below it is an oblong-shaped box labeled “Vetting.” Five upward arrows from it, which pass through Vetting, connect it to a line below that has five downward arrows that point to Safehaven. The bends of first three downward arrows that point to Safehaven are enclosed in circles and the bend of the fourth arrow is enclosed in a semicircle.>

Participation was specifically determined based on an opt-in basis. We did not do any sampling to ensure proportional representation. However, general practices from 11 out of 14 health boards were selected to reflect a mix of urban, rural, and areas of high socio-economic deprivation. We also included “Deep-end” practices (the 100 practices operating in the most socio-economically deprived areas of Scotland). As practices generally control access into treatment services in the U.K. health system and hold data on almost all health service encounters, they allow the examination of the association between appointment attendance patterns, long-term conditions, and a range of other health-related data.

This also allowed data to be linked from other sources with appointment patterns. For example, all-cause mortality (deaths) was linked using patient community health index numbers—a unique identifier for each patient—from our initial data set to Scottish death records databases using a predefined follow-up period. These identifiable community health index numbers were then anonymized again by the TTP, used for data extraction on this project, before being uploaded to a cloud-based system for analysis.

Data Analysis

Using analysis criteria from our pilot study, we allocated patients into the following groups: zero missed appointments (zero missed over the 3-year period), low missed appointments (less than one missed on average per year over the 3-year period), medium missed appointments (one to two missed on average per year), and high missed appointments (more than two missed on average per year). We calculated missed appointments on a per-year basis for each of the years within our 3-year study period. We calculated the mean missed appointment rate over 3 years to take account of varying appointment scheduling activity by illness episodes and social crises. Furthermore, we computed the relative contribution of patient and practice factors, both individually and collectively, to the variance in frequency of missed appointments. These data were typically modeled using Negative Binomial Regression Modeling. This type of regression is used for modeling count (frequency) data—in our design this was the number of appointments missed.

Section Summary

- The research aimed to understand how medical appointment attendance is associated with a variety of patient and system factors.
- The research design relied on the co-operation of multiple agencies and data providers throughout.
- All data for analysis was provided from a TTP who works closely with health services in Scotland.
- Final analyses were driven by key research questions relating to serial missed appointments.

Research Practicalities

Extracting secondary data from multiple sites, in this case general practices across Scotland, remains technically challenging. We were extremely fortunate to work with a TTP who have years of experience in working with NHS Scotland. While these data are used routinely to track national metrics of interest, it is only in recent years that this has also become available for research purposes. High-quality data of this nature is essential when making decisions relating to care. However, these data still require additional processing at various stages before attempting any analysis (see “Method in Action” section).

The data contained within this study did not require ethical approval due to it being regarded as a service evaluation. We obtained a letter of comfort from the West of Scotland NHS Ethics Committee and the University of Glasgow, College of Medical, Veterinary & Life Sciences Ethics Committee confirming that the full study did not need health service ethics permissions. However, we did need and obtain NHS R&D approvals and permissions to link the data from the Public Privacy Benefits Committee of NHS Scotland. Due to the sensitive nature of patient data, the data sets generated or analyzed during this study are not publicly available. Data have been made available only to the research team under controlled access and strictly for the purposes of this research study. All data were accessed and analyzed in a Safehaven, a secure cloud computer service, which researchers connect to remotely. Any results or figures generated by the research team had to be vetted and checked by staff at the Safehaven before being released for further use outside the Safehaven environment. For example, this clearance was required before sharing outputs between members of the research team and when finalizing figures for publication. This also occasionally involved data being aggregated where necessary to ensure individual patient privacy.

Section Summary

- Handling large quantities of patient data is technically challenging.
- Researchers can access these data via a safe cloud storage system.

Method in Action

The scale and depth of the GP data and linked data was widely regarded as novel and ground-breaking. For example, many analyses involved over 500,000 patients and 9,000,000 consultations, respectively, which included codes for health conditions diagnoses, social factors, practice information, and some prescription data. Linkage was then carried out with outpatient, inpatient, and A&E attendances and education data relating to school attendance and exclusions. Given the complexity of our design, we had anticipated challenges and delays, which were factored into our project planning. However, many additional issues were unexpected or became inherently more complex than expected. We worked hard to overcome these issues as they arose, but the research project also faced a number of other methodological challenges, which are also summarized below.

Data Availability and Permissions

After practice recruitment started, it quickly had to be paused as despite being advised it would not be necessary, one health board insisted we obtain NHS R&D permissions. These then had to be obtained from all participating health boards. This caused significant delay. In addition, significant amendments and new data sets were required following software updates to the systems used in general practice. As a result, practices that relied on that system had to generate new patient keys to continue providing access to this data. Our TTP had to re-index all these entries. Despite these issues, the research team remained busy by producing a categorization output plan, validating appointment data, cleaning data, and writing programming scripts for analysis.

We initially anticipated that obtaining Public Benefit and Privacy Panel for Health permissions (PBPP) would take several months following the successful completion of pilot work. However, both the research resource required to complete this process in terms of level of justification for each data item, correspondence, other paperwork changes, the requirement to conduct a public consultation, and the consequent time delays encountered had a significant impact on the research. This also triggered an internal discussion between health and education linkage colleagues at the Scottish Government and the preparation of revised legal agreements.

These delays led to a request for additional time from the funder. The specific conditions of this project being hosted in the National Safehaven meant that access to the data, the time required to access output for paper writing, and the potential further challenges we faced as we acquired our linked data meant that we had to seek additional funding to answer our research questions. These requests were ultimately successful and allowed the project to continue.

Data Processing and Computational Challenges

More time was spent on processing data for analysis than running the statistical models themselves. For example, data concerning a patient's attendance or absence is based on the duration of time their medical records were opened. However, these can also be opened for a variety of other reasons. Therefore, the research team had to make multiple decisions regarding what constituted a genuine appointment. Appointments were coded as attended or missed based on the duration of time between medical records being opened and then closed by a GP. However, medical records can be opened for a variety of reasons when a patient is not physically present. This involved exploring the distribution of appointment times across the entire data set alongside expert input from medical practitioners who use the system on a daily basis. As a result, we were able to generate a series of rules to determine what was a genuine appointment.

As stated previously, all our analysis took place while working in a cloud-based system (the Safehaven). However, we encountered several issues during our use of these systems. For example, during the early stages of the project, the research team were unable to install suitable software packages. These primarily concerned libraries required to support our analysis plan within R (a statistical programming language). In addition, all output had to be signed off by two members of a separate processing team which caused delays of a couple of days at times. To request output, a research assistant had to send a detailed email justifying why the output was required and the intended purpose. This again added significant workload and papers took longer to write as a result. The Safehaven also had unplanned and planned down time during working hours on a regular basis. It was impossible to foresee these technical and procedural limitations at the start of the project as these aspects of how the Safehaven operates were not made available to researchers ahead of submitting grant applications.

Computational power somewhat limited our choice of statistical analysis. However, in some respects, this became a net-positive in long run as it helped us communicate our findings clearly for readers working in a variety of fields (Ellis, McQueenie, McConnachie, Wilson, & Williamson, 2018). For example, following one publication, some researchers suggested that our analysis might benefit from a multilevel approach that would involve the use of zero-inflated negative binomial models (Smits & ter Riet, 2017). Such an approach might be particularly useful, given that 54% of patients did not miss any appointments. An initial analysis had attempted to use a mixed effects regression allowing for random practice effects, but even the simplest of models proved intractable in our data set (Ellis et al., 2018). To counter this limitation, we adjusted the analysis for available practice-level variables. The development or

application of statistical models was however less challenging when compared to the negotiation of data access and processing before any analysis.

Section Summary

- The research team encountered a number of setbacks regarding data availability throughout the project.
- Computational limitations shaped the final analysis strategy.

Practical Lessons Learned

An interdisciplinary team was essential to the project's success and while a substantial amount of communication was achieved via email and Skype, the research team found regular face-to-face meetings more beneficial. These were particularly important in relation to navigating our way through a variety of different systems used to manage and distribute sensitive patient data. The team included researchers with backgrounds in medicine, public health, statistics, psychology, and physics. While interdisciplinary teams often produce higher quality research, progress can sometimes be slower when compared to single disciplinary work due to the variety of perspectives and broad nature of discussions (Jones, 2010). These interdisciplinary teams rely on the regular transfer of knowledge between experts and organizations. This can be challenging when relying on numbers alone to convey the outputs of specific statistical models (Ellis & Merdian, 2015). Therefore, the team came to rely on a variety of innovative data visualizations throughout the project, some of which appear in subsequent publications (e.g., McQueenie, Ellis, McConnachie, Wilson, & Williamson, 2019). This was also important when establishing ground truth in data streams that were probably not originally designed to be used as

part of research. Combining these insights with expert knowledge from colleagues who are also employed as clinicians using the systems that generated our data helped reveal key nuances of appointment systems. This is something that is often missing in single discipline work that focuses on technical or predictive abilities in isolation (e.g., Computer Science), rather than considering how these data streams and insights will work in practice (Williamson et al., 2017). Therefore, an ideal research team for this type of research should always consider the suitable skills required for knowledge transfer at the outset. This includes communicating their own expertise to non-experts within that team and to outside organizations.

Our research successfully identified general practice and patient factors that help predict patterns of missed GP appointments in a large Scottish representative sample. Patients who serially miss appointments are more likely to be socially vulnerable and have high health needs. It is a strong risk factor for greatly increased mortality. However, due to the novelty of our research design, there were a number of risks associated with the project. As a research team, we were mindful of delays, but these went beyond what we expected. Fortunately, we kept our funder informed throughout and they, in turn, have been supportive to ensure the work was successful. Systems that we relied on were to some extent in their infancy when the project began and have improved over time. The lesson here: when attempting novel work assume delays and build these into the research timetable. However, this will never control for all eventualities, so expect the unexpected!

Section Summary

- Clear communication between members of an interdisciplinary team was essential for the project's success.
- Ground breaking or risky research is more susceptible to delays, but the rewards are often

likely to generate additional academic and applied impact.

Concluding Remarks

Ultimately, this case study highlights the value of and challenges associated with conducting novel, and by default, risky research. Our research observed that serial missed appointments imply low engagement in care. The work to date has identified a series of risk factors associated with this low engagement. We are now in the process of developing a risk model that can be used online and in general practices to identify high-risk patients for serial missed appointments so appointment and recall systems can be adapted.

This project was only possible after securing access to comprehensive data, which helped answer our key research questions. It was challenging and the quality of data in some cases will require attention if it is to be useful for future research. Going forward, accessing similar data may be more straightforward for researchers in the future, but improving the quality of some routinely collected patient data will also require additional resources. In addition, researchers may also need to consider how similar data sets can become freely available so other researchers can replicate their findings (Quintana, 2019).

In our analysis, we deliberately avoided computationally intensive machine learning techniques that rely on many variables to infer patterns in data without any direct input from researchers. This new trend in medicine is becoming popular as algorithms like these are, on paper, suitable for analyzing big data sets. For example, we could have used every piece of appointment and medical data to predict future health outcomes. However, for a variety of reasons, these algorithms tend to generate inflated optimism (DeMasi, Kording, & Recht, 2017). While these errors are less likely in research areas where there has been considerable technological progress (e.g., medical diagnoses from scans), predicting social outcomes (e.g.,

patients at risk) remains challenging. Indeed, traditional regression modelling with a handful of variables can outperform more computationally intensive methods (Dressel & Farid, 2018). Regression-based models also allow medical practitioners and patients to forensically examine how a system operates in practice. For example, in the case of our research, if variables used to train an algorithm were unclear to those who generate that data, then future predictions may have limited value (Nelson, Herron, Rees, & Nachev, 2019). Inaccurate predictions are of even greater concern, as they could have serious implications for patients, health practitioners and associated services. These practical limitations also sit alongside controversial data sharing issues exposed as part of other systems and research that used data without patient's consent. For example, Google recently used data from around 1.6 million patients, without asking permission (Iacobucci, 2017).

Even if computationally intensive machine learning (AI) approaches produced more accurate models on paper, they are unlikely to have helped answer our research questions. However, given our current understanding and regression-based modelling of data derived from this project, it is now possible to start developing systems that could guide clinical decision making. Therefore, working with our existing TTP, our next aim to develop a digital tool that can help GPs and practice managers predict patients at a high risk of adverse outcomes. We anticipate that this data-driven utility will initially use a traffic light system to help flag patients who are most at need. Such a tool would then be incorporated into software systems used by GPs and primary care practices across Scotland. In line with good practice and to avoid problems outlined previously, clinicians will continue to be involved with such developments throughout (Mistry, 2019).

In summary, while this research project has been challenging throughout, the research team plan to continue working together on future projects. Beyond paving the way for the development of new behavioral analytic tools, our research confirms that more attention needs paid to strategies that aim to increase engagement in care. Only then can stark inequalities in outcomes be addressed more effectively in health care provision.

Section Summary

- The research will feed into predictive models that can better identify patients with unmet need.
- Unlocking the potential for artificial intelligence in primary care should involve health care professionals, patients, and technology experts working together and engaging with policymakers and commissioners.

Classroom Discussion Questions

1. The research team had diverse backgrounds from different academic disciplines. What challenges are associated with interdisciplinary research?
2. What are potential ethical and privacy implications for large-scale projects using data linkage? How did the team mitigate these issues?
3. The researchers were unable to share their data openly as part of this project. However, open research data provides considerable scientific, societal, and economic benefits. Are there other ways in which resources from this project could be shared more widely?
4. What are the benefits and challenges of using more advanced machine learning techniques rather than the methods used as part of this research?

Multiple Choice Quiz Questions

1. Who determines the purposes for which and the manner in which any personal data are, or are to be processed?
 - a. A data controller
 - b. An ethics committee
 - c. The researchers

Correct answer: a

2. What type of analysis includes many techniques for predicting outcomes from other variables?
 - a. Regression
 - b. t-tests
 - c. Neither of the above

Correct answer: a

3. What name is commonly given to methods of data analysis that can automate analytical model building?
 - a. Artificial intelligence
 - b. Machine learning
 - c. Both of the above

Correct answer: c

4. What was the most challenging methodological aspect of the project?
 - a. Knowledge exchange
 - b. Acquiring data
 - c. Data analysis

Correct answer: b

Declaration of Conflicting Interests

The Authors declare that there is no conflict of interest.

Further Reading

Bartley, M. (2016). *Health inequality: An introduction to concepts, theories and methods*.

Hoboken, NJ: John Wiley.

Schmidt, E., & Cohen, J. (2013). *The new digital age: Reshaping the future of people, nations and business*. New York, NY: Hachette.

Web Resources

BBC (2019, January 11). *Concerns over missed doctors' appointments*. Retrieved from

<https://www.bbc.co.uk/news/av/uk-scotland-46838287/concerns-over-missed-doctors-appointments>

Ellis, D. A. (2019, August 24). *Dynamic data visualisation for psychology*. Retrieved from

<https://sites.google.com/site/psychvisualizations/>

ISD Scotland (2019, August 26). *About ISD*. Retrieved from <https://www.isdscotland.org/>

University of Glasgow (2019, August 23). *Serial missed appointments in the NHS*. Retrieved from <https://www.gla.ac.uk/researchinstitutes/healthwellbeing/research/generalpractice/research/serialmissedappts/>

References

- Campbell, K., Millard, A., McCartney, G., & McCullough, S. (2015). *Who is least likely to attend? An analysis of outpatient appointment “did not attend” (DNA) data in Scotland*. Edinburgh, UK: NHS Health Scotland.
- DeMasi, O., Kording, K., & Recht, B. (2017). Meaningless comparisons lead to false optimism in medical machine learning. *PLoS ONE*, *12*(9), e0184604.
- Ellis, D. A., & Jenkins, R. (2012). Weekday affects attendance rate for medical appointments: Large-scale data analysis and implications. *PLoS ONE*, *7*(12), e51365.
- Ellis, D. A., McQueenie, R., McConnachie, A., Wilson, P., & Williamson, A. E. (2018). Non-attending patients in general practice. *The Lancet Public Health*, *3*(3), e113.
- Ellis, D. A., McQueenie, R., McConnachie, A., Wilson, P., & Williamson, A. E. (2017). Demographic and practice factors predicting repeated non-attendance in primary care: national retrospective cohort analysis. *The Lancet Public Health*, *2*(12), e551–e559.
- Ellis, D. A., & Merdian, H. L. (2015). Thinking outside the box: Developing dynamic data visualizations for psychology with Shiny. *Frontiers in Psychology*, *6*, 1782.
- George, A., & Rubin, G. (2003). Non-attendance in general practice: A systematic review and its implications for access to primary health care. *Family Practice*, *20*, 178–184.
- Iacobucci, G. (2017). Patient data were shared with Google on an “inappropriate legal basis,” says NHS data guardian. *BMJ*, *357*, j2439.

- Jones, C. (2010). Interdisciplinary approach-advantages, disadvantages, and the future benefits of interdisciplinary studies. *ESSAI*, 7, 26.
- McQueenie, R., Ellis, D. A., McConnachie, A., Wilson, P., & Williamson, A. E. (2019). Morbidity, mortality and missed appointments in healthcare: A national retrospective data linkage study. *BMC Medicine*, 17, 2.
- Mistry, P. (2019). Artificial intelligence in primary care. *British Journal of General Practice*, 69, 422–423.
- Murdock, A., Rodgers, C., Lindsay, H., & Tham, T. C. K. (2002). Why do patients not keep their appointments? Prospective study in a gastroenterology outpatient clinic. *Journal of the Royal Society of Medicine*, 95, 284–286.
- Nelson, A., Herron, D., Rees, G., & Nachev, P. (2019). Predicting scheduled hospital attendance with artificial intelligence. *Npj Digital Medicine*, 2, Article 26.
- Quintana, D. (2019, August 7). *Synthetic datasets: A Non-technical primer for the biobehavioral sciences*. Retrieved from <https://psyarxiv.com/dmfb3/>
- Smits, F. T., & ter Riet, G. (2017). Non-attending patients in general practice. *The Lancet Public Health*, 2(12), e538.
- Waller, J., & Hodgkin, P. (2000). Defaulters in general practice: Who are they and what can be done about them? *Family Practice*, 17, 252–253.
- World Health Organization and Commission on Social Determinants of Health. (2008). *Closing the gap in a generation: Health equity through action on the social determinants of health: Final report of the commission on social determinants of health*. Retrieved from https://www.who.int/social_determinants/final_report/csdh_finalreport_2008.pdf

Williamson, A. E., Ellis, D. A., Wilson, P., McQueenie, R., & McConnachie, A. (2017).

Understanding repeated non-attendance in health services: A pilot analysis of administrative data and full study protocol for a national retrospective cohort. *BMJ Open*, 7(2), e014120.