Institute and Faculty of Actuaries

Using Predictive Analytics to Improve Health Care Demand Forecasting

by Lisa Altmann-Richer

November 2018

Title

Using Predictive Analytics to Improve Health Care Demand Forecasting

Author

L.K.A. Altmann-Richer

Abstract

Data from the health system, everyday data and genomic data have the potential to be used in predictive analytics models that help to improve forecasting of demand for health care services. A supportive infrastructure should be developed that allows interoperable multi-source data sets to be used in predictive analysis in a way that ensures individuals' data privacy requirements are met. Once this has been achieved, there may be the opportunity to use the insights from predictive models to inform network-level health system strategies. The outputs from predictive models could be used to direct the use of health services according to an individual's anticipated clinical need and taking into consideration the forecasted capacity constraints of the health care service.

Keywords

Predictive analytics; big data; health care

1. Introduction

Predictive analytics uses statistical techniques to determine patterns and predict future outcomes by utilising information from large data sets. There is a wealth of health data which could be analysed to help forecast demand for health care services.

This paper will give a brief overview of the predictive analytics process. It will then look at three broad categories of data sources that could be analysed to help forecast health care demand: health service data, everyday data and genomic data. Examples of how these sources of data have been used to create predictive analytic models will be presented and potential future uses of these sources of data considered.

Finally, the paper will consider how a supportive infrastructure could be developed to help leverage these data sources in a way that improves health care demand forecasting. Investment in technological capabilities, adherence to data protection regulations, and an appreciation for the limitations of predictive models are suggested as the 3 key requirements for advancing progress in this area. Once this infrastructure has been developed, there may be the opportunity to use the insights from predictive models to inform network-level strategies that help to direct the use of health services according to an individual's clinical need and take into account the capacity constraints of the health care service.

2. Overview of predictive analytics

In simple terms, predictive analytics is the process of learning from historical data to make predictions about future unknowns. A brief overview of predictive analytics will be given in this section to help contextualise the examples of predictive models outlined in later sections of the paper. For those interested in a more detailed overview of predictive analytics modelling, the IFoA's Modelling, Analytics and Insights Data working party produced some useful outputs in this area.

Predictive analytics can involve either supervised or unsupervised learning. With supervised learning, a target variable is identified upfront using traditional statistical techniques such as multi-variate regression. Training data is then used to create a model to help explain the underlying correlations between input variables and the key target variable. Unsupervised learning on the other hand, aims to discover trends and patterns in the data without making pre-defined assumptions about its structure and relationships. Attribute-based algorithms are constructed through the unsupervised learning process by identifying clusters and associations in the data. Unsupervised learning models can also be extended into more responsive reinforcement learning algorithms that allow the model to be dynamically updated as new data emerges. The additional flexibility of such a model means that it could be particularly useful in developing responsive forecasts of demand for health care services.

It is useful to think about the development of a responsive predictive model in six main steps. An example of how each step could be related to a model that predicts an individual's demand for health care is also given. The whole process should be iterative, with the performance captured in step six used to re-train the model and improve its predictive abilities.

Figure 1: Main steps of predictive modelling

Step	Description	Example of Health Care Demand Forecasting at an individual level
1. Obtain Data	Data to train the model is collected, cleaned and transformed into an appropriate structure for the model.	Obtain historic patient-level data e.g. electronic health records, data from wearable devices, patient admissions data.
2. Develop the model	Predictive attributes of the data are identified and algorithms utilising these attributes are developed.	Algorithms identify relationships in the patient-level data.
3. Collect case data	Collect predictive attributes for future cases of interest.	Collect patient-level data identified in step 1 from prospective patients.
4. Apply the model	Apply algorithms developed from the training data to the case attributes collected.	Apply the algorithms developed in step 2 to patients' data in step 3 to predict their probability of developing a health condition requiring medical attention.
5. Apply recommendations	Apply recommendations from the model to the case of interest.	Model may be able to forecast that primary/secondary care is needed due to likely onset of a health condition.
6. Capture performance	Record actual results from the case and compare this to the recommendations and forecasts in step 5. Use findings to help improve algorithm characterisation and the model's recommendations in the future.	Record the success of the model by identifying whether the patient required primary/secondary care as predicted by the model. Results can be fed back into the model so that it can be further refined.

Throughout this process, the crucial role played by platforms that process, store and analyse data should not be overlooked. The data sets obtained in step one above, are likely to be large data sets. Big data infrastructure will therefore be needed to store and analyse these. A recent review on 'Big Data Analytics for Genomic Medicine' identifies Hadoop, NoSQL databases and massively parallel processing as tools that can be used as part of the modelling process for this purpose (He, Ge, & He, 2017). A brief overview of these tools is given in the table below. Furthermore, once data has been stored and processed, it will be important to have a means of visualising and communicating the model's recommendations and its performance. In health care demand forecasting, outputs from predictive models can be used to help decision-makers ensure that the system is adequately resourced in line with projected demand for health services.

Figure 2: Example of tools for big data processing and storage

Hadoop	Processing of large data sets can be distributed across clusters of computers. This allows processing to be scaled up from single servers to thousands of machines. The library can recognise individual computer failures and re-distribute the processing of the data accordingly.
NoSQL	A NoSQL (non-relational) database allows data to be retrieved and stored even where models use non-tabular databases. The system can scale to clusters of machines and it can support a wide range of data structures.
Massively parallel processing	Massively parallel processing allows the co-ordinated processing of data by multiple processors working on different parts of the program. Each processor can utilise its own operating system and memory and messages are sent between processors. It can allow many databases to be searched in parallel which is useful in decision-support systems.

3. Key data sources in predictive models for forecasting health care demand

It is useful to have a wide variety of data sources for use in predictive modelling of demand for health care. This is because it can help to broaden the training data set and build up more representative patient profiles. Using a variety of data sources, predictive models can be developed in a way that more accurately predicts onset of and recommends treatment for disease.

Recently, the capability of algorithms to analyse not only structured data, but also semi-structured and unstructured data has risen. Traditionally only structured data, such as purpose-built electronic medical records or patient admissions data was available for use in predictive modelling. Developments are now enabling structured data to be analysed alongside data which is semi-structured (e.g. home monitoring, wearables) and/or unstructured (e.g. transcribed patient notes, videos, images). The use of semi-structured and unstructured clinical and behavioural data alongside structured genomic and health service data allows more sophisticated models of patient demand to be developed.

This section will draw on examples of the use of health service data, everyday data and genomic data in predictive models, considering each of these sources of data in turn. The power of combining these various sources of data should not be overlooked and this is considered in more detail in the later section on infrastructural requirements.

3.1. Health service data

There is the potential to leverage data directly from the health system to help forecast demand for health care. Data sources that have been use in predictive analytics models to date include electronic medical records, readmission rates, discharge rates and condition-specific data.

Predictive models that use data on hospital admissions, hospital discharges and condition-specific health indicators have begun to be used by health care providers. In four hospitals in Paris, 10 years' worth of hospital admission records were analysed using time series analysis provided by Intel's software engineers (Marr, 2016). The aim was to predict the arrival of patients at each hospital down to the hour. Some data, such as medical cause of admission, was restricted due to data protection laws. Nevertheless, the model was still able to leverage basic admissions data to successfully forecast admission rates up to 15 days in advance. This allowed the number of staff to be increased when a spike in admissions was projected. This example shows that even with basic hospital admissions data that does not divulge personalised patient-specific data, there is an opportunity to use predictive analytics to improve the forecasting of demand for hospital services.

Another example of the use of predictive analytics to forecast emergency hospital admissions comes from the Patient Admission and Prediction Tool (PAPT) in Australia (Jessup, Crilly, & Boyle, 2015). This model delivered real-time forecasting of hospital admissions based on 10 years' worth of admissions data from the hospital of interest itself in combination with data from surrounding hospitals. The model also incorporated data on public events, anticipated outbreaks of infectious disease and other attributes such as the day of the week. The model was tested across 27 hospitals over a five-year period in Queensland. A minimum number of admissions of 10 per day per was needed for a reliable forecast of admissions to be produced. Forecasts projected hourly admissions by gender, medical condition and severity. The model was used to help hospital bed managers be aware of the likely hospital occupancy for the coming week and to proactively plan resources accordingly.

There has also been interest in using this model for National Health Service planning. The model could help identify when hospital occupancy levels are likely to rise so high that bottlenecks in the health system would emerge and quality of care would be compromised (The Australian Hospital healthcare Bulletin, 2016). This example shows the potential value of predictive models in helping to improve patient flow. Predictive models could be used to proactively prevent overcrowding in hospitals by allowing hospital staff to discharge the lowest risk patients and thus free up capacity in the hospital system in times of a surge in demand for urgent and emergency care.

There is also the potential to use predictive models to help forecast condition-specific health care needs. Google's DeepMind set out to develop machine learning algorithms in partnership with University College London Hospital to detect differences between healthy and cancerous tissues. The aim was to use findings to improve radiotherapy treatment by targeting treatment more specifically towards cancerous tissue (University College London Hospital, 2016). In addition, researchers at the University of California developed a predictive model that used factors including gender, chest pain type and resting blood pressure to positively identify patients diagnosed with heart disease. The model could predict heart disease in patients with an accuracy greater than 90%. (Abdar & Arji, 2015). If in the future predictive models are able to predict the likely conditions for which patients will need treatment across a broad range of diseases, then health system decision-makers can try to ensure that the health system is adequately resourced to meet patient-specific needs.

Another data set that can help in this aim is electronic medical records of patients. This is arguably the most widespread health service data source currently used in predictive modelling. Electronic medical records contain details of patient-clinician interactions. This can include a whole range of information some of which will be structured and some of which will be unstructured. Data stored in electronic medical records includes notes from consultations between medical staff and the patient, the patient's medical outcomes, lab tests, vital signs, notes from carers and results from imaging such as MRI's and x-rays.

In the US, companies such as IBM Watson Health and Flatiron Health are using data from electronic medical records to try to build models that identify potential interventions and cost-saving opportunities for the health system (IBM, 2018) (Flatiron, 2018). Some further examples of the use of electronic medical records in predictive models are detailed in the table below. These examples show how electronic medical records are being used in predictive analytics models to help forecast admission, re-admission and discharge rates, as well as to streamline treatment pathways. If implemented across the health system in conjunction with other data sources considered in this paper, predictive models built from this data source could significantly improve patient outcomes and help to ensure health resources are allocated according to need and cost constraints. Furthermore, although the development of predictive models can be costly significant investment returns are possible. An example of this is the Veterans Health Administration (VHA) which is the largest integrated health care system in the US. The VHA invested an estimated $1bn in the development of a data warehouse to centralise patient's electronic data across its national health system. The VHA made a net return on this investment of an estimated $3bn over a 7-year period (Byrne, 2010). Other examples in the table below also show significant cost savings from the implementation of predictive models. Cost savings from predictive models are not unique to the health care sector. Other industries including financial services, telecommunications and retail have been able to make large cost savings by using predictive analytics models (FICO, 2018).

Figure 3: Examples of the use of electronic medical records (EMR) in predictive analytics

Example	Current use of EMR	Future potential
Parkland Health and Hospital system, Dallas Texas	Predictive analysis used to identify patients at high risk of readmission. The model was originally used for patients with congestive heart failure. It is estimated that the decline in readmissions among congestive heart failure patients saved the hospital an estimated $500,000 (Amarasingham, 2012). The model has since been extended to predict risk of readmissions for patients with a broader range of conditions including diabetes, pneumonia and acute myocardial infarction (AHRQ, 2014).	The start-up company Pieces Technologies has been set up off the back of this initial research (Pieces Technologies, 2018). The company is using artificial intelligence and natural language processing to provide software to improve patient outcomes across healthcare and community settings. The use of predictive models in this way to forecast patients at high risk of readmission can help physicians to decide which patients to discharge and at what point in time and/or which patients will require additional monitoring after discharge.
Collaboration with Google's DeepMind and Royal Free NHS Trust in North London	This collaboration developed clinical app Streams to improve prediction of outcomes for patients with kidney disease. DeepMind were able to access healthcare data on 1.6 million patients in order to help develop the model (Hodson, 2016).	The company is now focused on using AI systems to improve access to and increase the speed of care. Earlier diagnosis and treatment from using app-based clinical prediction models could help to reduce admissions by enabling conditions to be managed in a primary care setting before severity increases and hospitalisation is required.
Kaiser Permanente's Health Connect system, California	The health care company shares data across all their facilities, making it easier to access a patient's electronic medical records in a co-ordinated way. This integrated system has improved health outcomes and has saved an estimated $1bn from reduced patient visits and lab tests (McKinsey, 2013).	If electronic medical records can be shared across hospitals in the health system including between primary, secondary and community care then more powerful predictive models could be built using data from these centralised electronic medical records. Further improvements in health outcomes and cost-savings may be achieved by using centralised medical records to forecast healthcare demand and thereby adjust resources to meet anticipated patient needs.
Optum Labs, US	This research collaborative has collected electronic medical records of patients to create a database for predictive analytics tools that aim to improve the delivery of care (Optum Labs, 2018).	The predictive analytics models could help doctors to make data-driven decisions that improve patients' treatment. This could also be of particular value to patients with complex medical histories suffering from multiple co-morbidities. Tools could also be developed that use electronic medical record's data to predict those at high risk of conditions such as diabetes and intervene at an earlier stage.

The Veterans Health Administration (VHA), USA	The VHA is the largest health system in the United States. It has collected electronic data from its patients for over 30 years. Over the past 10 years or so, it has built a data warehouse to centralise patient-level data across its national sites. This has allowed the VHA to create algorithms that reliably predict outcomes such as risk of death and hospitalisation. Frontline care managers use these risk scores to guide outpatient services provided to patients (Fihn, 2014).	The data warehouse cost an estimated $1bn and cumulative net return on this investment was estimated to be approximately $3bn over a 7-year period (Byrne, 2010). This shows the potential cost savings that can be made from the ability to reliably predict patient-specific health care needs to improve health care delivery. If other health systems centralise electronic patient data, there is potential to improve health outcomes through ensuring delivery of care is targeted to those in most need.
AWARE Critical Care, produced by Ambient a Minnesota based company	AWARE's clinical decision support tool for intensive care has an algorithm that interfaces with electronic records to help reflect the clinical status of critically ill patients. Due to the critical risk of deterioration, real-time information on patient-risk is highly valuable in intensive care units. In clinical trials, intensive care units using AWARE were more efficient in gathering clinical information and made fewer cognitive errors (Pickering, 2015).	Predictive analytics on electronic medical records can be used to help inform clinical decision making in near real-time. Output could be centralised across health systems so that staffing of intensive care units meets demand.

There is also a potential role for utilising health service data in predictive analytics models that help to forecast epidemics. For example, analysis tool EpiFX developed by researchers at the University of Melbourne uses data collected by State health departments. This includes data on the arrival of patients at emergency departments and data on the symptoms from particular groups of patients where data is available such as the country's armed forces (Wells & Bennett, 2018) . Based on this data, the rate at which a pathogen is spreading can be forecast by the model.

In the past, researchers have been successful in accurately predicting flu outbreaks up to five weeks in advance. In extreme cases, there is the potential for the department of health to take action to minimise the spread of disease. For example, where appropriate local authorities could consider closing schools or public transport services to try to halt the transmission of a severe disease outbreak. Researchers are further developing the forecasting model to include infectious diseases such as Ebola and Zika. There has also been interest from the US Government's Department of Defence to adapt this tool for use in a response to bio-terrorism. This example highlights the potential importance of predictive models to not only forecast health care demand, but to enable decision-makers to respond more swiftly to the threat of large-scale disease outbreaks.

3.2. Everyday data

In recent times, technological advances have enabled health service indicators such as admission and discharge rates discussed in the previous section, to be more easily and cheaply measured. However, population-level health indicators are still less accurate and less well-validated. In addition to data that can be sourced directly from the health service, there is an increasing amount of health and lifestyle data that can be sourced from the everyday activities of individuals. The rise of 21st technologies including social media, smartphones, wearables and the Internet of Things has helped to create data sets that can be leveraged to provide population-level health indicators that improve health care demand forecasting.

Some technology companies are developing predictive models of disease based on dynamic crowd-sourced statistics on population health. These models make use of everyday data to predict an individual's risk of infectious disease. These forecasts may be more responsive than those presently available through the health service. For example, the app Sickweather scans social media networks and other 3rd party data sources for indicators of illness and provides users with their location-based probability of sickness. Based on this crowd-sourced approach, the app claims to be able to predict the rate of illness up to 15 weeks in advance with over 90% accuracy (Sick Weather, 2018). If health services are able to couple this location-based crowd-sourced sickness data with their own internal statistical data on patient admissions, this could help to create more flexible patient flow forecasts that are more dynamic in their response to changes in the external environment.

Location-based data from mobile devices can also be used to help predict how disease spreads in an epidemic. For example, phone location data was highly valuable in efforts to track population movements when predicting the spread of the Ebola virus during recent outbreaks of the disease (Wall, 2014). Data on predicted spread of epidemics can be used to direct the provision of treatment centres. It can also allow movement restrictions to be put in place when necessary to help halt the spread of the disease. The development of language recognition technologies may also have an important role to play in epidemic forecasting. Data from social media sites such as Twitter, could be used to help predict the spread of flu epidemics in near real-time.

Another example of the influence of the external environment on patient admissions is the predictive effect of the weather on medical conditions. An early example of a model that used this was the Healthy Outlook service provided by the Met Office in the winter of 2012/13 (Met Office, 2012). The service aimed to inform those with chronic obstructive pulmonary disease (COPD) about any potential adverse cold weather periods that might worsen their condition so they could take preventative action. A more recent example comes from the use of the Met Office's rainfall radar which can predict the amount of rain that will fall within a 10km radius. The rainfall forecasts have been used in a model developed by researchers that also incorporates population density, access to clean water and seasonal temperature to predict outbreaks of cholera in Yemen up to four weeks in advance (Ghosh, 2018).

Models such as this one allow health care professionals and individuals to be aware of potential health risk factors. In addition to being used to predict demand for health care under the probable onset of disease, demand forecasts can be used to guide preventative action. In the COPD example, patients can act through self-management or through seeking primary care before their condition worsens. This can help to control demand for health services by reducing the strain placed on urgent emergency care. In the case of cholera outbreaks, health workers have been sent to areas with the highest likelihood of an imminent disease outbreak to implement measures that limit the spread of infection including chlorine tablets and sanitation advice. If the prediction window could be increased to 8 weeks, then vaccination campaigns could also be targeted in accordance with the output of infectious disease forecasts.

The Internet of Things (IoT) is allowing data from wearables and other connected devices to improve health care demand forecasting. An early example of the use of connected technology in condition-specific health forecasting was Asthmapolis (Empson, 2013). The company aimed to use inhalers with GPS-enabled trackers to identify asthma trends on both an individual and population level. The idea was to use this data to help develop more personalised treatment plans for asthmatics.

The potential role for the Internet of Things in health care is expanding. It is now possible to use the IoT to keep track of patients after discharge, for example to verify that they have taken their prescribed medication or to assess the sleep and mobility patterns of patients. A recent example of this is the app Triggr Health which aims to aid addiction recovery (Triggr Health, 2018). The app can collect and measure users' screen engagement, texting habits, phone logs, sleeping patterns and location. Predictive algorithms use this data to identify potential risks of relapse and send a notification to the user's care team who would be able to check in with the user and intervene at an early stage.

Another example is the adoption of mobile and wearable technology by local NHS Trusts in partnership with Microsoft. A pilot scheme is aiming to apply machine learning to predict epileptic seizure episodes prior to occurrence. The predictive model uses data on sleep patterns, exercise and heart rate of those with epilepsy. Based on this data, the individual's doctor would be able to remotely contact the patient and alter their medication if the patient is predicted to be at high risk of epileptic episodes. Currently, there are 50 people enrolled in the myCareCentric Epilepsy programme at Poole Hospital NHS Foundation Trust and they have seen a 30% reduction in epilepsy-related hospital admissions in this group (Microsoft, 2018). Utilisation of technology in this way to improve self-management of chronic conditions could help to reduce hospital admissions as well as enable health services to make associated cost savings.

The future potential of everyday data to be integrated with health care is exciting. Other leading tech giants including Apple, IBM, Google and Amazon have also embarked on projects that aim to derive medical insights from user data in near-real time. Companies are aiming to create cloud-based healthcare analytics services that improve diagnosis and treatment of disease. Constructing a dynamic infrastructure that is capable of collecting and processing vast amounts of everyday data is likely to be the biggest future challenge. In addition, high levels of consumer engagement with and improved accuracy of connected devices are likely to be important in ensuring that data collected from these devices is reliable. Some of these challenges are considered in more detail in the later section on infrastructural requirements.

3.3. Genomic data

The potential opportunity to use genomic data to help forecast onset of disease or better target medical treatments is emerging. Implementing a large-scale population based genetic sequencing program requires a big data approach when it comes to data analysis. Classical machine learning techniques that allow researchers to identify groups or clusters of related variables may be less effective when analysing whole genome sequencing data sets. Early research indicates that semantic web technologies and deep learning may be better suited for genomic analysis (Karim, 2017). A recent review of the role of big data analytics in genomic medicine highlights the importance of cloud computing in helping to store genomic data alongside an individual's electronic medical record (He K. , 2017). The authors note the potential for over 100 million human genomes to be sequenced by 2025. It will therefore be important that technology is able to keep pace to store and analyse giant population-based data sets.

In the UK, obtaining insights from genetic data is a big area of focus with the 100,000 Genomes Project having helped to position the country as a world-leader in genome sequencing (NHS England, 2018). With the planned launch of the NHS Genomic Medicine Service in late 2018, the NHS is seeking in the long-term to transform patient care. The long-term aim is to utilise patient's genetic data to provide them with personalised medical intervention and treatment pathways. In the near-term the service will be seeking to increase the number of people receiving predictive genomic tests and to encourage patients to share their genomic data for the benefit of the research community. This data can be used to build up a comprehensive evidence-base with which to develop targeted treatments for diseases such as cancer.

Direct-to-consumer genetic tests have also entered the market. These aim to help predict individuals' risk of particular diseases. For example, the app MyGeneRank allows users to upload their genetic profile from gene sequencing companies such as 23andMe. The app then uses that data to estimate a genetic risk score for conditions such as coronary artery disease (My Gene Rank, 2018). Another company, Myriad Genetics, launched a commercial genetic test in September to use an individual's genetic data to estimate their risk of cancer including some breast, ovarian, pancreatic and colon cancers (Myriad, 2018). Technology analysts are predicting that direct-to-consumer genetic tests with accompanying risk scores for certain diseases will grow in popularity over the coming decades (Regalado, 2018). This opens up the long-term possibility of genetic report cards being issued in childhood that give the future risk of certain diseases emerging. Ethical and data privacy considerations will be crucial to consider here. If acceptable ethical and data privacy solutions are found, then there is the possibility for this data to help with long-term health resource planning and/or allow earlier intervention for those at high genetic risk of developing particular diseases.

Apps, such as the Apple Research Kit, are helping to facilitate progress in linking genomic data with health services and the medical research community. The post-natal depression app built off the Apple Research kit by the University of North Carolina's National Institute of Mental Health is the first app to use the iPhone to enable consent for DNA sample collection. It aims to help researchers understand whether there is a genetic predisposition for postnatal depression (Apple, 2018). Other apps are using the research kit to track behavioural and clinical outcomes for individuals with conditions such as Parkinson's, epilepsy, diabetes and melanoma. These apps have the potential to incorporate results of genetic tests in a similar way in the future. Combining health service, everyday and genomic data should allow a much deeper understanding of disease progression. With the right infrastructure in place, there is the potential for powerful diseases-prediction models based on a combination of these data sets to emerge in the future, that help to more reliably forecast an individual's health care needs.

4.Infrastructural requirements

Stakeholders from across the health care community, in both the public and private sectors, will need to come together to help implement predictive analytics models that successfully improve health care demand forecasting. Healthcare professionals, researchers, policy makers, insurers, educators, employers, individuals and others will all be important in achieving this aim. These stakeholders will need to work together to develop and implement the technological capabilities that allow a wide range of data to be shared and analysed across the health system. It will be important to ensure that data protection regulations are adhered to and that consumers' data privacy requirements are met. Key health care decision makers will also need to recognise the limitations of predictive models when using their insights for health care resource planning.

4.1. Creating an interoperable multi-source data system

As predictive algorithms increase in complexity and responsiveness, there's an opportunity for health systems to be more responsive to predicted changes in demand for health care. Technology needs to be developed in a way that helps to guide the decision-making of health service providers. This can be achieved by developing ways of linking outputs of predictive models into routine patient care. The clinical impact of doing this should also be evaluated and models refined where necessary to ensure that their use is helping to deliver cost-effective health care. Predictive analytics models could help to improve outcomes from the health system, for example by speeding up health care delivery and ensuring the health system is adequately resourced. In order for this to be achieved, a technological infrastructure capable of combining and analysing large data sets from a variety of sources will be required.

Ensuring that data can be accessed and pooled across the health system will be a key technological capability that needs to be developed. The benefits of achieving this have already been seen on a small scale. Six hospitals in California used the software communication tool PreManage ED. This allowed them to share patient records between emergency departments thereby enabling medical staff to understand what previous hospitals the patient was treated at, the patient's prior test results and the previous medical advice given to the patient. Treatment was therefore faster as some clinical tests did not need to be re-run. The patient's likely demand for health care on arrival could be more readily understood, allowing care to be delivered in a more consistent and effective manner (Xu, 2017). Another example of sharing data across the health system in the US, is the Undiagnosed Disease network. This is a collaboration with the National Institute of Health and various hospitals and universities to pool data on very rare conditions. This allows a greater evidence base to be built up for patient outcomes for these rare conditions.

When implementing such a system that collates patient data for use in research programs, infrastructure needs to be put in place to anonymise patient data at the point of entry. Anonymised patient-outcomes can then be measured in near real-time to link the health interventions with these outcomes according to different patient attributes. The results of this research can then be used to help inform treatment decisions for other patients with similar attributes. At present, there is a large lag between the intervention and the collection of data on health outcomes. Solutions that incorporate everyday data that lies outside of the health system could help to bridge this evidence gap.

A key challenge in using predictive models to affect change at the health-system level is that at a national level data is split across multiple entities and stored in different formats. A key technological requirement should be that relevant data is accessible across the health system and that data sets are able to interface with each other. An easy way to achieve interoperability would be to store data within a single central structure or warehouse. However, with the vast amounts of data available from outside the health system this seems unrealistic, especially as large companies are moving towards cloud-based health analytics. Therefore, in a cloud-based data storage age, data should be exchangeable by different players and devices across system.

A key driver of this will be that data is stored in a common programming language. This will allow fragmented systems to communicate and interpret any data exchanged across the system and then relay findings to clinicians and health-care decision makers (He, Ge, & He, 2017). Projects such as the Fast Healthcare Interoperability Resources (FHIR) specification project are helping to achieve this. This project has produced a data exchange and information modelling standard for electronic health records that can allow this data to be stored and correlated with other health system data. The aim is to facilitate interoperability that allows electronic medical records data, including data from legacy health systems, to interface with modern consumer facing devices such as mobile apps and wearables. The common language also allows for integration with third-party app developers who could build additional medical applications on top of the underlying data structure.

The importance of algorithms that can combine data from these various sources - health system data, everyday data and genomic data - should not be underestimated in achieving the goal of a health system that is responsive to predicted changes in demand. A study of 3,426 Swedish individuals showed that combining genetic data with environmental data has shown more potential in predicting the onset of type 2 diabetes than either data source alone (Zarkoob, 2017). Patient feedback metrics could be coupled with data from other sources such as social media and hospital admissions statistics to try to identify when services start to get overwhelmed. The distribution of patients between community, primary, secondary and tertiary care could then be adjusted in response to this.

An early example of combining patient feedback with data from other sources was published in the BMJ in 2017 using patient feedback on the NHS in England (Griffiths & Leaver, 2017). The researchers combined feedback from official NHS patient monitoring services, such as NHS Choices and Patient Opinion, with data from social media sites Facebook and Twitter. Using this data, they constructed a near real-time collective judgement score for acute hospitals. Logistic regression analysis showed a positive relationship between this aggregated patient feedback and the results of the Care Quality Commission's inspection outcomes on that day. This finding suggests that models combining patient-feedback from official channels with social media data could help to identify hospitals at high-risk of falling standards. This could allow health care decision makers to take corrective action.

Implementing population-level data driven health strategies will require a data-literate workforce with the technical-expertise to understand how to manage, analyse and interpret the complex data. To use predictive models to improve patient care, organisations will need to invest in infrastructure and analytical tools. An example of this comes from the Pittsburgh Health Data Alliance. The collaboration between Carnegie Mellon University, University of Pittsburgh and UPMC Enterprises is seeking to use data from electronic medical records, wearables, genomics and social media to try to build up a picture of the patient and tailor their health care accordingly (Pittsburgh Health Data Alliance, 2018). The individual patient's data isn't treated in isolation. By analysing it alongside data from other patients, it can highlight specific threats and issues through patterns that emerge during the comparison. Predictive models could enable doctors to assess the likely outcomes of the different treatment options available to the patient by using data from the outcomes of other similar patients to assess these.

4.2. Ensuring data privacy

To achieve a system that uses predictive analytics to guide patient care, individuals will need to feel empowered to become active participants in their health. To secure positive patient engagement, individuals will need to be assured of data protection. This will be important as patient data has the potential to be used in predictive models that draw on the evidence-base to attribute particular characteristics to the probability of other similar patients getting sick with a particular disease and/or to forecast the likely success of specific medical interventions.

With advances in cloud-computing technology offering solutions to storing and exchanging data, there is the potential to incorporate data from sources that are external to healthcare organisations to help predict illness and guide a patient's treatment. For individuals to be willing to share data from apps, wearables and other devices with health care providers, strong data privacy and security protocols will need to be implemented. In the example given earlier in this paper of NHS Poole using the myCareCentric epilepsy app, the patient's data is being stored in Microsoft's secure cloud platform Azure. This highlights the focus that tech companies and health services are already placing on ensuring that patient data is protected.

Threats to data privacy and confidentiality are a key risk. Some examples of recent data privacy leaks include the "WannaCry" malware attack on NHS England and some US medical device companies, a cyberattack on Singapore's government health database that obtained non-medical personal information of an estimated 1.5 million people, the hacking of identity information from US health insurer Anthem and the warning of Johnson & Johnson of a cyber-security bug in one of their insulin pumps that could potentially be exploited by hackers (Reuters, 2018) (Mathews, 2015) (Finkle, 2016). Given these historic examples, individuals may be concerned that their sensitive health-related data from electronic medical records, mobile health apps, wearable devices and genomic data could be divulged to unwanted third parties in the event of a cyber-attack.

Beyond the cyber-security risks, individuals are also keen to only authorise their health data to be used only for specific purposes. A recent survey by Accenture in the US looked at consumers' willingness to share data from wearables and mobile apps. Although, 90% of the approximately 2,300 people surveyed would be willing to share this data with their physician, there was greater resistance to sharing data with health insurers and even more resistance to sharing data in the workplace, with only 38% willing to share their data with their employer (Accenture Consulting, 2018). There appears to be even greater resistance by consumers to the sharing of their genomic data. The House of Commons Science and Technology report has highlighted patients' unwillingness to disclose genetic data as a key challenge that could hold back progress in using genomic sequencing to improve medical diagnosis and treatment (House of Commons Science and Technology Committee, 2018). Whilst most people surveyed by the British Science Foundation were willing to share their genomic data with university researchers and NHS workers, 95% of those surveyed would be unwilling to disclose their genomic data to private companies. Ensuring strong data protection processes that can secure and anonymise personal patient-data to allow it to be used for the benefit of population-level health care demand forecasting, without the necessity to disclose results to unwanted third parties is likely to be important.

Groups such as the USC Centre for Body Computing in California are looking at ways in which data privacy can be embedded into health IT solutions (USC Centre for Body Computing, 2018). Block chain technology is likely to provide a possible solution to both the cyber-security risks and the requirement of a system that allows the individual to be in control of who is able to access their data. If block chain is implemented in the right way, it could be used to provide an ultra-secure store of information across de-centralised systems. Technology could enable each individual to be the owner of their own personal health record. This record could contain medical data from physician interactions, everyday data from wearables and IoT devices and genomic data. The personal health record would be owned by the individual and this would allow them to manage the data that they share with each organisation.

Putting the individual in control of their own data management is likely to help them become active participants in their health, as well as alleviate some of their data privacy concerns. Individuals who share their anonymised data with the health service could help to contribute towards improving medical research and health care delivery. In turn, they could benefit from feedback on preventative health measures that they should undertake based on near real-time feedback of the data that their personal health record is collecting. They could also be more speedily guided towards appropriate health services in times of acute need based on a dynamic analysis of the data collected by the health record. Meanwhile, health service providers would be able to collate personal health record data from across the health system to help predict demand for services and ensure that the system is adequately resourced to meet patients' needs.

4.3. Recognising the limitations of predictive models

When implementing predictive analytics into healthcare demand forecasting, it's important to appreciate the limitations of predictive models. If health care decision-makers are not aware of the limitations of models, then this could lead to forecasts from models being applied out of context to health care resourcing decisions. A recent survey by the Society of Actuaries of 223 health payer and provider executives in the US found that 47% were currently using predictive analytics, with a further 42% planning to being using these models in the next 5 years (Society of Actuaries, 2017). Therefore, there could be an important role for the actuarial profession, not only in helping to build some of the predictive models described in this paper, but also in helping to highlight and quantify the limitations and uncertainties inherent in their output.

If data sources contain incomplete observations, selection bias or confounding medical interventions then this can decrease the accuracy of predictive models. For example, electronic medical records often contain systematic biases. For example, selection bias may be present whereby the sickest patients are the ones most likely to receive testing. One research study using electronic medical records data on more than half a million patients across two large hospitals has shown that the timing of a diagnostic tests can be more predictive than the result of the test (Agniel, 2018). The presence of these biases can violate assumptions made by off-the-shelf machine learning algorithms used to train predictive models (Paxton, 2013). Therefore, the use of dedicated health-care specific models that recognise the limitations of the data inputs will be important. Another example, the use of discharge-prediction processes in acute care hospitals. One team of researchers found that lack of descriptive detail in the data and its reporting could be holding back progress in developing a discharge prediction tools (De Grood, 2016). Actuaries could look to work with health service providers to highlight the key threats to data quality and help to suggest ways that the quality of the data being input into predictive models could be improved. This should help to improve the statistical validity and credibility of data sets and thereby improve the confidence that the user can place in the model's output.

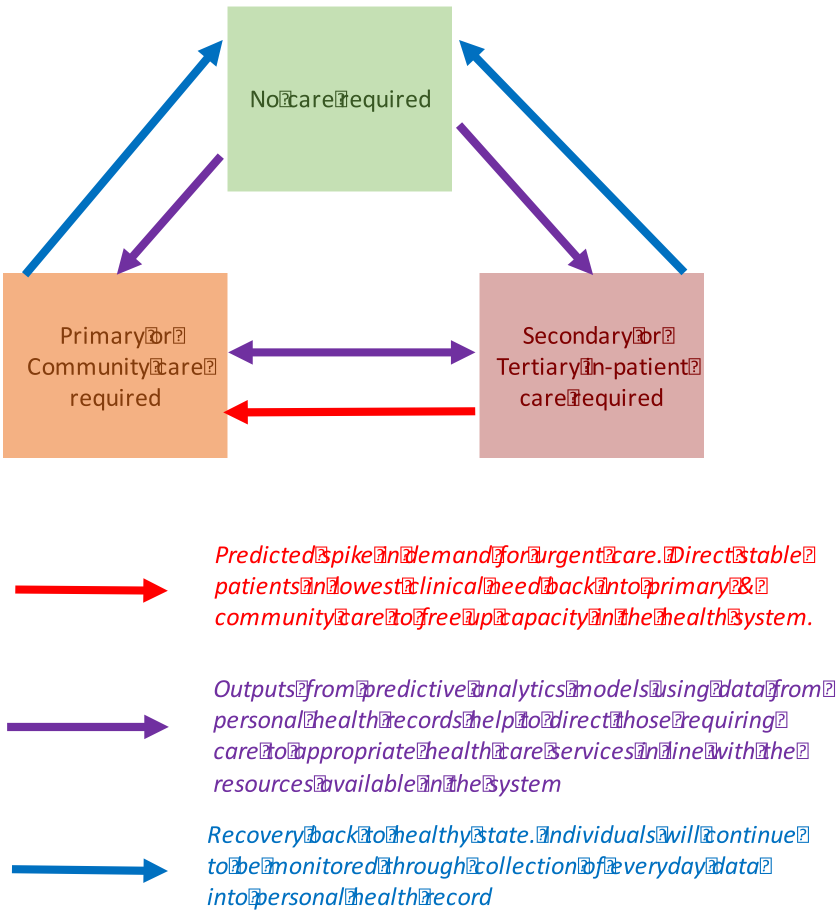
A further key limitation of predictive models is that history cannot always predict the future. In predictive analytics, we are combining and analysing data sources to help identify patterns that can be used in future prediction. However, as new medical threats and diseases emerge and medical technology advances even a highly responsive algorithm could be unable to accurately predict an outbreak of diseases, the impact of certain epidemics or a shift in the pattern of the development of chronic diseases. There could be an important role here for actuaries in terms of aiding risk management by clearly explaining the weaknesses and limitations inherent in the modelling process. Actuaries could look to quantify the uncertainty inherent in predictive models and communicate to health care decision makers the underlying uncertainties in the forecasts of demand for health care services.

5. Creating a network-level strategy for managing health service demand

It could be possible to use data from health systems, everyday patient data and genomic data to develop predictive-analytics models that are used in network-level strategies to manage patient demand across the health system. As discussed in the proceeding section, a suitable infrastructure will be required. This is one that ensures data privacy is achieved, allows an interoperable data system to be set up and helps to quantify and reduce uncertainties in demand forecasts. Setting up this infrastructure is likely to require significant financial and time investment. However, significant gains from investment in predictive models are likely. More than a quarter of the health care providers surveyed by the Society of Actuaries estimate predictive analytics process will enable them to achieve future cost savings of more than 25% over a 5-year period (Society of Actuaries, 2017). If predictive models are implemented across providers at a wider network level, then even greater financial savings may be possible.

As shown in the diagram below, a network-level strategy could direct patients between community-based or primary care and hospital-based secondary and tertiary inpatient care in line with their medical need and the capacity of the health system.

Figure 4: Diagram of a network-level strategy for using predictive analytics to respond to health care demand forecasts



Predictive models can allow those in the community setting at high risk of requiring medical care in the future to be identified. This could be achieved through predictive analysis of an individual's personal health record that collates everyday data in real-time from monitoring devices such as wearables and mobile apps. The record would also contain the individual's electronic medical records and/or genomic data. Based on the dynamic output of a multi-data source predictive model, the individual can be notified of an upcoming health risk. They can then be directed towards primary care before their condition deteriorates or be proactively triaged into secondary care if their condition requires more acute medical care.

In addition, personal health record data can be aggregated up to a higher level to provide insights including patient-feedback of hospital services, community-based sickness trends and overall hospital admissions. Predictive models can be used to identify when demand for secondary and tertiary in-patient care is likely to surge. Resources including such as staff, medical equipment and pharmaceutical supplies could be increased across the health system to help match demand. If hospitals are predicted to reach capacity, then data from personal health records including from electronic medical records and medical monitoring devices could be used help to identify those patients currently in secondary and tertiary care that are the most stable and have the lowest risk of developing health complications. These patients could then be directed back into primary and community care, with wearable and IoT devices allowing them to continue to be monitored outside of the traditional hospital setting. This would allow the system to free up resources in anticipation of a demand surge for urgent hospital care by prioritising patients in highest clinical need of treatment within specialised medical facilities and discharging patients who could be managed within the community/primary care setting.

Developing a network-level system that responds to health care demand forecasts in this way is likely to require significant investment of time and capital to build a supportive infrastructure. With a supportive infrastructure in place, predictive models that analyse historical trends and patterns in population-level patient data could be used to produce dynamic forecasts of the likely future healthcare demands across the health system. If health-care decision makers can successfully interpret this data, then they could build response mechanisms that help to allocate health care resources according to anticipated patient need and that help to prevent health care systems from becoming overwhelmed in times of projected peak demand.

6. Conclusion

This paper has given a brief overview of predictive analytics and big data infrastructures. It has then gone on to consider how health service data, everyday data and genomic data have been incorporated into early predictive analytics models for health care demand. It has also highlighted potential future developments for predictive models using these data sources. By implementing an interoperable data infrastructure that allows these various sources of data to be combined whilst respecting individuals' data privacy concerns, health care demand forecasting could be significantly improved. Actuaries could be involved in this process by helping to build highly responsive predictive models and by helping to manage the risks inherent in the modelling process through quantifying and helping to reduce the uncertainty of a model's output. Based on the predictions of these models, a network-level strategy could be implemented in response to healthcare demand forecasts. This would help to direct individuals towards health care services according to anticipated clinical need whilst taking into consideration the capacity constraints of the health system.

Acknowledgements

The author would like to thank Ewen Tweedie and Philippa Gillham for their guidance and support in producing this paper.

References

Abdar, M., & Arji, G. (2015, December 6). Comparing performance of data mining algorithms in prediction heart diseses. *International Journal of Electrical and Computer Engineering*.

Accenture Consulting. (2018). *2018 Consumer Survey on Digital Health.*

Agniel, D. (2018, April 30). Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. *BMJ*.

AHRQ. (2014). *Hospital Uses Data Analytics and Predictive Modeling To Identify and Allocate Scarce Resources to High-Risk Patients, Leading to Fewer Readmissions*. Retrieved from https://innovations.ahrq.gov/profiles/hospital-uses-data-analytics-and-predictive-modeling-identify-and-allocate-scarce-resources

Amarasingham, R. (2012, December). Applying data analytics and information exchange to improve care for patients. *Health Affairs*.

Apple. (2018). *Empowering medical researchers and doctors*. Retrieved from Apple: https://www.apple.com/uk/researchkit/

Byrne, C. (2010, April). The value from investments in health information technology at the U.S. Department of Veterans Affairs. *Health Affairs*.

De Grood, A. (2016, November 12). A Review of Discharge-Prediction Processes in Acute Care Hospitals. *Healthcare Policy*.

Empson, R. (2013, April 13). *Asthmapolis wants to hack the inhaler and help 26 million Americans better track and manage their asthma*. Retrieved from Tech Crunch: https://techcrunch.com/2013/04/05/asthmapolis-wants-to-hack-the-inhaler-and-help-26-million-americans-better-track-and-manage-their-asthma/

FICO. (2018). *Predicting behaviour puts Big data to work*. Retrieved from http://www.fico.com/en/predictive-analytics/understanding-predictive-analytics/who-uses-predictive-analytics

Fihn, S. (2014, July). Insights from advanced analytics at the Veterans Health Administration. *Health Affairs*.

Finkle, J. (2016, October 4). J&J warns diabetic patients: Insulin pump vulnerable to hacking. *Reuters*.

Flatiron. (2018). Retrieved from https://flatiron.com/

Ghosh, P. (2018, August 27). *Yemen cholera epidemic 'controlled' by computer predictions*. Retrieved from BBC News: https://www.bbc.co.uk/news/health-45259922

Griffiths, A., & Leaver, M. P. (2017, September 28). Wisdom of patients: predicting the quality of care using aggregated patient feedback. *BMJ*.

He, K. (2017, February 15). Big Data Analytics for Genomic Medicine. *International Journal of Molecular Sciences*.

He, K. Y., Ge, D., & He, M. M. (2017, February 15). Big Data Analytics for Genomic Medicine. *International Journal of Molecular Sciences*.

Hodson, H. (2016, April 29). Revealed: Google AI has access to huge haul of NHS patient data . *New Scientist*.

House of Commons Science and Technology Committee. (2018). *Genomics and genome editing in the NHS.*

IBM. (2018, April 22). *IBM Watson Health Finds Nation's Top-Performing Health Systems Deliver More Consistent Care Across Member Hospitals*. Retrieved from http://newsroom.ibm.com/2018-04-22-IBM-Watson-Health-Finds-Nations-Top-Performing-Health-Systems-Deliver-More-Consistent-Care-Across-Member-Hospitals

Jessup, M., Crilly, J., & Boyle, J. (2015, April 27). Users' experiences of an emergency department patient admission predictive tool: A qualitative evaluation. *Health Informatics Journal*.

Karim, R. (2017, May 28). *A Deep Learning Approach to Genomics Data for Population Scale Clustering and Ethnicity Prediction* . Retrieved from Insight Centre for Data Analytics: https://www.insight-centre.org/content/deep-learning-approach-genomics-data-population-scale-clustering-and-ethnicity-prediction

Marr, B. (2016, December 13). *Big Data In Healthcare: Paris Hospitals Predict Admission Rates Using Machine Learning*. Retrieved from Forbes: https://www.forbes.com/forbes/welcome/?toURL=https://www.forbes.com/sites/bernardmarr/2016/12/13/big-data-in-healthcare-paris-hospitals-predict-admission-rates-using-machine-learning/&refURL=https://healthactuarial.com/2018/06/17/using-data-science-to-help-reduce-hospital-admissions/&referrer=https://healthactuarial.com/2018/06/17/using-data-science-to-help-reduce-hospital-admissions/

Mathews, A. W. (2015, February 24). Anthem: Hacked Database Included 78.8 Million People. *The Wall Street Journal*.

McKinsey. (2013, April). *The big-data revolution in US health care*. Retrieved from McKinsey: https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/the-big-data-revolution-in-us-health-care

Met Office. (2012, November 1). *Healthy Outlook® – helping patients with COPD this winter.* Retrieved from Met Office: https://blog.metoffice.gov.uk/2012/11/01/healthy-outlook-helping-patients-with-copd-this-winter/

Microsoft. (2018, March 26). *Tech that helps epilepsy patients manage their condition 'could save NHS £250m'*. Retrieved from Microsoft: https://news.microsoft.com/en-gb/2018/03/26/tech-that-helps-epilepsy-patients-manage-their-condition-could-save-nhs-250m/

My Gene Rank. (2018). Retrieved from https://mygenerank.scripps.edu/

Myriad. (2018). Retrieved from https://new.myriadpro.com/products/myriad-myrisk/paneltests/

NHS England. (2018). *Genomics*. Retrieved from NHS England: https://www.england.nhs.uk/healthcare-science/personalisedmedicine/genomics/

Optum Labs. (2018). *Optum Labs*. Retrieved from https://www.optumlabs.com/

Paxton, C. (2013, November 16). Developing Predictive Models Using Electronic Medical Records: Challenges and Pitfalls. *AMIA Annual Symposium*.

Pickering, B. (2015, May). The implementation of clinician designed, human-centered electronic medical record viewer in the intensive care unit: a pilot step-wedge cluster randomized trial. *International Journal of Medical Informatics*.

Pieces Technologies. (2018). *Pieces Tech*. Retrieved from https://piecestech.com/

Pittsburgh Health Data Alliance. (2018). *Pittsburgh Health Data Alliance*. Retrieved from https://healthdataalliance.com/

Regalado, A. (2018, February 21). *Forecasts of genetic fate just got a lot morea ccurate*. Retrieved from MIT Technology Review: https://www.technologyreview.com/s/610251/forecasts-of-genetic-fate-just-got-a-lot-more-accurate/

Reuters. (2018, July 20). *Cyberattack on Singapore health database steals details of 1.5million*. Retrieved from https://uk.reuters.com/article/us-singapore-cyberattack/cyberattack-on-singapore-health-database-steals-details-of-1-5-million-including-pm-idUKKBN1KA14J

Sick Weather. (2018). Retrieved from http://www.sickweather.com/

Society of Actuaries. (2017). *Predictive Analytics in Healthcare Trend Forecast.*

The Australian Hospital healthcare Bulletin. (2016, September 5). *Whole-of hospital occupancy management through the ED*. Retrieved from The Australian Hospital healthcare Bulletin: https://www.hospitalhealth.com.au/content/clinical-services/news/whole-of-hospital-occupancy-management-through-the-ed-301776639

Triggr Health. (2018). *Triggr Health*. Retrieved from https://triggrhealth.com/

University College London Hospital. (2016, August 30). *Research begins to develop pioneering technology to plan radiotherapy treatment* . Retrieved from UCLH: https://www.uclh.nhs.uk/News/Pages/Researchbeginstodeveloppioneeringtechnologytoplanradiotherapytreatment.aspx

USC Centre for Body Computing. (2018). *California and Cyber-Health Security*. Retrieved from USC Body Computing: https://www.uscbodycomputing.org/cyber-health-it/

Wall, M. (2014, October 15). *Ebola: Can big data analytics help contain its spread?* Retrieved from BBC News: https://www.bbc.co.uk/news/business-29617831

Wells, L., & Bennett, H. (2018, February 13). *Forecasting Flu Outbreaks*. Retrieved from The University of Melbourne: https://pursuit.unimelb.edu.au/articles/forecasting-flu-outbreaks

Xu, X. (2017, August 10). PreManage ED™ Evaluation in Hospitals of Alameda County, California. *Journal of Patient-Centred Research and Reviews*.

Zarkoob, H. (2017, July 12). Utilization of genetic data can improve the prediction of type 2 diabetes incidence in a Swedish cohort. *PLoS One*.