Making the most of Big Data in Surgery:

Improving Outcomes, Protecting Patients & Informing Service Providers

John A. G. Gibson[1,2*], Thomas D. Dobbs[1,2], Ryan Kerstein[3], Ashley Akbari[5], Ronan Lyons[5], Hayley A. Hutchings[4], Iain S. Whitaker[1,2]

1. Reconstructive Surgery & Regenerative Medicine Research Group, Institute of Life Science, Swansea University Medical School, Swansea, UK
2. The Welsh Centre for Burns and Plastic Surgery, Morriston Hospital, Swansea, UK
3. Department of Plastic & Reconstructive Surgery, John Radcliffe Hospital, Oxford, UK.
4. Swansea University Medical School, Swansea University, Swansea, UK
5. Farr Institute of Health Informatics Research, Swansea University, Swansea, UK

Abstract

"Big data" refers to the interdisciplinary analysis of high volume, diverse clinical and lifestyle information on large patient populations. Recent advancements in data storage and electronic record keeping has enabled the expansion of research in this field. The use of such data in surgery has the potential to improve outcomes, increase safety and aid service planning. The potential has been harnessed in other specialties who are taking advantage of significant research council investment and centralisation of resources across the UK.

In this article, we provide an overview of the nascent field of big data analytics in surgery. We critically review the limitations and fears raised with large database studies, offer some suggestions regarding database optimisation creation, and suggest future directions for research in this exciting field.

Introduction

Health informatics or “big data” research in medicine refers to the interdisciplinary field of analysis of high volume, diverse clinical and lifestyle information on a large population of patients[1]. Recent advancements in data storage and the current drive to make patient records electronic has transformed the large volume of data created in healthcare, previously perceived as a by-product of healthcare delivery, to a central asset to improve patient care[2, 3]. In 2011, David Cameron, the then Prime Minister, famously quoted that every patient was a “research patient”[4]. Subsequently the Medical Research Council (MRC) has invested £90 million into health informatics[5]. Big data analysis provides a powerful approach to monitoring disease pattern and patient outcomes enabling targeted prevention strategies, improved patient safety and improved service delivery[6].

With patients becoming more and more informed, access to big data for practising surgeons will be of immense benefit. Accurate information based on representative data sets will be a vital tool in providing patient centred care. On a population level, such data could be used to develop targeted service delivery and improve cost efficiency.

As a community we have been faced with several challenges over the last few decades; notably the vaginal mesh complications in pelvic floor reconstruction[7], metal on metal joint implants in orthopaedic surgery[8] and the Silicon, Poly Implant Prothèse (PIP)[9] controversies and breast implant associated Anaplastic Large Cell Lymphoma (ALCL) in plastic surgery[10, 11]. High quality, real time data is essential to monitor the safety of medical implants.

Characteristics of Big Data

Characteristics of Big data

Big data comprises of three key characteristics: volume, variety and velocity[12, 13]. Volume refers to the volume of data per "transaction". This has gone from bytes (e.g traditionally recording observations every 6 hours), to kilobytes (e.g clinic letters in Electronic Health Records), to megabytes (e.g clinical photographs), to gigabytes (e.g CT & MRI scans, genomic sequencing). Subsequently, large volumes of data are being accumulated; in 2011 data from the U.S healthcare system reached 150 exabytes (161061273600 gigabytes)[14]. Storage and manipulation of such volumes of data are being facilitated by advances in data management, particularly in cloud computing and large-scale data centres [3, 12], as well as the reducing cost of the processors and storage media to analyse and store the data.

The variety of data available makes health analytics both exciting and challenging. Traditionally, it has been structured data (known as a Data Warehouse) that has been recorded electronically. Structured data (patient names, diagnostic codes, traditional databases etc) is standardised, pre-defined data that can easily be stored, manipulated and analysed[15]. In contrast, over 80% of data collected in the healthcare setting is in an unstructured format (known as a Data Lake) and on paper [12, 16]. The development of Natural Language Processing (NLP) software, which is able to read unstructured data and pick out specific aspects, has enabled diverse datasets of unstructured data such as operation and inpatient clinical notes linked to diagnosis and outcomes[17], prescriptions, radiology, pathology reports and primary care attendances to be critically analysed[18]. In addition, new data streams of both structured and unstructured data, are cascading into health care. Information from wearable devices (fitness devices, ECGs, Glucose monitoring in diabetics)[19] genetics and genomics[20], clinical trial data, social media feeds and smartphone applications are all examples of further sources of data becoming available.

Velocity, refers to the speed in which processes occur to make data available for use. Big data offers the ability for data to be processed in real-time. Traditionally, health care data has been mostly static, however real time accumulation of data has significant advantages, particularly in the monitoring of vital signs and blood results. This dimension of big data allows research to be undertaken at an unprecedented speed as compared to clinical trials[18].

Big data on its own is just the dataset; it is the analytics of such data that will have the greatest repercussions to research and clinical practice. As technology has developed for the storage of such data, so has the ability to analyse it. Using a form of artificial intelligence, machine learning involves generating algorithms capable of knowledge acquisition from historical examples [21]. Machine learning can be applied in a number of ways to improve clinical practice.

In this paper we provide an overview of big data research in surgery, the current resources available and its emerging role in research and clinical practice. We highlight how the successful use of this data can protect patients, improve outcomes and plan service delivery. To maintain balance, we also explore the limitations and fears raised with big data studies, offer some suggestions regarding how to optimise database creation, and suggest future directions for research in this exciting field.

Current Sources of Big data

Databases and Registries

There are a number of international databases and registries that would enable big data research in surgery. Whilst the creation of large databases and mining them for data is not innovative, it is the linkage of such data with Electronic Health Records that can provide novel insights. An advantage of utilising data from registries is the type of data stored such as information on medical devices and Patient Reported Outcome Measures (PROMs). These are often incorporated into disease specific databases, such as Breast-Q[22] in the Tracking Operations and Outcomes for Plastic Surgeons Program (TOPs)[23].

Electronic Health Records

Patient records have become increasingly computerised in both the primary and secondary settings to facilitate communication and improve patient care. Within England, The NHS aims to make patients' records entirely paperless by 2020[24]. As health informatics infrastructure has improved, the volume and detail of data available has grown exponentially. Information such as operation notes, laboratory and imaging data are now readily available digitally. In the United Kingdom, there are a number of databanks containing Electronic Health Records Available for analysis.

In Wales, The Secure Anonymised Information Linkage Databank (SAIL), a repository for a national collection of person-based anonymised health and socio-economic administrative data, was developed in 2007[25]. The databank incorporates a number of national databases among which are emergency department documentation, outpatient letters and primary care data. Similar databanks exist in England (The Clinical Practice Research Datalink[26]) and Scotland (Information Services Division[27]).

Clinical trial data

With the evolution of evidence-based medicine, a large number of clinical trials have been conducted in plastic surgery. Extensive data is collected often to answer a single research question; sharing of this data would allow secondary analysis. The concept of "open data" is becoming an increasing requirement for acceptance of a trial into medical journals[28]. Secondary analysis of such data would not only allow pooling of data, in the form of a meta-analysis but would facilitate novel research [29]. Through this approach, it has been discovered that tricyclic antidepressants could be efficacious against neuroendocrine tumours [30].

Social Media and web searches

The increased usage of social media by both patients and healthcare professionals provide data that can be utilised in both research and clinical practice. Analysis of twitter feeds for instance was two weeks faster at detecting the outbreak of cholera following the 2010 earthquake in Haiti than official reports [31]. In a similar fashion data from search terms on internet search engines was found to predict surges in emergency admissions from influenza [32].

Genetic Data

Following the Human Genome Project, the cost of genetic sequencing has improved significantly. The UK Government aims to be the first country in the world to sequence 100,000 whole human genomes for the treatment of patients with rare/inherited diseases or common cancers[20, 33].

Potential for Big data in Surgical Research

Advantages over traditional research methodologies

Big data research provides several advantages compared to traditional research methodologies [6, 34, 35]. The large volume of patient data accessible from big data can overcome some of the difficulties in recruiting patients with rare conditions and where there is significant heterogeneity in practice [35]. The size of such databases can also improve statistical power and reduce the risk of type II errors.

Patients recruited in prospective trials are often subject to selection bias, (eg: healthier and more compliant[36]) therefore reducing the translation into clinical practice. From an administrative perspective, big data analysis can be performed at a much faster rate, that is more cost efficient than clinical trials[12]. In fact using deep learning (artificial intelligence) algorithms, data can be autonomously mined to search for statistical associations and relationships.

Surgical Outcomes

The most powerful use of big data would be to inform patients and professionals alike of the accurate prevalence regarding post-operative complications, average length of stay, patient re-admissions, Venous Thromboembolism risks etc. To date, the majority of research incorporating big data has been published in this field.

Service planning

The comparison of outcomes between geographical regions would inform service planners, politicians and clinicians on best practices that could improve patient care. In the United

Kingdom, Professor Tim Briggs has set up the Getting It Right First Time (GIRST) programme[37]. The programme aims to save the National Health Service (NHS) £1.4 billion by using big data analysis to eliminate waste. In its recent review of general surgical patients, the average Length of Stay post appendicectomy was shown to be 3.5 days, however nearly half of trusts included discharged patients after two days. The report concludes that more than 30,000 bed days could be saved if all hospitals used the same approach and managed to meet the 'best' performing hospitals in such a metric. Whilst such results should be taken with caution, the role of big data in highlighting disparities and inefficiencies in care could lead to significant cost reduction[38].

Public Health

On a population level, big data can inform public health. The incidence of a variety of diseases such as skin cancer and trauma can be closely monitored and allow changing risk factors to be accurately identified. In plastic surgery, the international Burns Injury Database (iBID) has recently been used to identify detailed demographics of burn injuries within England and Wales, allowing targeted prevention strategies [39]. However traditional public health databases are not high volume, high variety or high velocity. In addition to structured data from databases, unstructured data such as, photographs, clinical notes and smartphone applications, would improve understanding of health behaviour allowing targeted prevention strategies.

Litigation

Big data can be effectively used in the field of litigation, to identify common causes of litigation and thus try to reduce these, along with hopefully reducing overall expenditure through reduced claims. Targeted educational courses, on common litigation pitfalls, could provide medical indemnity companies assurance of continued professional development in a similar approach to the National Speed Awareness Course aimed at motorists convicted of speeding in England and Wales[40].

Governance

Arguable one of the greatest benefits of big data within healthcare has been the monitoring of drugs and devices[41]. Safety information relating to medical devices is often confined to small trials in patients that are often not representative of the population[12, 42]. For example, the Non-Steroidal Anti-Inflammatory Refecoxib (Vioxx) was withdrawn after big data analysis identified a large number of patients sustained a cardiac arrest after the drug was commenced[43].

In this context, big data clearly has a significant role in clinical governance. For the PIP report, data was collected retrospectively and this delayed analysis and publication of the report. Real-time analysis of a robust prospective dataset would have enabled stricter regulation of the implants and more rapid identification of those at risk.

Data Driven Clinical Support Tools.

Whilst digitization of medical literature has greatly improved access, many surgeons struggle to keep up to date with the latest evidence. Even if all the relevant data were available, analysing it to develop a reasonable manageable plan for patients with multiple co-morbidities can be challenging. Data driven clinical decision support tools drawn from real time data analysis could be of immense benefit. IBM Watson for Oncology, a tool developed to help diagnose and propose treatment options for patients with cancer, is one such example[44]. Just as purchasers receive recommendations from retailers such as Amazon, this approach could inform clinicians of diagnostic and management choices made by other clinicians facing similar patient demographics. The acquisition of greater volumes of data and advances in technology are required to expand this field of research.

Data Analysis

The main components of big data processing include; data management, analysis and the production of suitable outputs which have been clearly visualised for their intended audience[18]. Clearly this involves close communication between data scientists, researchers and clinicians to understand the clinical setting and domain involved, as well as the technical requirements to match and assign valid outputs [12, 45]

Once data is extracted it is presented in a format to allow for analysis. Conceptually, big data can be analysed in a similar approach to traditional data using statistical software packages, however with increasing scale and scope of big data there is a need for larger and more sophisticated infrastructures to support these needs.

Privacy, ethics and security in big data

The benefits of big data are undeniable. However, as its use increases issues surrounding privacy, ethics and security in big data need addressing. There have already been a number of high profile concerns, such as security of electronic data in the NHS[46] and the 'inappropriate transfer' of patient data between the Royal Free Hospital and Google Deepmind[47].

Privacy is an important issue, especially with sensitive healthcare related information. While traditionally data has been de-identified using techniques such as anonymization, pseudonymisation and encryption it is increasingly realised that this may still not be enough to ensure privacy. It has been demonstrated on a number of occasions that de-identified data can be re-identified[48]. Furthermore, when data is multiply-linked and shared across the world, along with the integration of previously un-recognised data, the process of privacy protection becomes even harder. For example, the linking process may not only be to other healthcare data, but could also link healthcare to spending habits, travel destinations or exercise routines, increasing the amount of data exposed if privacy was broken.

Currently privacy and data protection laws are based around an individual's control over their information and on principles such as data minimization and purpose limitation. This however does not always fit that easily with the principles of maximizing available data and searching for patterns or associations that is common to big data analysis. This dichotomy is not helped by outdated laws and regulations, which are unable to keep up with the fast moving nature of computer science and healthcare innovation. In the UK the Data Protection Act 1996 and the Human Rights Act 1998 largely cover the area of big data, however there are considerable weaknesses. In May 2018, the European Union will implement the General Data Protection Regulation (GDPR) to strengthen data protection for all individuals in the EU [49]. The Regulation deems health related data as special categories of sensitive data subject to increased regulations. Researchers can use this data without consent providing appropriate safeguards are in place and that it is permitted under EU law [50].

Limitations of Big Data

Data Quality -Medical big data deals with data collected for different purposes, such as patient care in EHRs, inherently these data will have missing values[6]. Assuring quality of data will be essential as big data develops.

Coding of Data-Diagnosis and procedures may be documented using ICD codes or via Office of Population Censuses and Surveys Classification of Interventions and Procedures (OPCS-4) procedural codes. The knowledge of such coding systems amongst surgeons is limited making the quality of data inputted questionable. A study in St. George's Hospital, London demonstrated that 32% of operations were incorrectly coded resulting in an estimated financial loss of £17,000 from incorrect tariffs[51].

Electronic Health Records-The USA has been ahead of UK in this regard as administrative cultures lend themselves to capturing data in a common framework or language. In England, there has been high profile failures in creating EHR; notably NHS Care.data and NHS National Programme for Information Technology (NPfIT)[52, 53]. The down fall of such programmes were multifactorial however, the lack of collaboration between stake holders were common themes[53, 54]. In order to address such shortcomings there needs to be open discussion between policymakers, data scientists, clinicians and the general public. The general public need to be educated on the value of big data and shown that safeguards can be put in place to maintain their privacy and security as well as not feeling like their data is being monetized without them personally benefiting. Once the risks of privacy breaches, practical difficulties with obtaining consent for all uses and the potential benefits to individuals and society are set forth it is likely that there will be a better understanding and acceptance of the use of people's anonymised data.

The Future of Big Data

Big data analytics in surgery is currently in a nascent stage, however, this will not be for long. With advancement in technology, the use of Electronic Health Records and databases will rapidly progress not only within health organisations but also in the healthcare industry. Interdisciplinary collaboration between data scientists, researchers and clinicians is essential for this goal to be achieved[12].

Large patient databanks, such as SAIL,CPRD,and ISD are one such means of addressing some of the limitations associated with big data such as confidentiality. In addition to this, more disease-specific databases are required, giving the benefit of disease specific data including PROMs[55]. Device and drug registries need to become compulsory, to collect sufficient information to develop a basis for discussion on patient safety.

Conclusion

Big data analysis has the potential to improve surgical outcomes, protect patients and inform service providers. At present, the number of papers utilising big data is increasing. With this is mind, it is clear that the limitations surrounding big data analytics, such as confidentiality, need to be addressed.

Acknowledgements

There are no acknowledgements

Conflicts of Interests

The authors declare no conflicts of interest

References

1. de la Torre Diez I, Cosgaya HM, Garcia-Zapirain B, Lopez-Coronado M. Big Data in Health: a Literature Review from the Year 2005. J Med Syst. 2016;40(9):209.

2. NHS England. Five Year Forward View [Internet]. 2014. Available from: https://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf.

3. Kohn MS, Sun J, Knoop S, Shabo A, Carmeli B, Sow D, et al. IBM's Health Analytics and Clinical Decision Support. IMIA Yearbook. 2014:154-62.

4. British Broadcasting Corporation (BBC). Everyone 'to be research patient', says David Cameron [Internet]. 2011. Available from: http://www.bbc.co.uk/news/uk-16026827.

5. Medical Research Council. Initiatives in informatics research [Internet]. 2017. Available from: https://www.mrc.ac.uk/research/initiatives/health-and-biomedical-informatics/initiatives-in-informatics-research/.

6. Zhu VZ, Tuggle CT, Au AF. Promise and Limitations of Big Data Research in Plastic Surgery. Ann Plast Surg. 2016;76(4):453-8.

7. Wise J. NICE to ban mesh for vaginal wall prolapse. BMJ. 2017;359.

8. Carr A. Follow-up of patients with metal-on-metal hip replacements. BMJ. 2017;358.

9. Keogh SB. Poly Implant Prothèse (PIP) breast implants: final report of the Expert Group [Internet]. 2012. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/214975/dh_134657.pdf.

10. Kim B, Predmore ZS, Mattke S, van Busum K, Gidengil CA. Breast Implant–associated Anaplastic Large Cell Lymphoma: Updated Results from a Structured Expert Consultation Process. Plastic and Reconstructive Surgery Global Open. 2015;3(1):e296.

11. Ye X, Shokrollahi K, Rozen WM, Conyers R, Wright P, Kenner L, et al. Anaplastic large cell lymphoma (ALCL) and breast implants: breaking down the evidence. Mutation research Reviews in mutation research. 2014;762:123-32.

12. Gandomi A, Haider M. Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management. 2015;35(2):137-44.

13. Laney D. 3D Data Management: Controlling Data Volume, Velocity, and Variety. [Internet]. Meta Group Research; 2001. Available from: http://www.scirp.org/(S(351jmbntvnsjt1aadkposzje))/reference/ReferencesPapers.aspx?ReferenceID=1611280.

14. Institute for Health Technology Transformation. Transforming Health Care Through Big Data: Strategies for leveraging big data in the health care industry New York: 2013. Report No. Available from: http://c4fd63cb482ce6861463-bc6183f1c18e748a49b87a25911a0555.r93.cf2.rackcdn.com/iHT2_BigData_2013.pdf.

15. Koh HC, Tan G. Data mining applications in healthcare. J Healthc Inf Manag. 2005;19(2):64-72.

16. B F, Martin EM, Skotnes T. Big Data in Healthcare Hype and Hope [Internet]. Dr. Bonnie 3602012. Available from: http://www.west-info.eu/pdf/big-data-in-healthcare/.

17. Jordan SW, Mioton LM, Smetona J, Aggarwal A, Wang E, Dumanian GA, et al. Resident involvement and plastic surgery outcomes: an analysis of 10,356 patients from the American College of Surgeons National Surgical Quality Improvement Program database. Plast Reconstr Surg. 2013;131(4):763-73.

18. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Information Science and Systems. 2014;2:3.

19. Redmond SJ, Lovell NH, Yang GZ, Horsch A, Lukowicz P, Murrugarra L, et al. What Does Big Data Mean for Wearable Sensor Systems?: Contribution of the IMIA Wearable Sensors in Healthcare WG. Yearbook of Medical Informatics. 2014;9(1):135-42.

20. He KY, Ge D, He MM. Big Data Analytics for Genomic Medicine. International Journal of Molecular Sciences. 2017;18(2):412.

21. Kanevsky J, Corban J, Gaster R, Kanevsky A, Lin S, Gilardino M. Big Data and Machine Learning in Plastic Surgery: A New Frontier in Surgical Innovation. Plastic & Reconstructive Surgery. 2016;137(5):890e-7e.

22. Pusic AL, Klassen AF, Scott AM, Klok JA, Cordeiro PG, Cano SJ. Development of a new patient-reported outcome measure for breast surgery: the BREAST-Q. Plast Reconstr Surg. 2009;124(2):345-53.

23. Surgeons ASoP. Tracking Operations and Outcomes for Plastic Surgeons (TOPS), [Internet]. 2017. Available from: https://www.plasticsurgery.org/for-medical-professionals/quality-and-registries/tracking-operations-and-outcomes-for-plastic-surgeons.

24. NHS Trust Development Authority Five Year Forward View [Internet]. NHS England; 2014. Available from: http://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf

25. The Secure Anonymised Information Linkage Databank (SAIL). SAIL Databank [Internet]. Available from: https://saildatabank.com/.

26. Medicines and Healthcare Products Regulatory Agency (MHRA). Welcome to The Clinical Practice Research Datalink [Internet]. Available from: https://www.cprd.com/home/.

27. Information Services Division. Information Services Division [Internet]. Available from: http://www.isdscotland.org/.

28. Taichman DB, Sahni P, Pinborg A, Peiperl L, Laine C, James A, et al. Data sharing statements for clinical trials. BMJ. 2017;357.

29. Docherty AB, Lone NI. Exploiting big data for critical care research. Current Opinion in Critical Care. 2015;21(5):467-72.

30. Jahchan NS, Dudley JT, Mazur PK, Flores N, Yang D, Palmerton A, et al. A drug repositioning approach identifies tricyclic antidepressants as inhibitors of small cell lung cancer and other neuroendocrine tumors. Cancer Discov. 2013;3(12):1364-77.

31. Chunara R, Andrews JR, Brownstein JS. Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. Am J Trop Med Hyg. 2012;86(1):39-45.

32. Klembczyk JJ, Jalalpour M, Levin S, Washington RE, Pines JM, Rothman RE, et al. Google Flu Trends Spatial Variability Validated Against Emergency Department Influenza-Related Visits. J Med Internet Res. 2016;18(6):e175.

33. NHS England. Genomics [Internet]. 2007. Available from: https://www.england.nhs.uk/healthcare-science/personalisedmedicine/genomics/.

34. Taglang G, Jackson DB. Use of “big data” in drug discovery and clinical trials. Gynecologic Oncology. 2016;141(1):17-23.

35. Adkinson JM, Casale MT, Kim JY, Khavanin N, Gutowski KA, Gosain AK. So You Have a Research Idea: A Survey of Databases Available for Plastic Surgery Research. Plastic & Reconstructive Surgery. 2016;137(2):680-9.

36. Guller U. Surgical Outcomes Research Based on Administrative Data: Inferior or Complementary to Prospective Randomized Clinical Trials? World Journal of Surgery. 2006;30(3):255-66.

37. Getting It Right First Time (GIRFT). What we do [Internet]. 2017 [cited 2017]. Available from: http://gettingitrightfirsttime.co.uk/what-we-do/.

38. Abercrombie J. General Surgery: GIRFT Programme National Speciality Report [Internet]. 2017. Available from: http://gettingitrightfirsttime.co.uk/wp-content/uploads/2017/07/GIRFT-GeneralSurgeryReport-Aug17v1.pdf.

39. Stylianou N, Buchan I, Dunn KW. A review of the international Burn Injury Database (iBID) for England and Wales: descriptive analysis of burn injuries 2003-2011. BMJ Open. 2015;5(2):e006184.

40. The TTC Group. National Speed Awareness Course [Internet]. 2017. Available from: http://www.ttc-uk.com/police-referred-courses/national-speed-awareness-course/.

41. Bone A, Houck K. The benefits of data mining. Elife. 2017;6.

42. Labrinidis A, Jagadish HV. Challenges and opportunities with big data. Proc VLDB Endow. 2012;5(12):2032-3.

43. Krumholz HM, Ross JS, Presler AH, Egilman DS. What have we learnt from Vioxx? BMJ : British Medical Journal. 2007;334(7585):120-3.

44. IBM. Bringing confident decision-making to oncology [Internet]. 2017. Available from: https://www.ibm.com/watson/health/oncology-and-genomics/oncology/.

45. Meyfroidt G, Guiza F, Ramon J, Bruynooghe M. Machine learning techniques to examine large patient databases. Best Pract Res Clin Anaesthesiol. 2009;23(1):127-43.

46. Davies M. How hackers held the NHS to ransom [Internet]. 2017. Available from: http://www.newstatesman.com/politics/health/2017/05/how-hackers-held-nhs-ransom.

47. Hern A. Google DeepMind 1.6m patient record deal 'inappropriate' [Internet]. The Guardian2017. Available from: https://www.theguardian.com/technology/2017/may/16/google-deepmind-16m-patient-record-deal-inappropriate-data-guardian-royal-free.

48. Ohm P. Broken promises of Privacy: Responding to the Surprising Failure of Anonymization. UCLA Law Review. 2010;57:1701-77.

49. Privacy Shield. Privacy Shield [Internet]. 2017. Available from: https://www.privacyshield.gov/welcome

50. Dove ES, Thompson B, Knoppers BM. A step forward for data protection and biomedical research. The Lancet. 2016;387:1374-5.

51. Martinou E, Shouls G, Betambeau N. Improving the accuracy of operation coding in surgical discharge summaries. BMJ Quality Improvement Reports. 2014;3(1).

52. Godlee F. What can we salvage from care.data? BMJ. 2016;354.

53. Campion-Awwad O HA, Smith L, Vuaran M. The National Programme for IT in the NHS: A Case History: University of Cambridge; 2014.

54. Review of Data Security, Consent and Opt-Outs, (2016).

55. Black N. High-quality clinical databases: breaking down barriers. Lancet. 1999;353(9160):1205-6.