TITLE PAGE

Title: Can the NHS be a learning healthcare system in the age of digital technology?

Author names and affiliations:

Amitava Banerjee[1] ami.banerjee@ucl.ac.uk

Lydia N Drumright[2] lnd23@medschl.cam.ac.uk

Andrew R J Mitchell[3] mail@jerseycardiologist.com

[1]Farr Institute of Health Informatics Research, University College London, London NW1 2DA

[2]Department of Medicine, University of Cambridge, Cambridge CB2 0QQ

[3] Jersey General Hospital, Gloucester Street, St Helier, Jersey JE1 3QS

Corresponding author: Dr Amitava Banerjee

Present address: Farr Institute of Health Informatics Research

222 Euston Road

London

UK

NW1 2DA

Tel: 02035495449

Email: ami.banerjee@ucl.ac.uk

Word count: 2084 words (excluding figure legends and references)

References:30

Can the NHS be a learning healthcare system in the age of digital technology?

Abstract

Digital health technology, and predictive analytics can disrupt healthcare delivery. From direct-to-consumer genetic testing to mobile phone-based apps for prediction of heart attacks, the impact of big data is forecast as transformative for our ability to diagnose and treat our patients with personalised healthcare. With this abundance of opportunity, are healthcare systems ready for big data? The definition and scope of 'big data' remains unclear. The UK's National Health Service (NHS) is an example of a system caught between the promises and pitfalls of big data for healthcare and a deeper investigation may inform debate and implementation in other countries. In his recent review of health information technology in England, Professor Bob Wachter issued a reality check to the NHS:

"While there is great enthusiasm for using 'big data' to develop personalised approaches for individual patients ('precision medicine'), provide customised decision support to both clinicians and patients, and create 'learning healthcare systems (LHS)', today all these goals are more promise than reality."(1)

What is the gap between promise and reality in the NHS? Can the NHS face these challenges and deliver on the promises of big data? What will it take to develop LHS and provide better care? Can patient trust be gained to allow physicians and policymakers to deliver the potential of big data in a patient-centred healthcare model? Concomitance of "evidence-based healthcare", "right care" and "learning health systems" frameworks are required to drive high-quality, data-driven healthcare.

Introduction

"Big data" is defined by the "7-V's": volume (most frequently cited(2)), velocity, veracity, variety, volatility, validity and value. In healthcare, "big data" is associated with a step-change in the way information is gathered, analysed and used to facilitate disease management and prevention. With greater electronic data capture, there is enthusiasm for increased safety, efficiency and effectiveness in health and social care through machine learning, artificial intelligence(AI) and augmented intelligence. However, factors maintaining and widening the gap between the promise and the reality need to be addressed.

Can 'big' be evidence-based?

Current best practice has its foundation in evidence-based healthcare, with growth in publications, but poorly managed scientific insights, poor recording of care and poor use of evidence(3). Big data could improve the status quo and support learning health systems(4).

Computational methods can contribute to evidence management with automation of literature searching, critical appraisal and guidelines(5, 6). Similarly, big data already contributes to aetiologic, diagnostic, prognostic and therapeutic research, from -omics to electronic health records (EHR) trials(7). Critics emphasise lack of quality and validation of routinely collected clinical data, and risk of bias in observational studies, where scale cannot compensate for poor design(8). Conversely, data-driven approaches could transform a predominantly retrospective into a prospective or real-time paradigm, across disease boundaries.

Infrastructure and analytic tools are necessary but often poorly understood and underdeveloped. Automated extraction of necessary data fields in pseudonymised/anonymised format into curated warehouses is required with robust metadata catalogues and understanding of clinical context. Extraction, cleaning and processing of data in real-time are prerequisites for point-of-care information available to clinicians.

Preventing excesses and addressing deficiencies

Medicine grapples with conflicting challenges of over-diagnosis and over-treatment, as well as under-diagnosis and under-treatment. Widening inequalities also manifest across many health systems, despite improving medications and technologies. Efforts to optimise healthcare delivery have culminated in "right care"(9), which may be facilitated by big data methods and outputs.

Decision aids could help clinicians to optimise patient management, supported by large-scale EHR data. For example, among individuals who meet "appropriateness of use criteria" (AUC) for diagnostic coronary angiography, only 52.9% have obstructive coronary disease(10). Similarly, among elective surgery patients, routine tests of blood coagulation are unnecessary in over 94% of cases, with implications for cost and false positives(11). It is plausible that AUC for tests could be continuously evaluated using EHR, quality improvement initiatives and new trials. In this way, "an ideal system would be evidence-based, use uniform and comprehensive clinical data, provide point-of-care decision support, and aim to improve quality by reducing overuse and underuse"(12). The medical 'Internet of Things' describes the growing number of apps and devices which, through big data analytics, are linking patients and health professionals(13). An example of its use in tackling under-diagnosis is in screening and diagnosis of atrial fibrillation, although proper evaluation is required(14).

Similarly, AI, the mimicking of human cognition by computers, could change data use, analysis and interpretation in radiology, pathology, oncology and several other specialties(15). For example, AI may predict the grade and stage of lung cancer better than pathologists(16). A barrier to adoption may stem from fear of job loss and concerns around accountability, but people may be freed to work on the aspects which only humans can conduct(15). Machine learning has promise, particularly in tackling over- and under-diagnosis (17), with advantages of increased efficiency, reproducibility and scalability, e.g. diabetic retinopathy(18).

Risk prediction is of direct relevance in the era of personalised medicine. Both AI and machine learning aid not only in prediction of real-time risk, but also in the visualisation and communication of those risks. Risk scores can be automated within EHRs and multiple risk scores may be deployed in the same patient simultaneously, whether in primary care or intensive care. Moreover, risk prediction can be incorporated into data-driven management such as personalised radiotherapy for prostate cancer. However, there are only a handful of exemplars(19-21) and models that appear efficacious are yet to be applied(22, 23).

As Wachter says, there is currently "more promise than reality" and more hype than evidence. For example, intelligence augmentation is currently receiving much attention, but before widespread adoption, not only is evidence required; the pace and scale at which the research is conducted must keep up, i.e. big data methods are required in evidence generation(24). Moreover, there are primary research barriers related to implementing AI and machine learning into medicine, including: appropriate and secure access to patient data; platforms to collate personal and healthcare data, and linkage required to develop the models; missing, incomplete and unvalidated data(25); or development of interfaces to place predictive algorithms within EHR.

There are scenarios where big data and associated issues are compounding rather than solving problems. For example, direct-to-consumer genetic testing cannot yet provide genetic information which is useful to patients or clinicians for most complex diseases and may create more anxiety than health. Moreover, lack of integration of such services into disease management pathways may lead to high costs, confusion and inappropriate use of resources. The rise in patient use of wearables is currently not related with improved patient outcomes and may increase pressures on health services(26), thus adding to over-diagnosis and over-treatment. Inattention to "digital divides" may worsen health inequalities.

Big data, machine learning and AI all need to be used more intelligently for the individual and the health system. There is a trend towards piecemeal use of big data, which may partly be due to patchy availability of EHR and digital technologies across the NHS or patchy infrastructure for harnessing these data. The direction of travel should be throughout a patient pathway if use is to become widespread or universal.

This will require appropriate, secure, linked infrastructure that can collate, link, protect, anonymise and present data to the analyst for research, with capacity to be implemented back into the healthcare setting.

Learning healthcare systems

Health informatics describes the best use, albeit promise, of (big) data to improve health at the individual and population level, health care and biomedical research. Recognition of big data in healthcare has led to many related terms/disciplines, including “digital medicine”, “precision medicine” and “personalised medicine”, which overlap and inter-connect with different branches of informatics (Figure 1), but may confuse discussions. Taken collectively, health informatics would support a “learning healthcare system”, a framework proposed by the National Institute of Medicine in 2006 to capture the possibility of data from science, evidence and care feeding into one another in a “virtuous” cycle (3)(Figure 2). In his review, Wachter’s comments are timely and far-sighted:

“In particular, the data processing, analytics, and informatics research workforce must not be forgotten, particularly since the UK has an enormous opportunity to bring the promise of big data to life. Working through ambitious entities such as the Farr Institute, the UK could be the first nation to take data science to scale and build a truly national learning health system”(1).

However, “high-level” terminology stifles operationalisation. Many consider the LHS to be a “platform” of infrastructure, human resources, software and other inter-related components, and “learning health cycles” to describe the science-evidence-care relationship. Regardless, the wastage throughout the pathway from science through evidence to delivery of care is matched by strains on resources in every sector, from human and infrastructural to financial and managerial. Big data methods can aid in organisation of healthcare and public health services as well as research, education and training, but at least five major paradigm shifts are required at system level.

First, research and “service” remain two parallel and relatively unconnected activities in the NHS, despite the work of NIHR and other funders to create a research-active clinical environment throughout the NHS. An important consequence is under-performance by many hospitals in terms of research. Lack of academic activity in hospitals was correlated with poor quality care in the review of under-performing hospitals conducted by Bruce Keogh(27) and was also highlighted by the Francis report(28). The long timescale from research to impact as well as perceived and real gaps in translational research which truly changes acute care design and delivery are also barriers to research-driven care. In LHS, the provenance of data at the point-of-care encourages patient-centred research and patient-centred clinical care at the same time. On the other hand, the overall possibilities of precision medicine, whether in terms of diagnostics or therapeutics, are likely to be tested and used to their full potential in a patient-centred and data-centred environment, rather than a researcher-centric model.

Second, quality improvement and audit have been generally viewed as separate to clinical research and the overlap between the two is sub-optimally managed. LHS can make research scaleable and quality improvement a real-time activity. Issues such as workforce planning and predicting bed capacity are examples of particular needs where reframing as research questions may be of mutual benefit to clinicians and academics. Audit at scale could be continuous and go beyond government performance targets. There would be implications for development of the optimal metrics for patient outcomes, monitoring of burden of disease and service need, optimisation of care pathways, as well as design and implementation of EHR-based trials to test service-level interventions.

Third, the capacity in health informatics among clinicians and non-clinicians is a rate-limiting step. Undergraduate and postgraduate curricula are currently largely ignoring the impact of big data on healthcare delivery, and therefore the future workforce may not be adequately prepared for data-driven healthcare. Leadership, participation and long-term vision at the national and local level are required from both academia (e.g. Farr Institute of Health Informatics Research, independent health informatics researchers) and the NHS (e.g. the Chief Clinical Informatics Officer, CCIO, network) to ensure that the rich and overlapping tapestry of quality improvement, implementation science, translational research and audit can be realised through LHS. Within the NHS, extra efforts to support co-creation of capacity between academic and NHS sectors will be critical to keep up with University-owned hospitals internationally.

Fourth, the trust of patients and the public in the use of their data is central to developing a LHS where routine health data are optimally used. There have been historic challenges in developing large-scale IT programmes in the NHS, whether Connecting for Health, or more recently, care.data, and lack of public trust in secondary use of data played a significant role(29). Currently, Google DeepMind is working with the Royal Free Hospitals NHS Trust to develop risk prediction tools for acute kidney injury, but there have been concerns from regulatory bodies, clinicians and public regarding ethical and data security implications for this research model. A new framework for public understanding of science and trust, with involvement and co-creation at the centre, is required to make sure that these concerns are allayed before implementation so that research, innovation and care delivery are not prevented(30).

Fifth, investment and engagement from patient-facing healthcare professionals (HCP) is essential for maximising LHS. Assessment of data can only be as good as the data used. Continuous feedback and improvement in how data are recorded, captured and standardised is necessary to ensure that the results of assessments provide the most pertinent information for continuous improvement. This will only come from sharing evaluations with HCP and discussing where, how, and why inconsistencies and missing data persist and empowering HCP to take responsibility for continuous improvement in both healthcare provision and the recording of that provision.

These five and other barriers to LHS implementation in the NHS are highlighted in Box 1. Scale-up is limited by a lack of examples of LHS in the published literature and therefore, publication, education and dissemination regarding successful projects are crucial, especially from within the NHS (Box 2), rather than trying to emulate examples from other health systems which may not be completely transferable or relevant to the NHS and the UK. The costs of not implementing LHS in terms of human resources, patient safety and patient-centred outcomes have not been quantified in existing research and represent a powerful tool for engaging and convincing policymakers.

Conclusions

The learning health system is a natural progression if the full potential of the wealth of big data throughout the health sector is to be realised. Clinicians, academics, citizens and private sector can align to develop a patient-centred, data-driven model of care but multiple barriers exist in the NHS at present, including infrastructure, workforce, training and culture. Trust and evidence will be important levers for patients, public and health professionals and need be addressed urgently, along with the other barriers in the NHS. Without the learning health system approach, big data is likely to remain an entity used in discrete projects rather than across the NHS.

Box 2. Examples of real world opportunities for learning healthcare systems in the NHS

The 100 000 Genomes Project: Genomics England are sequencing 100,000 genomes from 70,000 people with a rare disease, plus their families, as well as patients with cancer. A new genomic medicine service for the NHS is being delivered via this project. https://www.genomicsengland.co.uk/the-100000-genomes-project/

Connected Care in Berkshire: Using Graphnet's CareCentric™ software, 17 health and social care organisations across Berkshire will share the care records of 855,000 patients, enabling shared EHR which will allow care professionals to create and update care plans for patients. http://www.digitalhealth.net/shared_care_records/47246/berkshire-to-share-patient-records-with-graphnet

East London Genes and Health: A long-term study of 100,000 people of Bangladeshi and Pakistani origin, linking genomic data with EHR to study disease and treatments. http://www.genesandhealth.org/

Hospital Electronic Prescribing and Medicines Administration systems in Scotland: The implementation of e-prescribing and medicines administrations will enable improvements in care and in the secondary use of data for research. Comprehensive stakeholder analysis has led to early identification of barriers and facilitators to a learning health system approach. https://www.ncbi.nlm.nih.gov/pubmed/28059691

Royal Free NHS Trust/Google DeepMind collaboration:
Artificial intelligence company, Google Deepmind and the Royal Free NHS Trust are working together on real-time EHR data to develop apps for use by health professional in the early detection of patient deterioration. http://www.bbc.co.uk/news/health-38055509

Key messages

- Big data has big potential
- The definition and the focus of big data need to be clear
- Evidence needs to be gathered before and after data-driven interventions in health
- Research should be embedded within healthcare continuously

Acknowledgments

This manuscript was influenced by the outcomes of three workshops which AB designed and coordinated on the theme of big data in healthcare: (i) “Big data and Evidence-Based Mediciine-the promises and the pitfalls”at Evidence Live, Oxford in June 2016; (ii) “Learning health systems-the only way to do translational data science?” at the International Population Data Linkage Network (IPDLN) Conference, Swansea in August 2016; and “Big data-part of the problem of over-diagnosis or part of the solution?” at the Preventing Overdiagnosis Conference, Barcelona in September 2016.

Contributorship statement

AB produced the original draft manuscript and all authors were responsible for revisions and the final version.

Competing interests

AB has received honoraria from Novo Nordisk and Boehringer Ingelheim. The other authors have no competing interests to report.

Funding

No funding was necessary for this work.

References

1. UK Department of Health. Making IT work: harnessing the power of health information technology to improve care in England. UK Department of Health; 2016 07/09/2016.
2. Baro E, Degoul S, Beuscart R, Chazard E. Toward a Literature-Driven Definition of Big Data in Healthcare. Biomed Res Int. 2015;2015:639021.
3. In: Olsen LA, Aisner D, McGinnis JM, editors. The Learning Healthcare System: Workshop Summary. The National Academies Collection: Reports funded by National Institutes of Health. Washington (DC)2007.
4. Heneghan C, Godlee F. Where next for evidence based healthcare? BMJ. 2013;346:f766.
5. Toga AW, Foster I, Kesselman C, Madduri R, Chard K, Deutsch EW, et al. Big biomedical data as the key resource for discovery science. J Am Med Inform Assn. 2015;22(6):1126-31.
6. Elliott J, Sim I, Thomas J, Owens N, Dooley G, Riis J, et al. #CochraneTech: technology and the future of systematic reviews. Cochrane Db Syst Rev. 2014(9).
7. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst. 2014;2:3.
8. Khoury MJ, Ioannidis JP. Medicine. Big data meets public health. Science. 2014;346(6213):1054-5.
9. Elshaug AG, Rosenthal MB, Lavis JN, Brownlee S, Schmidt H, Nagpal S, et al. Levers for addressing medical underuse and overuse: achieving high-value health care. Lancet. 2017.
10. Mohareb MM, Qiu F, Cantor WJ, Kingsbury KJ, Ko DT, Wijeysundera HC. Validation of the appropriate use criteria for coronary angiography: a cohort study. Ann Intern Med. 2015;162(8):549-56.
11. Capoor MN, Stonemetz JL, Baird JC, Ahmed FS, Awan A, Birkenmaier C, et al. Prothrombin Time and Activated Partial Thromboplastin Time Testing: A Comparative Effectiveness Study in a Million-Patient Sample. PLoS One. 2015;10(8):e0133317.
12. Doll JA, Patel MR. Self-regulation in the era of big data: appropriate use of appropriate use criteria. Ann Intern Med. 2015;162(8):592-3.
13. Dimitrov DV. Medical Internet of Things and Big Data in Healthcare. Healthc Inform Res. 2016;22(3):156-63.
14. Cifu A, Prasad V. Wearables, Smartphones and Novel Anticoagulants: We Will Treat More Atrial Fibrillation, but Will Patients Be Better Off? J Gen Intern Med. 2016;31(11):1367-8.
15. Jha S, Topol EJ. Adapting to Artificial Intelligence: Radiologists and Pathologists as Information Specialists. JAMA. 2016;316(22):2353-4.
16. Yu KH, Zhang C, Berry GJ, Altman RB, Re C, Rubin DL, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. Nat Commun. 2016;7:12474.
17. Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. Artif Intell Med. 2001;23(1):89-109.
18. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA. 2016;316(22):2402-10.
19. Brady TM, Neu AM, Miller ER, 3rd, Appel LJ, Siberry GK, Solomon BS. Real-time electronic medical record alerts increase high blood pressure recognition in children. Clin Pediatr (Phila). 2015;54(7):667-75.
20. Hye RJ, Smith AE, Wong GH, Vansomphone SS, Scott RD, Kanter MH. Leveraging the electronic medical record to implement an abdominal aortic aneurysm screening program. J Vasc Surg. 2014;59(6):1535-42.
21. Motwani M, Dey D, Berman DS, Germano G, Achenbach S, Al-Mallah MH, et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. Eur Heart J. 2017;38(7):500-7.

22. Sinha N, Dauwels J, Kaiser M, Cash SS, Westover MB, Wang YJ, et al. Predicting neurosurgical outcomes in focal epilepsy patients using computational modelling. Brain. 2017;140:319-32.
23. Kipnis P, Turk BJ, Wulf DA, LaGuardia JC, Liu V, Churpek MM, et al. Development and validation of an electronic medical record-based alert score for detection of inpatient deterioration outside the ICU. J Biomed Inform. 2016;64:10-9.
24. Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. N Engl J Med. 2016;375(13):1216-9.
25. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JP. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. J Am Med Inform Assoc. 2017;24(1):198-208.
26. NPR. Sure You Can Track Your Health Data, But Can Your Doctor Use It? 2015 [Available from: http://www.npr.org/sections/health-shots/2015/01/19/377486437/sure-you-can-track-your-health-data-but-can-your-doctor-use-it.
27. Keogh B. Review into the quality of care and treatment provided by 14 hospital trusts in England: overview report. . 2013.
28. The Mid Staffordshire NHS Foundation Trust Public Inquiry. Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry. 2013.
29. Hays R, Daker-White G. The care.data consensus? A qualitative analysis of opinions expressed on Twitter. BMC Public Health. 2015;15:838.
30. Chan T, Di Iorio CT, De Lusignan S, Lo Russo D, Kuziemsky C, Liaw ST. UK National Data Guardian for Health and Care's Review of Data Security: Trust, better security and opt-outs. J Innov Health Inform. 2016;23(3):627-32.