

Leading article

Learning health systems need to bridge the 'two cultures' of clinical informatics and data science

Philip J. Scott

Centre for Healthcare Modelling and Informatics, School of Computing, University of Portsmouth, Portsmouth, UK

Rachel Dunscombe

Chief Executive, NHS Digital Academy Salford Royal NHS Foundation Trust, Salford, UK

David Evans

BCS, The Chartered Institute for IT, Swindon, UK

Mome Mukherjee

Usher Institute of Population Health Sciences and Informatics, The University of Edinburgh, Edinburgh, UK

Jeremy C. Wyatt

Wessex Institute, University of Southampton, Southampton, UK

Author address for correspondence:

Philip J. Scott Senior Lecturer Centre for Healthcare Modelling and Informatics School of Computing, University of Portsmouth Portsmouth, UK

Cite this article: Scott PJ, Dunscombe R, Evans D,

informatics and data science. J Innov Health Inform.

Mukherjee M, Wyatt J. Learning health systems

need to bridge the 'two cultures' of clinical

http://dx.doi.org/10.14236/jhi.v25i2.1062

Copyright © 2018 The Author(s). Published by BCS,

license http://creativecommons.org/licenses/by/4.0/

The Chartered Institute for IT under Creative Commons

2018;25(2):126-131.

Accepted May 2018

Email: Philip.scott@port.ac.uk

ABSTRACT

Background UK health research policy and plans for population health management are predicated upon transformative knowledge discovery from operational 'Big Data'. Learning health systems require not only data, but feedback loops of knowledge into changed practice. This depends on knowledge management and application, which in turn depends upon effective system design and implementation. Biomedical informatics is the interdisciplinary field at the intersection of health science, social science and information science and technology that spans this entire scope.

Issues In the UK, the separate worlds of health data science (bioinformatics, 'Big Data') and effective healthcare system design and implementation (clinical informatics, 'Digital Health') have operated as 'two cultures'. Much National Health Service and social care data is of very poor quality. Substantial research funding is wasted on 'data cleansing' or by producing very weak evidence. There is not yet a sufficiently powerful professional community or evidence base of best practice to influence the practitioner community or the digital health industry.

Recommendation The UK needs increased clinical informatics research and education capacity and capability at much greater scale and ambition to be able to meet policy expectations, address the fundamental gaps in the discipline's evidence base and mitigate the absence of regulation. Independent evaluation of digital health interventions should be the norm, not the exception.

Conclusions Policy makers and research funders need to acknowledge the existing gap between the 'two cultures' and recognise that the full social and economic benefits of digital health and data science can only be realised by accepting the interdisciplinary nature of biomedical informatics and supporting a significant expansion of clinical informatics capacity and capability.

Keywords: Big Data, health informatics, bioinformatics, evidence-based practice, health policy, programme evaluation, education, learning health systems

INTRODUCTION

C. P. Snow famously characterised the gulf between the 'two cultures' of science and the humanities as a serious barrier to progress. In our field, at least in the UK, there appears to be an analogous gap between the policy and funding programmes of data science (bioinformatics, 'Big Data') and effective system design and implementation (clinical informatics, 'Digital Health').

Data science in healthcare is subject to strong regulatory and ethical controls, minimum educational qualifications, well-established methodologies, mandatory professional accreditation and evidence-based independent scrutiny. By contrast, 'Digital Health' has minimal substantive regulation or ethical foundation, no specified educational requirements, weak methodologies, a contested evidence base and negligible peer scrutiny. Yet, the 'Big Data' vision is to base its science on the data routinely produced by digital health systems.

This paper is focussed on the UK context. We bring together experience from the frontline National Health Service (NHS) clinical informatics and epidemiological research to present the operational realities of health data quality and the implications for data science. We argue that to build a successful learning health system, data science and clinical informatics should be seen as two parts of the same discipline with a common mission. We commend the work in progress to bridge this cultural divide, but propose that the UK needs to expand its clinical informatics research and education capacity and capability at much greater scale to address the substantial gaps in the evidence base and to realise the anticipated societal aims.

ROUTINE CLINICAL DATA IS HIGHLY PROBLEMATIC

Data quality in the frontline health and care system faces a dual challenge in our current environment. First is the lack of standard data sets and adoption of reference values, though work is progressing in this area.2 The second is the lack of data quality due to unreliable adherence to process³ and poor system usability.⁴ Embarking on the implementation of clinical terminology including Systematised Nomenclature of Medicine Clinical Terms (SNOMED CT) and Logical Observation Identifiers Names and Codes (LOINC) shows us that our historical environment and the complexity of these standards always causes long debate and significant amounts of implementation effort. So far, little progress has been made even by the 'Global Digital Exemplars'5 in implementing SNOMED CT in any depth. Furthermore, complexity is introduced when interoperating with other care settings such as social care and mental health. GP data is far from consistent. Different practices will use different fields in different ways and usage varies from clinician to clinician. Historically, the system has not forced users to standardise their recording or practice. This results in varying data quality between GP practices, which affects not just epidemiological studies but operational processes. Failure to enter accurate data into health and care systems occurs for a number of reasons including poor usability, overly complex systems, lack of data input logic to check errors and poor business change leadership.

Most epidemiological research with routine clinical data uses coded data, rather than free text. Thus, there is over reliance on codes used during clinical consultations. A national evaluation of usage of codes in primary care in Scotland, taking allergy as an example, found that 50% usage in over 2 million consultations, over 7 years, were from eight codes used to report for an incentive programme for GPs, 95% usage was from 10% of the 352 allergy codes (n = 36) and 21% codes were never ever used.⁶ A systematic review found that there are variations in completeness (66%-96%) and correctness of morbidity recording across disease areas. For instance, the quality of recording in diabetes is better than asthma in primary care. There are also changes in case definition and diagnostic criteria across disease areas over time, which are seldom mentioned in the databases. A recent primary care study found that choice of codes can make a difference to outcome measures, for example, the incidence rate was found to be higher when non-diagnostic codes were used rather than with diagnostic codes.8 Since there is variability of coding of data across GP practices, when practices with poor quality of recording were included in the analysis, there was significant difference in incidence rate and trends, with lower incidence rate and decreasing trends when they were included. This study highlights the effect of miscoding and misclassification. It also shows that when data are missing, they might not be missing at random. Furthermore, there could be unavailability of codes that were needed during consultation and thus were recorded in free text. All these salient features around coding of data are often ignored when interrogating patient databases for research and thus could lead to erroneous conclusions. No amount of data cleansing could sort the inherent discrepancies involved in coded data.

There could be confounding by indication or severity, for example, when severely ill patients receive more intensive treatment and could have poor outcomes compared to other patients.9 Clinical databases only comprise patients who attended healthcare services. A UK-wide study showed the difference in asthma prevalence when asthma was reported from population surveys compared to clinical databases. 10 Besides quality of coded data, there could be lack of key variables in clinical databases, since their primary purpose was not designed for research, for example, the absence of diagnoses in outpatient hospital attendances.

Furthermore, significant variance is seen in the success of electronic patient record deployments from the same commercial vendor in different localities. For example, the Arch Collaborative from KLAS research¹¹ shows variance in all aspects of success including data quality of the deployments by Cerner, Epic and Allscripts. US experience has shown a particular risk from 'copy and paste' errors. 12

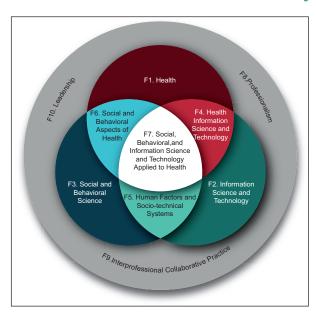


Figure 1 The intersecting knowledge and practice domains of biomedical informatics

BIOMEDICAL INFORMATICS IS AN INTERDISCIPLINARY FIELD WITH A **COMMON MISSION**

The 'two cultures' are both embraced by the widely adopted American Medical Informatics Association definition of biomedical informatics as: 'the interdisciplinary field that studies and pursues the effective uses of biomedical data, information, and knowledge for scientific inquiry, problem solving and decision making, motivated by efforts to improve human health'.13 Biomedical informatics can be visualised as the intersection of health science, social science and information science and technology (Figure 1, reproduced with permission from AMIA 14).

In this definition, biomedical informatics has sub-fields such as health informatics (comprising clinical and public health informatics) and bioinformatics (also called computational biology). Whereas bioinformatics deals with data science, clinical informatics 'covers the practice of informatics in healthcare' (emphasis added). Therefore, getting clinical informatics right is more about people than it is about technology or data. As Coiera said, informatics is 'as much about computers as cardiology is about stethoscopes'. 15

Of course, biomedical informatics must be aimed at a grand outcome - the betterment of health - rather than a contained body of knowledge or an abstract philosophy. The sole axis of interest is whether or not health is ultimately improved.

This has a number of implications. In pursuit of a better health outcome, a clinician may employ nuclear physics or big data analytics. Similarly, an informatician needs to be multi-disciplinary and citizen-centred as they play their part in a shared mission. Maintaining a system-wide view of outcomes is an ethical imperative for everyone involved, from research to application. 16

Treating the 'two cultures' within biomedical informatics as separate disciplines, rather than as a shared mission, may be professionally attractive and tractable for funders and policymakers, but risks maintaining silos and working against the public interest. Instead, biomedical informatics researchers and practitioners - including clinicians - need to be part of a single professional organism made of interlocking professional communities; able to work together in a single systemic view of citizen benefit and harm, and able to implement the best scientific, engineering and medical disciplines available. To do otherwise is simply unethical.

This ethical perspective opens up an exciting vista of fruitful, high impact, applied research and professional practice. Global health public policy is united in its view that digital systems, data and digital transformation are vital tools for the advancement of health and care. Learning health systems¹⁷ require not only the Big Data 'engine' but also the feedback loop of knowledge into changed practice. This crucially depends on knowledge management and application, which in turn depends on effective system design and implementation: clinical informatics. Figure 2 (adapted from Rouse et al. 18 originally based on ONC19) illustrates how much of the learning health system depends on clinical informatics and how much on data science.

STEPS TOWARDS CONVERGENCE

There are several encouraging steps towards convergence. We highlight and commend several excellent initiatives that are taking a collaborative and aligned approach:

- The NHS Digital Academy²⁰
- Health Education England's 'Building a digital ready workforce' programme²¹

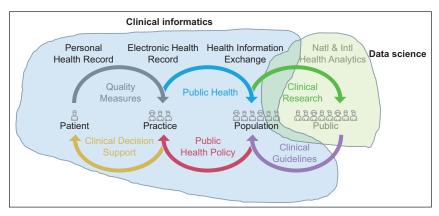


Figure 2 Learning health systems vitally depend on clinical informatics

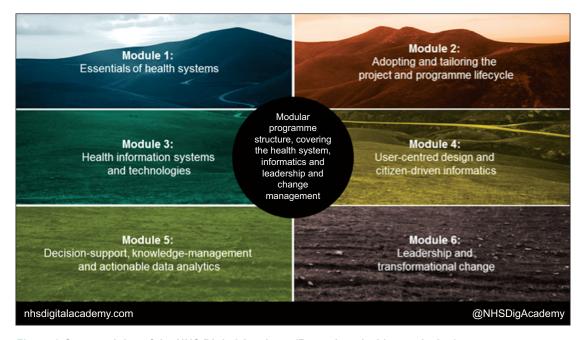


Figure 3 Core modules of the NHS Digital Academy (Reproduced with permission)

- The UK Faculty of Clinical Informatics²²
- The Federation of Informatics Professionals²³

In addition, some of the Academic Health Science Networks²⁴ are helping to bring together the practitioner and research communities in both data science and clinical informatics initiatives and the 'Global Digital Exemplars'5 are to participate in a national evaluation programme²⁵. The invitation to participate in the recently launched 'Local health and care record exemplar' programme²⁶ includes several references to 'research', but unfortunately this seems to be solely the 'Big Data' aspect not the clinical informatics research needed to improve frontline usage and data quality.

One focus of the NHS Digital Academy (Figure 3) will be to unpick the currently secret recipe for deriving user satisfaction, productivity and good quality data from clinical systems. There is a significant focus on user-centred design, interoperability and healthcare system standards within the modules. The aim is to ensure that the cohort of 'digital leaders' understand the role of the end-to-end technology from data standards to usability in achieving good data for direct care and research.

EXPANDING CLINICAL INFORMATICS RESEARCH AND EDUCATION CAPACITY AND CAPABILITY

However, we suggest that the UK needs increased clinical informatics research and education capacity and capability at much greater scale and ambition to be able to address the fundamental gaps in the discipline's evidence base and mitigate the absence of regulation.4 Numerous basic clinical informatics research questions remain to be satisfactorily addressed,27 including in the fields of:

- Cost effectiveness^{28,29}
- Efficiency/productivity30-32
- Impact on service utilisation33
- Patient empowerment/outcomes³⁴
- Decision support³⁵
- Usability and human factors^{36,37}
- Unintended consequences^{38–41}
- Application of safety-critical software engineering methods.42

This realisation has led to the 'Evidence-Based Health informatics' movement, which is well described in an open access textbook. 43 The way to build our discipline's evidence base is to identify and test relevant theories using rigorous evaluation studies.44 A key measure that would bring the 'two cultures' of data science and clinical informatics closer is to make independent evaluation of digital health interventions the norm, not the exception. 45,46 These studies need to be carried out by independent evaluators, not system developers, because there is clear systematic review evidence that even randomised controlled trials (RCTs) carried out by system developers are three times as likely to generate positive results than RCTs carried out by independent evaluators.47

CONCLUSIONS

We have highlighted serious issues with the quality of routine data and how that can be addressed beyond nugatory 'data cleansing'. We submit that policy makers and research funders need to acknowledge the existing gap between the 'two cultures' and recognise that the full social and economic benefits of digital health and data science can only be realised by accepting the interdisciplinary nature of biomedical informatics and supporting a significant expansion of clinical informatics capacity and capability.

REFERENCES

- 1. Snow CP. The Two Cultures and the Scientific Revolution. The Rede Lecture, 1959. Cambridge, UK: University Press, 1959.
- 2. Scott P, Bentley S, Carpenter I, Harvey D, Hoogewerf J, Jokhani M, et al. Developing a conformance methodology for clinically-defined medical record headings: a preliminary report. European Journal for Biomedical Informatics 2015;11(2):23-30.
- 3. Burnett S, Franklin BD, Moorthy K, Cooke MW and Vincent C. How reliable are clinical systems in the UK NHS? A study of seven NHS organisations. BMJ Quality & Safety 2012;21(6):466-72. Available from: https://doi.org/10.1136/ bmjqs-2011-000442. PMid:22495099; PMCid:PMC3355340.
- 4. Koppel R. The health information technology safety framework: building great structures on vast voids. BMJ Quality & Safety 2016;25(4):218-20. Available from: https://doi.org/10.1136/ bmjqs-2015-004746. PMid:26584580.
- 5. NHS England. Global digital exemplars. 2018. Available from: https://www.england.nhs.uk/digitaltechnology/info-revolution/ exemplars/. Accessed 26 March 2018.
- 6. Mukherjee M, Wyatt JC, Simpson CR and Sheikh A. Usage of allergy codes in primary care electronic health records: a national evaluation in Scotland. Allergy 2016;71(11):1594-602. Available from: https://doi.org/10.1111/all.12928. PMid:27146325.
- 7. Jordan K, Porcheret M and Croft P. Quality of morbidity coding in general practice computerized medical records: a systematic review. Family Practice 2004;21(4):396-412. Available from: https://doi.org/10.1093/fampra/cmh409. PMid:15249528.
- 8. Tate AR, Dungey S, Glew S, Beloff N, Williams R and Williams T. Quality of recording of diabetes in the UK: how does the GP's method of coding clinical data affect incidence estimates? Cross-sectional study using the CPRD database. BMJ Open 2017;7(1):e012905. Available from: https://doi.org/10.1136/ bmjopen-2016-012905. PMid:28122831; PMCid:PMC5278252.
- 9. Kyriacou DN and Lewis RJ. Confounding by indication in clinical research. JAMA 2016;316(17):1818-9. Available from: https:// doi.org/10.1001/jama.2016.16435. PMid:27802529.
- 10. Mukherjee M, Stoddart A, Gupta RP, Nwaru BI, Farr A, Heaven M, et al. The epidemiology, healthcare and societal burden and costs of asthma in the UK and its member nations: analyses of standalone and linked national databases. BMC Medicine 2016;14(1):113. Available from: https://doi.org/10.1186/s12916-016-0657-8. PMid:27568881; PMCid:PMC5002970.
- 11. KLAS. What is the Arch Collaborative? KLAS Research [electronic document]. 2017. Available from: https://klasresearch. com/usability-studies.
- 12. Koppel R. Illusions and delusions of cut, pasted, and cloned notes: ephemeral reality and pixel prevarications. Chest 2014;145(3):444-5. Available from: https://doi.org/10.1378/ chest.13-1846.
- 13. Kulikowski CA, Shortliffe EH, Currie LM, Elkin PL, Hunter LE, Johnson TR, et al. AMIA Board white paper: definition of biomedical informatics and specification of core competencies for graduate education in the discipline. Journal of the American Medical Informatics Association 2012;19(6):931-8.

- Available from: https://doi.org/10.1136/amiajnl-2012-001053. PMid:22683918; PMCid:PMC3534470.
- 14. AMIA. Health Informatics Core Competencies for CAHIIM [electronic document]. 2017. Available from: https://www. amia.org/sites/default/files/AMIA-Health-Informatics-Core-Competencies-for-CAHIIM.PDF. Accessed 26 March 2018.
- 15. Coiera E. Guide to Health Informatics, 2nd edition. Abingdon, UK: CRC press, 2003. Available from: https://doi.org/10.1201/ b13618.
- 16. Heathfield HA and Wyatt J. The road to professionalism in medical informatics: a proposal for debate. Methods of Information in Medicine 1995;34(5):426-33. Available from: https://doi. org/10.1055/s-0038-1634627.
- 17. Friedman CP, Rubin JC and Sullivan KJ. Toward an Information Infrastructure for Global Health Improvement. Yearbook of Medical Informatics 2017;26(1):16-23. Available from: https:// doi.org/10.15265/IY-2017-004. PMid:28480469.
- 18. Rouse WB, Johns MME and Pepe KM. Learning in the health care enterprise. Learning Health Systems 2017;1(4):e10024. Available from: https://doi.org/10.1002/lrh2.10024.
- 19. ONC. A 10-year vision to achieve an interoperable health IT infrastructure [electronic document]. from: https://www.healthit.gov/sites/default/files/ ONC10yearInteroperabilityConceptPaper.pdf. Accessed 17 April 2018.
- 20. NHS England. NHS digital academy. 2017. Available from: https://www.england.nhs.uk/digitaltechnology/info-revolution/ nhs-digital-academy/. Accessed 26 March 2018.
- 21. Health Education England. Building a digital ready workforce. 2018. Available from: https://hee.nhs.uk/our-work/building-digital-ready-workforce. Accessed 26 March 2018.
- 22. Faculty of Clinical Informatics. Safe, effective and efficient healthcare achieved through the best use of information and information technology. 2018. Available from: https://www.facultyofclinicalinformatics.org.uk/. Accessed 26 March 2018.
- 23. De Lusignan S, Barlow J and Scott PJ. Genesis of a UK Faculty of Clinical Informatics at a time of anticipation for some, and ruby, golden and diamond celebrations for others. Journal of Innovation in Health Informatics 2018;24(4):344-6. Available from: https://doi.org/10.14236/jhi.v24i4.1003. PMid:29334353.
- 24. NHS England. Academic health science networks. 2018. Available from: https://www.england.nhs.uk/ourwork/part-rel/ ahsn/. Accessed 26 March 2018.
- 25. The University of Edinburgh. Global digital exemplar programme evaluation. 2018. Available from: https://www.ed.ac. uk/usher/digital-exemplars. Accessed 26 March 2018.
- 26. Farenden J and Singh I. Local Health and Care Record Exemplars. Invitation to Participate. UK; NHS England, 2018.
- 27. Haux R, Kulikowski CA, Bakken S, de Lusignan S, Kimura M, Koch S, et al. Research strategies for biomedical and health informatics. Some thought-provoking and critical proposals to encourage scientific debate on the nature of good research in medical informatics. Methods of Information in Medicine 2017;56:e1-10. Available from: https://doi.org/10.3414/ME16-01-0125. PMCid:PMC5388922.

- 28. Reis ZSN, Maia TA, Marcolino MS, Becerra-Posada F, Novillo-Ortiz D and Ribeiro ALP. Is there evidence of cost benefits of electronic medical records, standards, or interoperability in hospital information systems? Overview of systematic reviews. JMIR Medical Informatics 2017;5(3):e26. Available from: https://doi.org/10.2196/medinform.7400. PMid:28851681: PMCid:PMC5596299.
- 29. Dranove D, Forman C, Goldfarb A and Greenstein S. The Trillion Dollar Conundrum: Complementarities and Health Information Technology. US National Bureau of Economic Research Working Paper Series, No. 18281, 2012.
- 30. Friedberg MW, Chen PG, Van Busum KR, Aunon F, Pham C, Caloyeras J, et al. Factors affecting physician professional satisfaction and their implications for patient care, health systems, and health policy. Rand Health Quarterly 2014;3(4):1. PMid:28083306; PMCid:PMC5051918.
- 31. Hill RG Jr, Sears LM and Melanson SW. 4000 clicks: a productivity analysis of electronic medical records in a community hospital ED. The American Journal of Emergency Medicine 2013;31(11):1591-4. Available from: https://doi.org/10.1016/j. ajem.2013.06.028. PMid:24060331.
- 32. Heponiemi T, Hypponen H, Vehko T, Kujala S, Aalto AM, Vanska J, et al. Finnish physicians' stress related to information systems keeps increasing: a longitudinal three-wave survey study. BMC Medical Informatics and Decision Making 2017;17(1):147. Available from: https://doi.org/10.1186/s12911-017-0545-y. PMid:29041971: PMCid:PMC5646125.
- 33. Kash BA, Baek J, Davis E, Champagne-Langabeer T and Langabeer JR 2nd. Review of successful hospital readmission reduction strategies and the role of health information exchange. International Journal of Medical Informatics 2017;104:97-104. Available from: https://doi.org/10.1016/j.ijmedinf.2017.05.012. PMid:28599821.
- 34. Rigby M, Georgiou A, Hypponen H, Ammenwerth E, de Keizer N, Magrabi F, et al. Patient portals as a means of information and communication technology support to patient-centric care coordination—the missing evidence and the challenges of evaluation. A joint contribution of IMIA WG EVAL and EFMI WG EVAL. Yearbook of Medical Informatics 2015;10(1):148-59. Available from: https://doi.org/10.15265/IY-2015-007. PMid:26123909; PMCid:PMC4587055.
- 35. Ammenwerth E, Nykanen P, Rigby M and de Keizer N. Clinical decision support systems: need for evidence, need for evaluation. Artificial Intelligence in Medicine 2013;59(1):1-3. Available from: https://doi.org/10.1016/j.artmed.2013.05.001. PMid:23810731.
- 36. Marcilly R, Peute L and Beuscart-Zephir MC. From usability engineering to evidence-based usability in health IT. Studies in Health Technology and Informatics 2016;222:126-38. PMid:27198098.

- 37. Turner P, Kushniruk A and Nohr C. Are we there yet? Human factors knowledge and health information technology-the challenges of implementation and impact. Yearbook of Medical Informatics 2017;26(1):84-91. Available from: https://doi. org/10.15265/IY-2017-014. PMid:29063542.
- 38. Coiera E, Ash J and Berg M. The unintended consequences of health information technology revisited. Yearbook of Medical Informatics 2016;10(1):163-9. Available from: https://doi.org/10.15265/IY-2016-014. PMid:27830246; PMCid:PMC5171576.
- 39. Schiff GD, Amato MG, Equale T, Boehne JJ, Wright A, Koppel R, et al. Computerised physician order entry-related medication errors: analysis of reported errors and vulnerability testing of current systems. BMJ Quality & Safety 2015;24(4):264-71. Available from: https://doi.org/10.1136/bmjqs-2014-003555. PMid:25595599; PMCid:PMC4392214.
- 40. Amato MG, Salazar A, Hickman TT, Quist AJ, Volk LA, Wright A, et al. Computerized prescriber order entry-related patient safety reports: analysis of 2522 medication errors. Journal of the American Medical Informatics Association 2017;24(2):316-22. PMid:27678459
- 41. Cresswell KM, Bates DW, Williams R, Morrison Z, Slee A, Coleman J, et al. Evaluation of medium-term consequences of implementing commercial computerized physician order entry and clinical decision support prescribing systems in two 'early adopter' hospitals. Journal of the American Medical Informatics Association 2014:21(e2):e194–202. Available from: https://doi.org/10.1136/amiajnl-2013-002252. PMid:24431334; PMCid:PMC4173168.
- 42. Thomas M. Making Software Correct by Construction. Oxford, UK: Gresham College, 2017.
- 43. Ammenwerth E and Rigby M (Eds). Evidence-Based Health Informatics. Amsterdam, Netherlands: IOS Press, 2016.
- 44. Wyatt JC. Evidence-based Health Informatics and the Scientific Development of the Field. Studies in Health Technology and Informatics 2016;222:14-24. PMid:27198088.
- 45. Sheikh A, Atun R and Bates DW. The need for independent evaluations of government-led health information technology initiatives. BMJ Quality & Safety 2014;23(8):611-3. Available from: https://doi.org/10.1136/bmjqs-2014-003273. PMid:24950693.
- 46. Scott P. Exploiting the information revolution: call for independent evaluation of the latest English national experiment. Journal of Innovation in Health Informatics 2015;22(1):244-9. Available from: https://doi.org/10.14236/jhi.v22i1.139. PMid:25924557.
- 47. Garg AX, Adhikari NK, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. JAMA 2005;293(10):1223-38. Available from: https://doi.org/10.1001/ jama.293.10.1223. PMid:15755945.