Artificial Intelligence in Primary Health Care: Perceptions, Issues, and Challenges
Primary Health Care Informatics Working Group Contribution to the Yearbook of Medical Informatics 2019

Harshana Liyanage[1], Siaw-Teng Liaw[2], Jitendra Jonnagaddala[2], Richard Schreiber[3], Craig Kuziemsky[4], Amanda L Terry[5], Simon de Lusignan[1]

[1] Department of Clinical & Experimental Medicine, University of Surrey, Guildford, Surrey, UK
[2] School of Public Health & Community Medicine, UNSW Medicine Australia, Ingham Institute of Applied Medical Research, NSW, Australia
[3] Geisinger Holy Spirit, Camp Hill, PA, USA
[4] Telfer School of Management, University of Ottawa, Ottawa, ON, Canada
[5] Centre for Studies in Family Medicine, Department of Family Medicine, Interfaculty Program in Public Health, Department of Epidemiology & Biostatistics, Schulich School of Medicine & Dentistry, Western University, London, ON, Canada

Summary

Background: Artificial intelligence (AI) is heralded as an approach that might augment or substitute for the limited processing power of the human brain of primary health care (PHC) professionals. However, there are concerns that AI-mediated decisions may be hard to validate and challenge, or may result in rogue decisions.

Objective: To form consensus about perceptions, issues, and challenges of AI in primary care.

Method: A three-round Delphi study was conducted. Round 1 explored experts' viewpoints on AI in PHC (n=20). Round 2 rated the appropriateness of statements arising from round one (n=12). The third round was an online panel discussion of findings (n=8) with the members of both the International Medical Informatics Association and the European Federation of Medical Informatics Primary Health Care Informatics Working Groups.

Results: PHC and informatics experts reported AI has potential to improve managerial and clinical decisions and processes, and this would be facilitated by common data standards. The respondents did not agree that AI applications should learn and adapt to clinician preferences or behaviour and they did not agree on the extent of AI potential for harm to patients. It was more difficult to assess the impact of AI-based applications on continuity and coordination of care.

Conclusion: While the use of AI in medicine should enhance healthcare delivery, we need to ensure meticulous design and evaluation of AI applications. The primary care informatics community needs to be proactive and to guide the ethical and rigorous development of AI applications so that they will be safe and effective.

Introduction

Health systems around the globe are under stress due to many socio-political factors. The delivery of health services using optimal resources without compromising patient safety is in demand more than ever before. The rise in ageing population with multiple chronic diseases together with the increase of healthcare spending worldwide are some of the key factors putting strain on healthcare systems [1]. Primary health care (PHC) to some extent can respond to these demands at both population and community levels [2]. PHC is rapidly evolving not only in terms of health policies but also technologically. The majority of PHC providers are now digitized and use health information systems as part of care provision. With the advances in computational and informatics technologies it is now possible to exploit these health information systems using Artificial intelligence (AI) concepts such as machine learning and deep learning [3, 4].

AI is not a new concept and has been around for more than 50 years, popularized in the 1980s and 1990s with the advent of neural networks. However, this trend did not last long mainly due to bottleneck in computational capabilities of hardware at the time. With the latest advances in Graphics Processing Units (GPUs), we can now overcome these computational limitations allowing us to develop more efficient neural networks in the form of deep learning. Deep learning is a machine learning technique where the models are trained using artificial neural networks with many layers (sometimes around 1,000). Deep learning has demonstrated significant results in various non-health and health-related applications using computer vision and natural language processing [5, 6]. However, very few of health-related AI systems are actually incorporated into clinical practice [7].

In recent years, deep learning has been used in PHC. Abramoff *et al.* developed an AI system, approved by the Federal Drug Administration of USA, to detect diabetic retinopathy in PHC centres [8]. A similar AI system using deep learning was developed in 2016 for the same purpose – automated

diagnosis of diabetic retinopathy [9]. The key limitations of these two specific systems included a need for external validation, integration into clinical workflow, and the attitudes of clinicians [10, 11]. AI systems based on deep learning and other similar machine learning techniques are heavily critiqued for their 'black-box' paradigm wherein some of the intrinsic estimations are not clinically interpretable in biological terms. Additionally, various ethical issues are observed in the application of AI in PHC. One such ethical issue is the risk of introducing bias. An AI system can incorporate the biases inherent to the training data set, and propagate them into the validation set [12]. Collective knowledge from clinicians might be able to avoid these biases and subsequently help making appropriate clinical decisions. Another ethical issue is that dependence on AI by clinicians might change the patient-clinician relationship dynamic.

The above-mentioned studies together with several others give us a glimpse into the future on how PHC can leverage AI. However, despite all the methodological and computational advances in AI, very few are translated into routine clinical practice. We believe the issues and challenges surrounding the use of AI in PHC are one of the key reasons. Additionally, there is significant variability of opinions on the use of AI in PHC among various stakeholders - clinicians, informaticians, AI researchers, and AI practitioners. In this context, the International Medical Informatics Association (IMIA) Primary Health Care Informatics Working Group undertook this Delphi study to seek consensus on the perceptions, issues, and challenges of AI in PHC.

Methods

Consensus Exercise

We recruited volunteer health informatics experts and clinicians involved in the Primary Health Care Informatics Working Group of IMIA to conduct a three-round Delphi study. The study was conducted during the months of October and November 2018. Each round lasted for about two weeks.

a. Round 1: Identifying the global perspectives of issues and challenges associated with using artificial intelligence in primary care – an online survey

Round 1 was an online survey which aimed to explore clinicians' and health informatics experts' awareness of typical uses of AI in primary care setting. We also inquired about the role of AI in fulfilling requirements of safety, interoperability, data quality, and ethics. Finally, we inquired about the future potential of AI in primary care to enhance health care. The recruitment was done mainly within the primary care but was also extended to other related professional networks based on their interest and exposure to the topic. The response period for the survey was two weeks with a reminder being sent to the invitees during the last week.

b. Round 2: Rating statements using the RAND/UCLA appropriateness method – an online survey

The responses from Round 1 were tabulated and analysed by the authors. Responses were organised according to a series of themes. We created 14 consensus statements based on the responses and across the themes identified. The 14 consensus statements were sent to the panel of 20 experts who responded to Round 1. In addition to the consensus statements, we also included two open ended questions to capture additional information on specific uses of AI and the clinician's role as a learned intermediary. The two open ended questions were:

- Are there any other AI use cases or scenarios that you would like to see included?
- In the AI environment, does the clinician still have a role as the "learned intermediary" between the system/ knowledge source and the patient?

Twelve participants (60%) responded to the Round 2 survey. The list of statements is given in Table 1. We replaced the standard terms used in the UCLA/RAND appropriateness method, "Highly appropriate" and "Highly inappropriate", with "Strongly agree" and "Strongly disagree".
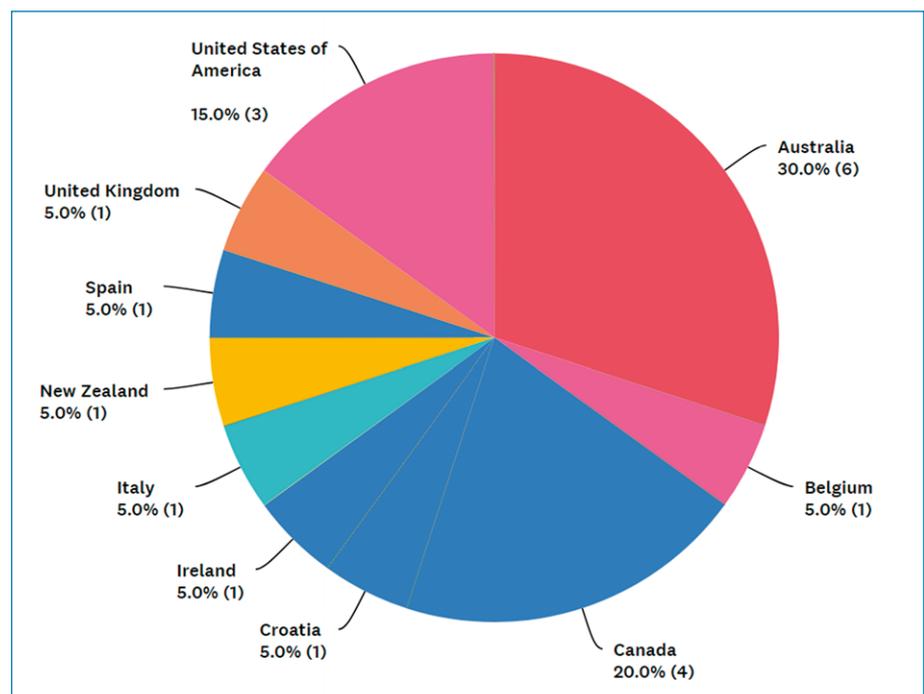

Fig. 1 Distribution of the health informatics experts who participated to the Round 1.

c. Round 3: Discussion of the findings by health informatics experts – an online panel discussion

The final round of the consensus process was an online panel discussion conducted as a video conference with a shared screen to present results. Three separate online meetings were organised to engage panel members in different time zones. Thirteen experts (65%) participated to this round.

Results

The process involved inviting and consulting with an international panel of 20 experts from 9 countries: Australia, Belgium, Canada, Croatia, Italy, New Zealand, Spain, United Kingdom, and USA.

a. Round 1: Identifying the global perspectives of issues and challenges associated with using artificial intelligence in primary care – an online survey

Panel Characteristics
The panel included experts from a range of professions including clinicians (7), academics (9), informaticians (2), and researchers (2). The majority of panel members were knowledgeable about AI although they did not have substantial hands-on experience with utilising AI applications in practice.

Benefit Use Cases for AI in Primary Care
The panel provided a range of benefit use cases where they considered AI to be a useful addition in the primary care setting. The responses received are generalised across several themes in Table 1.

Risks Associated with Using AI in Primary Care
The panel was asked about potential risks associated with the use of AI as an integral part of primary health care. We have grouped the use cases to generalise situations that could potentially be harmful to patients in Table 2.

In order to enable safe use of AI applications in primary care, it was believed that input data (for AI application), output data, and access protocols should be kept within a secure infrastructure. For a safe processing of patient data, compliance with data protection regulations such as the General Data Protection Regulation (GDPR) was considered to be important.

Adoption of AI in Primary Care
The panel members unanimously agreed that using common standards in computerised medical records such as common data models, common metadata standards, common terminologies, and common data quality metrics would facilitate effective implementation of AI across various primary care providers. To encourage the adoption of AI in primary care, the panel strongly believed that AI applications need to be usable within the practitioner's workflow and relevant to clinical practice. Respondents expressed the desire for a strong evidence to support AI. The panel, however, had mixed views about cost efficiency being a factor for encouraging adoption of AI.

Ethical and Lawful Processing of Patient Data by AI Applications
The panel agreed that AI applications require close monitoring when processing patient data. There was an agreement for the need for compliance with standards for AI applications and the need for transparency regarding data processing. Most panel members considered informed consent to be important for lawful processing. There was less agreement that the principle of the "learned intermediary" applies to AI applications (the "learned intermediary" principle holds clinicians responsible to use technology in combination with their professional knowledge and experience providing care) [13]. Similarly, there was less agreement for the requirement of an ethics committee to have well defined processes for dealing with inconsistent outputs from AI applications.

Implications of AI in Learning Health Systems
The expert panel members provided a range of implications from using AI in learning health systems. They indicated that learning health systems should include a useful collection of methodologies that will help reflect data back to the system to drive quality improvement. They also suggested

Table 1 Examples of benefit use cases in which AI can be leveraged in a primary care setting as suggested by the panel members.

Themes	Examples of benefit use cases of AI in primary care setting
Decision support to improve primary health care processes	a) Improving accessibility by triaging primary care patients and conduct a preliminary analysis suggesting likely diagnosis. b) Learning preferred prescribing patterns of clinicians that use AI-enhanced computerised medical records c) Assisting the prototype development of decision support tools
Pattern recognition in imaging results	a) Automatic detection of tumours using whole slide digital pathology images
Predictive modelling performed on primary care health data	a) Detection of high risk for mental health disorders/ cardiovascular disease b) AI-driven tools for clinicians e.g. prediction of mortality c) Assistance with diagnosis of obscure cases using iterative algorithms of accumulated case histories d) Assistance with management of complex cases, using iterative accumulation of outcome data (big data repositories with complex neural networks) e) Early diagnosis of diseases in primary care patients
Business analytics for primary care provider	a) AI applications that operate on routinely collected administrative data could provide regular feedback to practice managers, business owners, and individual clinicians (doctors, nurses, and others) to reduce variability and improve quality of care b) AI modelling of administrative data could assist in finding organizational models for an effective comparison among different countries

that AI may optimise and tailor best practice to the local environment. This is positive at first but over time the system may begin to overfit, meaning that the learning system reached a saturation point from which it may be difficult to learn new changes, especially if they are contradictory to what was previously learnt. AI systems must be open-minded over time to adopt and perhaps even challenge contradictory rules and behaviours.

Future of AI in Primary Care

The increasing use of electronic medical record systems in the last few decades means that there is a large volume of data available for AI applications to utilise. AI can help by augmenting (supporting) tasks such as decision making to reduce cognitive burden on clinicians. This would be particularly helpful for challenging diagnostic or therapeutic decision-making. It can also do the background data analysis to enable providers to have a more integrated record of their patient during a consultation. AI

may also have an important role in identifying populations of patients at risk. The panel members expressed optimism that AI would be most promising to learn new risk stratification models and rules from GP data. In particular, AI systems may help reduce health inequalities by surfacing the most vulnerable patients. The need for clinicians to drive care delivery will not go away, and in fact will become more critical since various outcomes suggested by AI applications required physician validation for the particular patient. Widespread acceptance of AI outputs requires considerable further work to assure it a place as an additional and completely trusted source for direct patient care. As an example, panel members speculated that as physicians learn to validate or refute deep learning decisions which may initially appear non-plausible, this will increase physicians' trust in AI processes. Over time, we can either accept these as good AI decisions, or learn when the human brain may need to override a proposition for a final decision.

b. *Round 2: Rating statements using the RAND/UCLA appropriateness method – an online survey*

There was a good degree of consensus, as defined by the RAND/UCLA method [14] by the end of the final round (see Table 3). The statements for Round 3 (Table 2) had agreement on 8 out of 14 statements.

c. *Round 3: Discussion of the findings by health informatics experts – an online panel discussion*

The expert panel discussed various possible reason for the variability in agreement levels for the statements in Round 2. The discussion section incorporates feedback received during these meetings.

Discussion
Principal Findings

The participants suggest that AI has potential to improve primary health care but unsupervised machine learning is currently not sufficiently mature or robust to be confidently used without checks in place. They were mostly in agreement that advances in AI application in primary care can lead to improvement of managerial and clinical decisions and processes. The primary care community needs to be proactive and guide the ethical and rigorous development of AI applications so that they will be safe and effective in the workplace.

The most established use of AI in primary care reported is suggested to be predictive modelling [15]. This is likely to be because the respondents do not have substantial clinical experience with AI tools - their suggested use cases may be more academic and non-clinical such as predictive modelling. Similarly, their responses to the statements are likely to be more academic.

Participants also agreed that formal processes need to be developed and Ethics Committees (or Institutional Risk Management Committees) be trained to assess the ethical processing of data in AI applications. Data governance committees should contribute the oversight of AI applications and have processes in place to monitor data process-

Table 2 Examples of risk use cases in which AI could result in a potential risk to patients in primary care as suggested by the panel members.

Themes	Examples of risk use cases of AI in primary care setting
AI technology currently available to deploy in primary care is still not competent to replace human decision making in clinical scenarios	a) Interpreting the results of an analysis using AI without an understanding of the primary health care context b) Overreliance on what AI can do. Using AI as a substitute for due clinical diligence c) Missing competencies/willingness in using AI properly d) In AI, few techniques such as deep neural networks are incapable of explaining the underlying models completely. This makes it hard to interpret the interplay between covariates in a model e) Relying on AI and not using human skills to ensure it is correct f) Going down the primrose path. One of the most dangerous aspects of black-box algorithms is not knowing the source of the data. To take an extreme example, if the AI is built for fever of unknown origin at a major referral hospital in the US, it will not be applicable to a patient with fever in sub-Saharan Africa who in fact has malaria.
Risk of medical errors	a) Potential for errors in prescribing. If a doctor prescribes a medication using adult doses for a child, and the AI doesn't have a guideline to spot the error, the AI could propagate the error into the child's future and that of other children on the same medication. This happens with humans (who are experts and specialists) and can happen in a learning AI scenario b) Incorrect diagnosis leading to unnecessary treatment c) Assumed effectiveness before proper trials undertaken
Risk of bias	a) That the data behind the constructed AI knowledge model was biased, or not compatible with the patient to whom the clinician applies the AI: e.g., a model learned in a population with specific sub-phenotypes may not be adequate to another population, or a model learned with past data models (ICD-9) may not be adequate/generalizable to new data models (ICD-10)
Risk of secondary effects of utilising AI	a) Insurance providers using AI for higher premiums or even excluding certain people for insurance

Table 3 Consensus statements generated from the analysis of Round 1's responses (with Agreement written in green, Equivocation in brown, and Disagreement in red according to responses from Round 1).

Statement 1 - The most prevalent use of AI currently in primary care is for predictive modelling (e.g. detection of high risk for mental health disorders / cardiovascular disease) based on knowledge inferred from large clinical datasets.

Statement 2 - AI in primary care is currently needed more to manage provision of care (e.g. triage) than for clinical decision support.

Statement 3 - AI applications can be incorporated more easily in business analytics in primary care than analytics to support the clinical process.

Statement 4 - AI applications should be capable of assessing and adapting to the preferences of a clinician (e.g. learning about preferred medication that a clinician prescribes for male adult hypertensive).

Statement 5 - (Over) reliance on AI applications to make clinical decisions can be harmful to patients.

Statement 6 - Current AI applications mainly operate as black boxes (from the perspective of clinicians) and therefore need regular scrutiny by users (e.g. clinicians and managers)

Statement 6 - Excessive patient data will reduce the effectiveness of patients' online experience. [Inhibitor] [Equivocation]

Statement 7 - Current datasets used to train and testing AI applications are not representative of patient services enhances shared decision-making. [Enabler] [Disagreement]

Statement 9 - Access to patient data such as radiology results or lab results will not be cost-beneficial as it will not be used by the wider patient population. [Inhibitor] [Disagreement]

>   **(a) the real world (e.g., a patient wearing fitness monitoring devices may be healthier than the general population (worried well)).**
>
>   **(b) specified population (e.g., a model learned in a population with specific sub-phenotypes may not be applicable to other populations).**
>
>   **(c) the underlying terminological system (e.g., a model learned with past data models (ICD-9) may not be adequate/generalizable to new data models (ICD-10 or SNOMED-CT)).**

Statement 8 - Clinical decisions made by AI applications may lead to unnecessary treatment which may not be those recommended by evidence-based guidelines.

Statement 9 - Ethics committees (or institutional risk management committees) should be trained in formal processes to assess the ethical processing of data in AI applications.

Statement 10 - Data governance committees should also oversee AI applications.

Statement 11 - Data processing in AI applications needs to be monitored closely.

Statement 12 - Data output display needs to be assessed for fidelity and quality.

Statement 13 - Mechanisms to identify biases in unsupervised algorithms need to be implemented in all AI applications.

Statement 14 - Advances in AI application in primary care will lead to improvement of a) clinical decision making; b) risk assessment; c) care processes; d) continuity of care; e) coordination of care; f) safety of care; and g) managerial processes in health care.

---

ing in and outputs from AI applications for fidelity, bias, and quality.

There was less agreement on whether AI applications should be focused on service provision or decision support or whether it was easier to support the managerial or clinical process. It was not agreed that AI applications should learn and adapt to clinician preferences or behaviour or the extent of the potential for harm to patients.

Implication of Findings

The clinical and informatics community need to establish the professional rules for the initial and on going use of AI applications to support managerial or clinical practice. Specific legislation may be needed to address some of the more intractable issues such as the liability for "black box" approaches of AI or even the liability of the clinician as a learned intermediary.

There is an agreed need for regular scrutiny by users (e.g., clinicians and managers) because the accuracy, fidelity, or relevance of the output of AI is not guaranteed, the current training datasets for AI applications may not be representative of specific populations or of the underlying terminological system or data models, and there is a need for mechanisms to identify biases in unsupervised algorithms. Identification of biases should be followed up with "unlearning" processes that increase the accurate functionality of AI applications [16, 17].

Caution is needed as it may be more difficult to assess the impact of AI-based applications on continuity and coordination of care.

The panel members noted the unexpected finding that there was a lack of consensus regarding the potential for AI to assist and adapt to clinician preferences. Neural networks can continuously learn, which could assist primary care clinicians to define their particular patient population as well as include PHC's individual treatment preferences. Yet respondents appeared to not agree with this. Perhaps this was due to a misunderstanding of the question, or perhaps the panel was diversely versed in the promise of AI. The findings of our study closely mirrored outcomes of a recent

qualitative survey involving a large cohort of general practitioners in the UK in which they expressed both scepticism and optimism on the notion of replacing human roles in health care using AI [18].

Limitations of the Method

We used an opportunistic sample of health informatics experts drawn from international Primary Care Health Informatics Working Groups. While a globally representative list of experts was invited, there was no response from the African, South Asian, and Middle Eastern countries. Because respondents did not have substantial clinical experience with AI tools, their suggested use cases may be more academic and non-clinical such as predictive modelling. Similarly, their responses to the statements are likely to be more academic. In addition, as with most self-reported methods, the phrasing of questions may have an effect on the responses obtained.

Conclusions

PHC and informatics experts reported that AI has the potential to improve managerial and clinical decisions and processes. However, unsupervised machine learning is currently not sufficiently mature or robust to be used confidently without checks in place. The primary care informatics community needs to be proactive to guide the ethical and rigorous development of AI applications so that they will be safe and effective in the workplace.

References

1. Beard JR, Bloom DE. Towards a comprehensive public health response to population ageing. Lancet 2015;385(9968):658-61.
2. Starfield B, Shi L, Macinko J. Contribution of primary care to health systems and health. The Milbank Q 2005;83(3):457-502.
3. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform 2018 Nov 27;19(6):1236-46.
4. Xiao C, Choi E, Sun J. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. J Am Med Inform Assoc 2018;25(10):1419-28.
5. Powles J, Hodson H. Google DeepMind and healthcare in an age of algorithms. Health Technol (Berl) 2017;7(4):351-67.
6. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, van den Driessche G, et al. Mastering the game of Go with deep neural networks and tree search. Nature 2016;529:484.
7. The LJL. Artificial intelligence in health care: within touching distance. 2018;390(10114):2739.
8. Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ Digit Med 2018;1(1):39.
9. 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.
10. Keane PA, Topol EJ. With an eye to AI and autonomous diagnosis. NPJ Digi Med 2018;1(1):40.
11. Stead WW. Clinical implications and challenges of artificial intelligence and deep learning. JAMA 2018;320(11):1107-8.
12. Char DS, Shah NH, Magnus D. Implementing Machine Learning in Health Care — Addressing Ethical Challenges. N Engl J Med 2018;378(11):981-3.
13. Sullivan HR, Schweikart SJ. Are Current Tort Liability Doctrines Adequate for Addressing Injury Caused by AI? AMA J Ethics 2019;21(2):160-6.
14. Fitch K, Bernstein SJ, Aguilar MD, Burnand B, LaCalle JR. The RAND/UCLA appropriateness method user's manual. RAND CORP SANTA MONICA CA; 2001.
15. Alanazi HO, Abdullah AH, Qureshi KN. A Critical Review for Developing Accurate and Dynamic Predictive Models Using Machine Learning Methods in Medicine and Health Care. J Med Syst 2017;41(4):69.
16. Sullivan K, ElMolla A, Squires B, Luke S, editors. Unlearning from demonstration. Twenty-Third International Joint Conference on Artificial Intelligence; 2013.
17. Ashraf A, Khan S, Bhagwat N, Chakravarty M, Taati B. Learning to Unlearn: Building Immunity to Dataset Bias in Medical Imaging Studies. arXiv preprint arXiv:181201716. 2018.
18. Blease C, Bernstein MH, Gaab J, Kaptchuk TJ, Kossowsky J, Mandl KD, et al. Computerization and the future of primary care: A survey of general practitioners in the UK. PloS One 2018;13(12):e0207418.

Correspondence to:
Harshana Liyanage
Department of Clinical & Experimental Medicine
University of Surrey
GUILDFORD
Surrey GU2 7XH. UK
E-mail: h.s.liyanage@surrey.ac.uk