

From Patient Engagement to Precision Oncology: Leveraging Informatics to Advance Cancer Care

Ashley C. Griffin¹, Umit Topaloglu², Sean Davis³, Arlene E. Chung^{1, 4-5}

¹ University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

² Wake Forest University School of Medicine, Winston-Salem, NC, USA

³ National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

⁴ University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, USA

⁵ UNC Lineberger Comprehensive Cancer Center, Chapel Hill, NC, USA

Summary

Objectives: Conduct a survey of the literature for advancements in cancer informatics over the last three years in three specific areas where there has been unprecedented growth: 1) digital health; 2) machine learning; and 3) precision oncology. We also highlight the ethical implications and future opportunities within each area.

Methods: A search was conducted over a three-year period in two electronic databases (PubMed, Google Scholar) to identify peer-reviewed articles and conference proceedings. Search terms included variations of the following: neoplasms[MeSH], informatics[MeSH], cancer, oncology, clinical cancer informatics, medical cancer informatics. The search returned too many articles for practical review (23,994 from PubMed and 23,100 from Google Scholar). Thus, we conducted searches of key PubMed-indexed informatics journals and proceedings. We further limited our search to manuscripts that demonstrated a clear focus on clinical or translational cancer informatics. Manuscripts were then selected based on their methodological rigor, scientific impact, innovation, and contribution towards cancer informatics as a field or on their impact on cancer care and research.

Results: Key developments and opportunities in cancer informatics research in the areas of digital health, machine learning, and precision oncology were summarized.

Conclusion: While there are numerous innovations in the field of cancer informatics to advance prevention and clinical care, considerable challenges remain related to data sharing and privacy, digital accessibility, and algorithm biases and interpretation. The implementation and application of these findings in cancer care necessitates further consideration and research.

Keywords

Cancer informatics, precision medicine, oncology, machine learning, digital health

Yearb Med Inform 2020:235-42

<http://dx.doi.org/10.1055/s-0040-1701983>

1 Introduction

Advances in technology, hardware, and computing have created new opportunities to improve the quality of cancer care and research by leveraging informatics innovations. Digital health, machine learning, and precision oncology are key areas where there are noteworthy advances in the field of cancer informatics. As the adoption of smartphones and wearable technologies increases, patient-generated health data such as physical activity and electronic patient-reported outcomes (PROs) are being leveraged for prevention and in interventions, for remote monitoring during treatment and survivorship, and to predict health outcomes. Moving beyond cancer clinical trials, health systems are increasingly seeking to integrate PROs into routine cancer care processes. As part of the National Cancer Institute's (NCI) Cancer Moonshot Initiative, six healthcare systems will systematically integrate PROs into clinical workflows and electronic health records (EHRs) [1]. Additionally, much attention surrounds the hype and challenges of the application of machine learning algorithms in clinical care, including for a number of oncology use cases. However, there are concerns around biases and homogeneous training data sets. Thus, machine learning has had limited implementation into clinical care. Considerable progress in genomics and computational medicine has ushered in a new era in precision oncology, generating vast amounts of data, much of which is publicly accessible. For example, the NCI Precision Medicine Initiative-On-

cology has had recent success defining genetically targeted therapies, understanding tumor treatment resistance, and developing a U.S. cancer knowledge system [2]. While the opportunities within cancer informatics have much potential for positive impact, if not applied thoughtfully, these innovations risk propagating disparities and inequities. Operationalizing these informatics discoveries into routine cancer care also continues to present a diverse set of challenges.

2 Objectives

Given the rapid pace of cancer informatics research, our objective was to conduct a survey of the literature for advancements over the past several years in three key areas of growth: 1) digital health; 2) machine learning; and 3) precision oncology. We also highlight the ethical implications, challenges, and opportunities for each topic.

3 Methods

A literature search was conducted in two electronic databases (PubMed and Google Scholar) to identify peer-reviewed articles and conference proceedings. Search terms included variations of the following: neoplasms[MeSH], informatics[MeSH], cancer, oncology, clinical cancer informatics, medical cancer informatics. The search covered the period from October 15, 2016 to October

15, 2019 and returned too many articles for practical review (23,994 from PubMed and 23,100 from Google Scholar). Thus, we narrowed the search to key PubMed-indexed informatics journals and proceedings such as the *Journal of the American Medical Informatics Association (JAMIA)*, *JAMIA Open*, *Applied Clinical Informatics*, *Journal of Medical Internet Research*, *Bioinformatics*, *Journal of Biomedical Informatics*, *JCO Clinical Cancer Informatics*, and *Nature Digital Medicine*. We also performed a manual review of cancer informatics journals that were not yet indexed in PubMed. Additionally, we searched the proceedings of the AMIA Annual Symposia and AMIA Informatics Summits.

Given the sheer number of results, we focused on manuscripts that demonstrated a clear focus on clinical or translational cancer informatics. Manuscripts were selected based on methodological rigor, scientific impact, innovation, and contributions towards cancer informatics as a field, or on impact on cancer care. The reviewed manuscripts were grouped into three topics including digital health, machine learning, and precision oncology. For each topic, we describe the current state, implications, challenges, and opportunities.

4 Results

4.1 Digital Health

Digital health is the intersection of technologies with health and wellness to enhance communication and delivery of health services. It includes mobile health, health information technologies, wearable devices, telehealth or virtual health, and personalized medicine [3]. As health systems transition from traditional office-based care to more continuous interactions with patients and caregivers, emerging technologies are key drivers of the capture of patient-generated health data [4]. Currently, it is still challenging to bring together disparate data streams generated by patients from technologies such as smartphones, social media platforms, wearable activity trackers, and Internet of Things (IoT) devices such as smart home

sensors. However, multiple studies have begun to demonstrate the value of employing digital health technologies to improve cancer prevention and care [5-19].

For example, apps and web-based interventions have been used to support exercise [5] and in digital outreach efforts focused on cancer screening [6]. In a web-based exercise intervention to mitigate the effects of chemotherapy, breast cancer patients demonstrated improvements on the six-minute walk test, and abdominal, back, lower body strength compared to the control group over the eight-week study [5]. One study demonstrated that Google search and Facebook advertisements could be used to promote cancer screening [6]. These advertising campaigns were associated with increased visits to institutional websites and scheduled screening exams compared to the week before and after the campaign [6]. These studies reveal the potential for digital health apps to engage patients, reach broad populations, and improve physical activity outcomes.

In addition to being a vehicle to deliver interventions, social media can be used for behavioral surveillance. Several studies explored communications within patient communities with the most common posts or tweets on the topics of clinical trials, treatment, and support [7-10]. Taylor et al. found evidence of differing levels of emotional, informational, and social support for patients with lung cancer across platforms (*i.e.*, Facebook, Twitter, Macmillan.org.uk) [9], which may be due to differences in user characteristics or the nature of self-moderating communities. In terms of engagement, Cho *et al.*, examined the relationship between melanoma-related Instagram posts and engagement outcomes (*i.e.*, likes, comments, social support). They found that posts about physical consequences decreased the number of “likes” but increased comments and perceived emotional support, and that the inclusion of images increased the number of comments made about the posts [10]. Overall, these highlighted studies contribute to an understanding of participative engagement and experiences on social media platforms for written and visual cancer-related content that is user-generated, and suggest that patients may have varied experiences related to social support across platforms and communities.

There is mounting evidence that using PROs for cancer care symptom assessments can improve survival, quality of life, and help patients remain longer on treatment when compared to usual care [11]. While there is a wide range of PROs that could be assessed during cancer treatment and in survivorship, several efforts have aimed to identify key domains relevant to improving outcomes in cancer care. For example, a clinical advisory group identified priority domains and measures to enhance cancer care, which included distress and symptoms [12-14] and unmet needs [15]. Girgis *et al.*, then developed clinical recommendations from these domains, mapped them into well-being categories, and demonstrated that these recommendations could algorithmically provide automated clinical feedback [15]. There is also concern that the use of structured questionnaires to assess PROs may limit the ability to capture the broad range of symptoms experienced by patients during treatment. Chung *et al.*, examined whether patient-entered free-text narratives of symptomatic adverse events could be mapped to established terminologies from the National Cancer Institute (NCI) Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) [16]. Among 1,760 patients enrolled in three multicenter cancer trials, over half (58%) provided unstructured free-text entries, and most of these entries could be mapped *post hoc* to a PRO-CTCAE or Medical Dictionary for Regulatory Activities (MedDRA) term. This suggests that mapping to existing terminologies is not only feasible, but also provides the opportunity to supplement trial-specific questionnaires to better capture diverse patient experiences during cancer treatments.

Use of wearable activity trackers and other IoT sensors to capture additional patient-generated health data is expanding. Several feasibility studies have had favorable results, although sample size was generally small [17-19]. Studies noted that higher daily steps recorded from activity trackers were correlated with PROs such as increased performance status [17], lower distress [18, 19], fatigue [18], depressive symptoms [18, 19], adverse events, hospitalizations, and death [17]. Despite high drop-off rates for

wear time for activity trackers [18, 19], this research highlights that wearable devices could provide accurate assessments of performance status and physical function to inform treatment selection and could serve as a viable proxy for PROs.

4.2 Machine Learning

Recent advancements in machine learning (ML) reveal opportunities for ML to transform cancer care. The field has seen progress in early detection of cancers, and improved diagnostic accuracy, personalized therapeutics, and clinical workflows [20–46]. In the last two years, the field has also experienced substantial growth in FDA approvals for algorithms including oncology applications for the detection of suspicious lesions and clinical decision support (CDS) [47].

Recent reviews have highlighted a number of promising ML-based knowledge extractions from unstructured and semi-structured oncology data from EHRs, social media platforms, and online health communities [20, 21]. Using EHR clinical documents, Savova *et al.*, created a natural language processing (NLP) system, DeepPhe, to generate cancer phenotypes [22], which could help reduce time spent reviewing clinical documents. Guan *et al.*, used NLP techniques to extract genomic sequencing information from clinical progress notes and then classified the notes on whether sequencing led to a treatment change [23]. Qiu *et al.*, also leveraged NLP to extract the primary site of tumor origin from pathology reports [24]. Given that cancer stage and origin are rarely captured in structured format within EHRs, many of the recent approaches have focused on NLP to extract these data. However, Seneviratne *et al.*, used the Observational Medical Outcomes Partnership (OMOP) common data model [25] to create a feature vector framework to classify patients with prostate cancer by stage using structured EHR data [26]. This approach may be more generalizable than NLP methods since clinical notes vary widely in content depending on the type and author [26].

There have also been a number of studies that apply NLP to online health community and social media datasets [21, 27, 28]. Zhang *et al.*, used a convolutional neural

network to classify topics within an online health community focused on breast cancer and found that patients with different cancer stages have different topics of conversations [27]. Lee *et al.*, assessed conversations about living with cancer among sexual minority women using latent Dirichlet allocation, which revealed that the most common topic was about coping and connecting to others with shared gender identities [28]. Utilizing NLP techniques to analyze and understand topics in online communities could help to optimize interventions and deliver valuable support.

Recognizing the ongoing challenges with concept extraction and the scarcity of annotated datasets, researchers have employed open science efforts to foster interdisciplinary collaborations in several areas. For example, Temporal History of Your Medical Events (THYME) corpora [29, 30], Informatics for Integrating Biology & the Bedside (i2b2) challenge sets [31], and Medical Information Mart of Intensive Care (MIMIC-III) [32] contain collections of de-identified clinical text for research. DREAM Challenges have also hosted several cancer-focused competitions that yielded a number of R packages and established benchmarks for prognostic models [33, 34].

In medical imaging, ML has been deployed for early detection, improved diagnosis, and better prognostic accuracy. Ardila *et al.*, used low-dose chest computed tomography images to perform end-to-end lung cancer screening with high accuracy [35]. Similarly, predictive models to calculate breast cancer risk from mammograms demonstrated improved risk discrimination over current clinical standards [36]. In dermatology, several deep learning methods were found to be superior to dermatologists in diagnosing and detecting skin cancers [37, 38]. Wei *et al.*, also found employing deep learning approaches to classify lung pathology images yielded more accurate results when compared to pathologists [39]. Additionally, advancements have been made in ML-based CDS tools for classification of skin lesions, estimates of colon cancer outcomes based on various therapies, and management of uterine cervical abnormalities [40]. Given the lack of large imaging datasets for training models, investigators

have also focused on producing realistic synthetic data [41, 42]. Research by Senaras *et al.*, revealed that generative adversarial networks are a promising technique to generate labeled synthetic immunohistochemical stained breast tissue [41].

For precision cancer care, genomics tools have also demonstrated the potential to advance the discovery of novel targeted cancer therapies. DeepMind's AlphaFold tool has shown promise in predicting protein structure from its genetic sequence [43]. CRISPR/Cas9 screens have also facilitated tremendous progress in identifying essential genes across cancer cell lines [44]. Several studies have used ML to identify cancer subtypes using gene expression data [45] and pathological images [46]. For example, Coudray *et al.*, established that ML algorithms could predict six of the most commonly mutated genes within adenocarcinoma lung cancer [46]. ML could also be used to assist pathologists in detecting gene mutations or cancer subtypes. Given the variety of data sources necessary to compare personalized cancer care pathways, incorporating genetic, clinical, imaging, cost, and quality of life data into multifaceted CDS tools could aid in reducing the cognitive load of clinicians [40].

4.3 Precision Oncology

Precision medicine is an evolving set of diagnostic and treatment paradigms which seeks to optimize and tailor safe and effective care for patients based on their individual characteristics. Precision medicine has been advanced considerably by cancer genomics. Increasingly, discoveries using genomic technologies that have fueled translational research are impacting clinical settings [48–64]. Single cell level assays are proving transformational in the understanding of tumor heterogeneity, disease biology, and treatment resistance, and will likely play an increasingly important role in clinical research and precision medicine applications [48].

Several studies have demonstrated the feasibility of large cohorts prospectively assayed using genomic technologies with turnaround times short enough to impact

clinical decisions [49, 50]. For example, in an institutional precision oncology initiative at Memorial Sloan Kettering Cancer Center (MSKCC-IMPACT), approximately 10,000 patients, sequenced at a rate of more than 500 patients per month, underwent paired tumor and germline DNA sequencing of a panel of approximately 400 known cancer genes [49]. Turnaround time to final report was less than 21 days on average. Most (69%) of the somatic variants were not present in the large, public database of somatic variants called the Catalogue Of Somatic Mutations In Cancer (COSMIC) [51] at the time of the study, suggesting a gap between finding a variant and interpreting its relevance to a patient's clinical context. As with other similar studies, researchers relied on publicly available data resources, such as My Cancer Genome [52], OncoKB [53], Clinical Interpretation of Variants in Cancer (CIViC) [54], HemOnc.org [55], and community-based resources [56] to facilitate interpretation of results. The clinical impact of large-scale precision oncology sequencing efforts suggest mixed results. An overview suggests that only a minority of sequenced patients have at least one actionable molecular finding [57]. Specifically, only 18% and 11% of patients respectively from the NCI-MATCH (approximately 6,000 patients) and the MSKCC-IMPACT trials received targeted therapies [49, 58]. Increasing the number and effectiveness of available therapeutics remains a central challenge to broadening the impact of genomic precision medicine [59]. However, it will also be critical to identify treatments that may yield low therapeutic benefit.

Biomarker definition is another growth area in precision oncology. Precision immuno-oncologics, a class of relatively new therapeutics, have shown remarkable efficacy for some cancers. Several defined prognostic and predictive indices suggest which patients are most likely to benefit from, for example, immune checkpoint inhibitors [60, 61]. For Chimeric Antigen Receptor T-cell (CAR-T) therapy, there are over 100 trials targeting at least 25 different biomarkers in multiple cancer types [62]. Public health projects are also beginning to adopt strategies to capitalize on biomarkers and genomic information to enhance precision medicine applications. For instance, the U.S. Centers for Disease

Control and Prevention maintains the Public Health Genomics Knowledge Base (PHGKB) which includes cancer-specific resources to “provide systematically curated and updated information that bridges population-based research on genomics with clinical and public health applications [63].” The NCI has also created an online resource that announces and coordinates “precision cancer surveillance” efforts including the addition of patient-level genomic linkages into cancer registries [64].

5 Discussion

Innovations in cancer informatics are abundant in the areas of digital health, machine learning, and precision oncology and have great potential in improving cancer care. While there are considerable challenges, there are also many opportunities for future research to advance cancer informatics as a field.

5.1 Opportunities and Implications for Digital Health in Cancer Care

Given the ubiquitous adoption of smartphones and growing ownership of wearable and IoT devices, digital health technologies have the unique capability to gather longitudinal data outside of clinical settings that could yield deeper insights into a patient's symptom experience or help identify modifiable risk factors for cancer. Interoperability standards and numerous U.S. initiatives, such as the 21st Century Cures Act and the Promoting Interoperability Program (formerly named the Meaningful Use Program), are transforming the way individuals access medical services and manage their health, thereby accelerating the meaningful application of digital health technologies.

The evidence base for digital health in cancer care is currently limited by small sample size, short study duration, and limited focus on digital accessibility, underserved or underrepresented populations, or the needs and psychosocial outcomes of caregivers of patients with cancer. Patients with cognitive or physical disabilities may not be able to use

some technologies, and different modalities are needed to reach patients based on their abilities, preferences, cultural norms, level of training, and lived environment [65]. Caregivers may be an important bridge to reach patients who have technical, physical, or cognitive limitations, but this is an understudied area. Many studies specifically exclude participants based on non-ownership of a smartphone, lack of Internet access, severity of medical conditions, being on chemotherapy treatment, or a recurrence of cancer. Yet, these populations may be those who could benefit the most from digital health technologies and interventions.

As companies are increasingly engaged in acquiring and sharing individuals' health-related information, there are noteworthy concerns about data privacy, security, and the ethical use of digital health data. Given the growing number of partnerships between health systems, pharmaceutical companies, insurers, and technology companies, and the increased access by third parties to cloud-based patient data and associated analytical tools, there is growing unease about data linkages and the potential for misuse and data breaches. For example, Google recently sparked a federal investigation following their partnership with Ascension [66] due to concerns for sharing protected health information and a Department of Justice inquiry over their acquisition agreement for Fitbit [67]. Apple has several partnerships with insurers and pharmaceutical companies such as Aetna [68] and Eli Lilly [69], and Fitbit has partnerships with Bristol-Myers Squibb-Pfizer Alliance [70] and Humana [71]. Recently, it was revealed that the U.K.'s National Health Service has given Amazon free use of all health data under an Alexa advice deal [72]. Consumers may be unaware or inadequately informed about how their data are being used and by whom [73], suggesting the need for greater transparency of privacy and data sharing practices and policies to be consumable at the patient level. Although the FDA regulates some digital health apps or technologies, the majority of apps available in the marketplace are not validated through rigorous research. Additionally, it remains challenging for patients and caregivers to determine which apps are trustworthy, effective, and useful.

5.2 Opportunities and Implications for ML in Cancer Care

Machine learning research has had meaningful advances in cancer prevention, diagnostics, and prediction. This progress has been driven by increased availability of data and scalable computing infrastructures. Cloud-based platforms, such as Microsoft Azure Databricks [74], Google Cloud AutoML [75], and Amazon Web Services SageMaker [76] have simplified building, training, and deploying models, though developers should be cognizant of the computational and environmental costs associated with cloud-based platforms [77]. Real-time NLP and ML innovations within EHR systems are also being explored by companies like Nuance and Epic, which could lead to better capture of tumor staging information as structured data elements through ambient computing [78].

Despite the promise of ML to improve cancer care, there are substantial challenges that must be addressed before ML is implemented into routine cancer care. For instance, 450,000 women missed screening mammograms in the U.K. due to algorithm failure [79]. The IBM Watson for Oncology project highlighted a number of issues in the ML and oncology communities related to lack of validation and benchmarking, data quality, and subjectivity in interpretation of results [80]. Data quality and availability challenges in healthcare are an acknowledged barrier to research, particularly the paucity of data for underrepresented populations [81]. This may lead to biases since training data does not reflect the attributes of these populations yet may be applied in clinical care. There are also growing concerns about the “de-skilling” of physicians that could occur when some or all of the tasks become automated, such as a drop in a clinician’s diagnostic accuracy [82]. This has future implications for graduate medical education and how curricula may require transformation to address these emerging issues.

Given the large volume and diversity of training data needed for ML research, federated learning approaches, which do not require direct data sharing, have strong potential but remain difficult to implement due to privacy preservation challenges [83]. Furthermore, the “black box” nature

of many algorithms renders interpretation and benchmarking performance difficult. To improve algorithm transparency, Price *et al.*, proposed a three-step framework for validating “black box” algorithms, which involves: 1) having high quality training data and development procedures, 2) testing algorithm performance against independent test data, and 3) evaluating performance continuously [84]. Even with benchmarking, appropriate selection of realistic and bias-controlled test cases may still be an issue. The “intrinsic uncertainty” in medicine introduces variations in results interpretation [82], which also suggests that model performance criteria should be use case specific rather than based on standard scoring metrics [85].

In the era of Deepfake, ethical issues are inevitable if there are no appropriate regulatory frameworks for the deployment of ML algorithms. Health disparities could be widened due to lack of representative training data and the possible consequences from algorithm failures limit the current utility of implementing ML algorithms in real-world cancer care settings. Governance and policies for deployment, audits of performance, and implementation of best practices will be critical for safe implementation, but these are not yet widely used within health systems. To mitigate some of the challenges facing the field of ML, several organizations have proposed ethical and regulatory frameworks. The FDA has proposed policies for ensuring safe and effective use of ML-based software for medical purposes, including regulatory frameworks for software as a medical device [86], for CDS [87] as well as a pre-certification program [88]. In real-world deployment into clinical care, it is critical to have rigorous change protocols for algorithm modifications to ensure safety and to provide transparency to users during updates of algorithms [86]. However, these may not be routinely implemented. While ML-based CDS tools can assist in automated detection, classification, or reporting, safeguards are essential to proactively mitigate errors that may arise from these complex systems. Professional societies such as the American College of Radiology, along with several other U.S. and international radiology organizations, have released an exemplar consensus and guidance document on the importance of developing ethical

standards for ML [89]. Toolkits, such as the American College of Radiology’s AI-LAB framework, that promote a vendor-neutral approach to develop algorithms based on patient populations [90] may also allow for extensibility of algorithms.

5.3 Opportunities and Implications for Precision Oncology

Precision medicine utilizes characteristics of individual patients, but relies on populations of patients to actually guide treatment decisions. Even when high-quality data are available, experts may disagree about interpretation and actionability applied within a clinical setting [91]. Perhaps more problematic is the bias that can arise when extrapolating findings to ethnic, racial, or age groups not well-represented in research cohorts [92, 93]. Despite the growing application of genomic testing in research and precision oncology, use cases are still limited by lack of availability of codified data elements from molecular testing on tumor samples, germline DNA, and serum biomarkers, even for healthcare systems with advanced EHRs and cancer research efforts [94].

Currently, there are numerous open-source cancer genomics tools and ML training platforms that have the potential to accelerate cancer informatics research, such as OncoSim, OncoWiki, and Google Collaboratory [95, 96]. The NCI Informatics Technology for Cancer Research Program has a number of open-access tools that support the analysis of genomic, imaging, and clinical data [97]. Yet, despite the momentum in precision oncology discoveries and actionability of results, their application to the context of the treatment selection for individual patients remains an open challenge in accelerating the implementation of precision oncology.

6 Conclusions

Bringing informatics innovations within digital health, machine learning, and precision oncology into cancer care will require thoughtful approaches to operationalize the collection and meaningful summarization

of disparate data sources to actualize the promise of cancer informatics. Moving from patient engagement in collecting and sharing health data with care teams and researchers, to the delivery of precision cancer care necessitates leveraging informatics innovations. The research highlighted in this survey paper reflects the fast-paced, ever-evolving field, and its challenges as we move discoveries into cancer care.

Acknowledgements

Ashley Griffin is supported by NLM training grant 5T15LM012500-03.

References

- SIMPRO Study Team. Electronic symptom management implementation of patient reported outcomes in oncology research; 2019. Available from: <https://www.esyncancermoonshot.org> [Accessed 14 Nov 2019].
- U.S. National Cancer Institute. NCI and the Precision Medicine Initiative; 2017. Available from: <https://www.cancer.gov/research/areas/treatment/pmi-oncology> [Accessed 11 Nov 2019].
- U.S. Food and Drug Administration. Digital Health. Available from: <https://www.fda.gov/medical-devices/digital-health> [Accessed 14 Nov 2019].
- Chung AE, Jensen RE, Basch EM. Leveraging emerging technologies and the “Internet of Things” to improve the quality of cancer care. *J Oncol Pract* 2016;12(10):863-6.
- Ariza-Garcia A, Lozano-Lozano M, Galiano-Castillo N, Postigo-Martin P, Arroyo-Morales M, Cantarero-Villanueva I. A web-based exercise system (e-CuidateChemo) to counter the side effects of chemotherapy in patients with breast cancer: randomized controlled trial. *J Med Internet Res* 2019;21(7):e14418.
- Jessup DL, Glover Iv M, Daye D, Banzi L, Jones P, Choy G, et al. Implementation of digital awareness strategies to engage patients and providers in a lung cancer screening program: retrospective study. *J Med Internet Res* 2018;20(2):e52.
- Sedrak MS, Salgia MM, Decat Bergerot C, Ashing-Giwa K, Cotta BN, Adashek JJ, et al. Examining public communication about kidney cancer on Twitter. *JCO Clin Cancer Inform* 2019;3:1-6.
- Chen L, Wang X, Peng TQ. Nature and diffusion of gynecologic cancer-related misinformation on social media: analysis of tweets. *J Med Internet Res* 2018;20(10):e11515.
- Taylor J, Pagliari C. The social dynamics of lung cancer talk on Twitter, Facebook and Macmillan.org.uk. *NPJ Digit Med* 2019;2:51.
- Cho H, Silver N, Na K, Adams D, Luong KT, Song C. Visual cancer communication on social media: an examination of content and effects of #melanomasucks. *J Med Internet Res* 2018;20(9):e10501.
- Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. *JAMA* 2017;318(2):197-8.
- Roth AJ, Kornblith AB, Batel-Copel L, Peabody E, Scher HI, Holland JC. Rapid screening for psychologic distress in men with prostate carcinoma: a pilot study. *Cancer* 1998;82(10):1904-8.
- National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology: Distress Management. Available from: https://www.nccn.org/professionals/physician_gls/default.aspx [Accessed 17 Nov 2019].
- Richardson L, Jones G. A review of the reliability and validity of the Edmonton Symptom Assessment System. *Curr Oncol* 2009;16(1):55.
- Girgis A, Durcinoska I, Koh ES, Ng W, Arnold A, Delaney GP, et al. Development of health pathways to standardize cancer care pathways informed by patient-reported outcomes and clinical practice guidelines. *JCO Clin Cancer Inform* 2018;2:1-13.
- Chung AE, Shoenbill K, Mitchell SA, Dueck AC, Schrag D, Bruner DW, et al. Patient free text reporting of symptomatic adverse events in cancer clinical research using the National Cancer Institute’s Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). *J Am Med Inform Assoc* 2019;26(4):276-85.
- Gresham G, Hendifar AE, Spiegel B, Neeman E, Tuli R, Rimel BJ, et al. Wearable activity monitors to assess performance status and predict clinical outcomes in advanced cancer patients. *NPJ Digit Med* 2018;1:27.
- Gupta A, Stewart T, Bhulani N, Dong Y, Rahimi Z, Crane K, et al. Feasibility of wearable physical activity monitors in patients with cancer. *JCO Clin Cancer Inform* 2018;2:1-10.
- Chung IY, Jung M, Lee SB, Lee JW, Park YR, Cho D, et al. An assessment of physical activity data collected via a smartphone app and a smart band in breast cancer survivors: observational study. *J Med Internet Res* 2019;21(9):13463.
- Savova GK, Danciu I, Alamudun F, Miller T, Lin C, Bitterman DS, et al. Use of natural language processing to extract clinical cancer phenotypes from electronic medical records. *Cancer Res* 2019;79(21):5463-70.
- Gonzalez-Hernandez G, Sarker A, O’Connor K, Savova G. Capturing the patient’s perspective: a review of advances in natural language processing of health-related text. *Yearb Med Inform* 2017;214-17.
- Savova GK, Tseytlin E, Finan S, Castine M, Miller T, Medvedeva O, et al. DeepPhe: a natural language processing system for extracting cancer phenotypes from clinical records. *Cancer Res* 2017;77(21):e115-e118.
- Guan M, Cho S, Petro R, Zhang W, Pasche B, Topaloglu U. Natural language processing and recurrent network models for identifying genomic mutation-associated cancer treatment change from patient progress notes. *JAMIA Open* 2019;2(1):139-49.
- Qiu JX, Yoon H, Fearn PA, Tourassi GD. Deep learning for automated extraction of primary sites from cancer pathology reports. *IEEE J Biomed Health Inform* 2018;22(1):244-51.
- Observational Health Data Sciences and Informatics OMOP Common Data Model. Available from: <https://www.ohdsi.org/data-standardization/the-common-data-model/> [Accessed 16 Dec 2019].
- Seneviratne MG, Banda JM, Brooks JD, Shah NH, Hernandez-Boussard TM. Identifying cases of metastatic prostate cancer using machine learning on electronic health records. *AMIA Annual Symp Proc* 2018;2018:1498-504.
- Zhang S, Grave E, Sklar E, Elhadad N. Longitudinal analysis of discussion topics in online breast cancer community using convolutional neural networks. *J Biomed Inform* 2017;69:1-9.
- Lee YJ, Kamen C, Margolies L, Boehmer U. Online health community experiences of sexual minority women with cancer. *J Am Med Inform Assoc* 2019;26(8-9):759-66.
- Temporal History of Your Medical Events (THYME) Corpus. Available from: https://clear.colorado.edu/TemporalWiki/index.php/Main_Page [Accessed 18 March 2020].
- Styler WF, Bethard S, Finan S, Palmer M, Pradhan S, de Groen PC, et al. Temporal annotation in the clinical domain. *Trans Assoc Comput Linguist* 2014(2):143-54.
- Informatics for Integrating Biology & the Bedside (i2b2). NLP Research Datasets; 2020. Available from: <https://www.i2b2.org/NLP/DataSets/Main.php> [Accessed 19 March 2020].
- Johnson AEW, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, et al. MIMIC-III, a free accessible critical care database. *Sci Data* 2016(3).
- Laajala TD, Murtojärvi M, Virkki A, Aittokallio T. ePCR: an R-package for survival and time-to-event prediction in advanced prostate cancer, applied to real-world patient cohorts. *Bioinformatics* 2018;34(22):3957-9.
- DREAM Challenges. Available from: <http://dream-challenges.org/> [Accessed 18 Dec 2019].
- Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med* 2019;25(6):954-61.
- Yala A, Lehman C, Schuster T, Portnoi T, Barzilay R. A deep learning mammography-based model for improved breast cancer risk prediction. *Radiology* 2019;292(1):60-6.
- Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542(7639):115-8.
- Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann Oncol* 2018;29(8):1836-42.
- Wei JW, Tafe LJ, Linnik YA, Vaickus LJ, Tomita N, Hassanpour S. Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks. *Sci Rep* 2019;9(1):3358.
- Walsh S, de Jong EE, van Timmeren JE, Ibrahim A, Compter I, Peerlings J, et al. Decision support systems in oncology. *JCO Clin Cancer Inform*

- 2019(3):1-9.
41. Senaras C, Niazi MKK, Sahiner B, Pennell MP, Tozbikian G, Lozanski G, et al. Optimized generation of high-resolution phantom images using cGAN: application to quantification of Ki67 breast cancer images. *PLoS One* 2018;13(5):e0196846.
 42. Guan S, Loew M. Breast cancer detection using synthetic mammograms from generative adversarial networks in convolutional neural networks. *J Med Imaging (Bellingham)* 2019;6(3):031411.
 43. AlQuraishi M. AlphaFold at CASP13. *Bioinformatics* 2019;35(22):4862-5.
 44. Zhan T, Rindtorff N, Betge J, Ebert MP, Boutros M. CRISPR/Cas9 for cancer research and therapy. *Semin Cancer Biol* 2019;55:106-19.
 45. Poudel P, Nyamundanda G, Patil Y, Cheang MCU, Sadanandam A. Heterocellular gene signatures reveal luminal-A breast cancer heterogeneity and differential therapeutic responses. *NPJ Breast Cancer* 2019;5(1):21.
 46. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med* 2018;24(10):1559-67.
 47. American College of Radiology. FDA Cleared AI Algorithms. Available from: <https://www.acrdsi.org/DSI-Services/FDA-Cleared-AI-Algorithms> [Accessed 26 March 2020].
 48. Wiedmeier JE, Noel P, Lin W, Von Hoff DD, Han H. Single-cell sequencing in precision medicine. *Cancer Treat Res* 2019;178:237-52.
 49. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med* 2017;23(6):703-13.
 50. Coyne G, Takebe N, Chen A. Defining precision: the precision medicine initiative trials NCI-MPACT and NCI-MATCH. *Curr Probl Cancer* 2017;41(3):182-93.
 51. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res* 2019;47(D1):D941-d7.
 52. My Cancer Genome. Available from: <https://www.mycancergenome.org/>. [Accessed 18 March 2020].
 53. Chakravarty D, Gao J, Phillips SM, Kundra R, Zhang H, Wang J, et al. OncoKB: A precision oncology knowledge base. *JCO Precis Oncol* 2017.
 54. Griffith M, Spies NC, Krysiak K, McMichael JF, Coffman AC, Danos AM, et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat Genet* 2017;49:170-4.
 55. HemOnc.org. Available from: https://hemonc.org/wiki/Main_Page [Accessed 17 March 2020].
 56. Levit LA, Kim ES, McAneny BL, Nadauld LD, Levit K, Schenkel C, et al. Implementing precision medicine in community-based oncology programs: three models. *J Oncol Pract* 2019;15(6):325-9.
 57. Remon J, Dienstmann E. Precision oncology: separating the wheat from the chaff. *ESMO Open* 2018;3(e000446).
 58. Harris L, Chen A, O'Dwyer P, Flaherty K, Hamilton S, McShane L, et al. Update on the NCI-Molecular Analysis for Therapy Choice (NCI-MATCH/EAY131) precision medicine trial. ACR-NCI-EORTC International Conference: Molecular Targets and Cancer Therapeutics. Philadelphia, PA; 2018.
 59. Cutler D. Early returns from the era of precision medicine. *JAMA Forum* 2019.
 60. Schachter J, Ribas A, Long GV, Arance A, Grob JJ, Mortier L, et al. Pembrolizumab versus ipilimumab for advanced melanoma: final overall survival results of a multicentre, randomized, open-label phase 3 study (KEYNOTE-006). *Lancet* 2017;390(10105):1853-62.
 61. Kazandjian D, Gong Y, Keegan P, Pazdur R, Blumenthal GM. Prognostic value of the lung immune prognostic index for patients treated for metastatic non-small cell lung cancer. *JAMA Oncol* 2019;5(10):1481-5.
 62. Townsend MH, Shrestha G, Robison RA, O'Neill KL. The expansion of targetable biomarkers for CAR T cell therapy. *J Exp Clin Cancer Res* 2018;37(1):163.
 63. Yu W, Gwinn M, Dotson WD, Green RF, Clyne M, Wulf A, et al. A knowledge base for tracking the impact of genomics on population health. *Genet Med* 2016;18(12):1312-4.
 64. U.S. National Cancer Institute. Toward Precision Cancer Surveillance. Available from: <https://surveillance.cancer.gov/blog/home-page> [Accessed 14 Dec 2019].
 65. Nguyen MH, Smets EM, Bol N, Loos EF, van Laarhoven HW, Geijsen D, et al. Tailored web-based information for younger and older patients with cancer: randomized controlled trial of a preparatory educational intervention on patient outcomes. *J Med Internet Res* 2019;21(10):e14407.
 66. Cohen J. Google, Ascension data partnership sparks federal probe. *Modern Healthcare*. 2019. Available from: <https://www.modernhealthcare.com/information-technology/google-ascension-data-partnership-sparks-federal-probe> [Accessed 14 Dec 2019].
 67. Kosman J. Google's \$2.1B acquisition of Fitbit faces DOJ probe. *New York Post*. 2019. Available from: <https://nypost.com/2019/12/10/googles-2-1b-acquisition-of-fitbit-faces-doj-probe-sources/> [Accessed 15 Dec 2019].
 68. Farr C. Apple and Aetna are teaming up on a new app to help track and reward healthy behavior. *CNBC*. 2019. Available from: <https://www.cnbc.com/2019/01/28/apple-aetna-team-up-on-attain-health-tracking-app.html> [Accessed 15 Dec 2019].
 69. Farr C. Apple and Eli Lilly are studying whether data from iPhones and Apple Watches can detect signs of dementia. *CNBC* 2019. Available from: <https://www.cnbc.com/2019/08/07/apple-eli-lilly-studying-if-iphones-apple-watches-can-spot-dementia.html> [Accessed 15 Dec 2019].
 70. Reuters. Fitbit in healthcare partnership to take on Apple Watch. 2019. Available from: <https://www.reuters.com/article/us-fitbit-health/fitbit-in-healthcare-partnership-to-take-on-apple-watch-idUSKBN1WW263> [Accessed 15 Dec 2019].
 71. Fitbit. Fitbit and Humana expand strategic partnership to drive healthy habits and prevent and manage chronic disease. 2018. Available from: <https://investor.fitbit.com/press/press-releases/press-release-details/2018/Fitbit-and-Humana-Expand-Strategic-Partnership-to-Drive-Healthy-Habits-and-Prevent-and-Manage-Chronic-Disease/default.aspx> [Accessed 15 Dec 2019].
 72. Walker A. NHS gives Amazon free use of health data under Alexa advice deal. *The Guardian*; 2019. Available from: <https://www.theguardian.com/society/2019/dec/08/nhs-gives-amazon-free-use-of-health-data-under-alexa-advice-deal> [Accessed 15 Dec 2019].
 73. Study reveals how well consumers feel health data is protected. *HIPAA Journal*. 2019. Available from: <https://www.hipaajournal.com/study-reveals-how-well-consumers-feel-health-data-is-protected/> [Accessed 15 Dec 2019].
 74. Etaati L. Azure Databricks. In: Etaati L, editor. *Machine Learning with Microsoft Technologies: Selecting the Right Architecture and Tools for Your Project*. Berkeley, CA: Apress; 2019. p. 159-71.
 75. Bisong E. Google AutoML: Cloud Vision. In: Bisong E, editor. *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*. Berkeley, CA: Apress; 2019. p. 581-98.
 76. Amazon Web Services. Machine Learning on AWS. 2019. Available from: <https://aws.amazon.com/machine-learning/> [Accessed 20 Dec 2019].
 77. Strubell E, Ganesh A, McCallum A. Energy and policy considerations for deep learning in NLP. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy; 2019. p. 3645-50.
 78. HealthITAnalytics. Epic systems: machine learning is the EHR usability solution. 2017. Available from: <https://healthitanalytics.com/features/epic-systems-machine-learning-is-the-ehr-usability-solution> [Accessed 15 Dec 2019].
 79. Charette RN. 450,000 women missed breast cancer screenings due to "algorithm failure." *IEEE Spectrum* 2018. Available from: <https://spectrum.ieee.org/riskfactor/computing/it/450000-woman-missed-breast-cancer-screening-exams-in-uk-due-to-algorithm-failure> [Accessed 4 Dec 2019].
 80. Ross C, Swetlitz I. IBM's Watson supercomputer recommended 'unsafe and incorrect' cancer treatments, internal documents show. *STAT* 2018. Available from: <https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/> [Accessed 8 Nov 2019].
 81. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 2019;366(6464):447-53.
 82. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. *JAMA* 2017;318(6):517-8.
 83. Hitaj B, Ateniense G, Perez-Cruz F. Deep models under the GAN: information leakage from collaborative deep learning. *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. Dallas, Texas; 2017. p. 603-18.
 84. Price WN. Big data and black-box medical algorithms. *Sci Transl Med* 2018;10(471):eaa05333.
 85. Shah NH, Milstein A, Bagley P, Steven C. Making machine learning models clinically useful. *JAMA* 2019;322(14):1351-2.
 86. U.S. Food and Drug Administration. Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD). 2019. Avail-

- able from: <https://www.fda.gov/media/122535/download> [Accessed 11 Nov 2019].
87. U.S. Food and Drug Administration. Clinical decision support software guidance for industry and FDA staff; 2019. Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-decision-support-software> [Accessed 11 Nov 2019].
 88. U.S. Food and Drug Administration. Digital health software precertification (Pre-Cert) program; 2019. Available from: <https://www.fda.gov/medical-devices/digital-health/digital-health-software-precertification-pre-cert-program> [Accessed 11 Nov 2019].
 89. Geis JR, Brady AP, Wu CC, Spencer J, Ranschaert E, Jaremko JL, et al. Ethics of artificial intelligence in radiology: summary of the Joint European and North American Multisociety Statement. *J Am Coll Radiol* 2019;16(11):1516-21.
 90. Allen B, Agarwal S, Kalpathy-Cramer J, Dreyer K. Democratizing AI. *J Am Coll Radiol* 2019;16(7):961-3.
 91. Saltz LB. Precision oncology giveth and precision oncology taketh away. *Lancet Oncol* 2019;20(4):464-6.
 92. Gopishetty S, Gupta S, Kota V, Guddati AK. Analysis of racial distribution amongst patients in phase III cancer clinical trials. *J Clin Oncol* 2019;37(15_suppl):6588.
 93. Vo KT, Parsons DW, Seibel NL. Precision medicine in pediatric oncology. *Surg Oncol Clinics* 2019;29(1):63-72.
 94. Burnett-Hartman AN, Udaltsova N, Kushi LH, Neslund-Dudas C, Rahm AK, Pawloski PA, et al. Clinical molecular marker testing data capture to promote precision medicine research within the cancer research network. *JCO Clin Cancer Inform* 2019;3:1-10.
 95. Schoenborn P, Osborne R, Toms N, Johnstone K, Milsom C, Muneer R, et al. OncoSim and OncoWiki: an authentic learning approach to teaching cancer genomics. *BMC Med Educ* 2019;19(1):407.
 96. Bisong E. Google Collaboratory. In: Bisong E, editor. *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*. Berkeley, CA: Apress; 2019. p. 59-64.
 97. National Cancer Institute. *Informatics Technology for Cancer Research*; 2020. Available from: <https://itcr.cancer.gov/informatics-tools> [Accessed 18 March 2020].

Correspondence to:

Arlene E. Chung, MD, MHA, MMCI
 5034 Old Clinic Building, CB 7110
 Chapel Hill, NC 27599-7110
 USA
 Tel: +1 919 966 2276
 Fax: +1 919 966 2274
 E-mail: arlene_chung@med.unc.edu