Sharing Clinical and Genomic Data on Cancer - The Need for Global Solutions

mary care practices for this demonstration, which qualifies as an Alternative Payment Model under the Medicare Access and CHIP Reauthorization Act (MACRA).

Calls for fundamental payment and EHR reform are likely to trigger considerable discomfort and resistance. Many practice managers and payers and some clinician groups will mourn the loss of work documentation and process-based payment afforded by current EHRs and fee-for-service payment. Their substantial investment of time and money in the current systems and ability to benefit financially from them are not to be underestimated. These groups argue that paying for outcomes could be achieved with less disruption by simply grafting some value-based payments and penalties onto a fee-for-service model. This type of evolutionary approach is embodied in the Merit-Based Incentive Payment option of MACRA. Although it is superficially appealing and encourages cost containment, this so-called pay-for-value plan relies even more heavily on EHR documentation and rewarding of achievement on process measures. It does little to reduce documentation burden or correct the misplaced emphases on care processes and service volume.

Delivery from EHR purgatory is within our reach if we as a profession have the will to take more responsibility for outcomes. Not all outcomes are under our control, but focusing on those we can reasonably affect represents a genuine opportunity to restore meaning to our daily work and the social contract we have with our patients and society.

1. Wachter R. The digital doctor: hope, hype, and harm at the dawn of medicine's computer age. New York: McGraw-Hill, 2015.
2. Sinsky C, Colligan L, Li L, et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. Ann Intern Med 2016;165:753-60.
3. Mendelson A, Kondo K, Damberg C, et al. The effects of pay-for-performance programs on health, health care use, and processes of care: a systematic review. Ann Intern Med 2017;166:341-53.
4. Goroll AH, Berenson RA, Schoenbaum SC, Gardner LB. Fundamental reform of payment for adult primary care: comprehensive payment for comprehensive care. J Gen Intern Med 2007;22:410-5.
5. Bitton A, Schwartz GR, Stewart EE, et al. Off the hamster wheel? Qualitative evaluation of a payment-linked patient-centered medical home (PCMH) pilot. Milbank Q 2012; 90:484-515.

Sharing Clinical and Genomic Data on Cancer — The Need for Global Solutions

The Clinical Cancer Genome Task Team of the Global Alliance for Genomics and Health

The Cancer Moonshot initiative has emphasized the importance of breaking down data silos to create a comprehensive and effective "cancer knowledge network" that would accelerate the combining of genomic, epidemiologic, and clinical information to improve patient outcomes. The real value of genomic data will be realized only when they are linked to high-quality, longitudinal, computationally amenable clinical information, allowing researchers to identify precise genotype–phenotype associations. If we don't concentrate our efforts (and dedicate substantial resources) to robustly improve data sharing, we risk undermining precision oncology's capacity to deliver substantive advances for people with cancer. Patients are demanding that their data be shared, and funding agencies are specifying that institutions provide broad access to research-generated information. Cancer doesn't respect national borders, so we need effective global strategies for sharing cancer-related data. But getting to that point presents various challenges.

Some barriers are technical. The lack of effective methods for extracting data from electronic health records (EHRs) has made it difficult to obtain relevant clinical information for data amalgamation. Incompatible data formats and a shortage of interoperable data-harmonizing informatics tools also compromise researchers' ability to mine multiple data sets. Finally, the absence of a single standardized cancer ontology (a machine-readable set of defined descriptors of clinical manifestations) limits the ability to capture clinical data and retards cross-study data analysis, a central requirement for a cancer knowledge network.

Legal, regulatory, and ethical barriers also pose daunting chal-

lenges for effective data sharing, particularly across borders. In the European Union, the General Data Protection Regulation has raised concerns among cancer researchers[1] because of its potential to undermine patient-oriented collaborative research, international data-sharing efforts, and the clinical applications of new treatments — although lobbying by the biomedical research community has led to a more research-friendly regulation than was originally proposed.[2] Ethical considerations related to cross-border data sharing sometimes require asking patients to complete additional consent forms because of concerns about individual patient identification. The lack of consistent regulatory policies further compromises the ability to share data among countries.

The Global Alliance for Genomics and Health is an international coalition of over 470 stakeholder organizations from more than 60 countries that aims to develop interoperable solutions that promote sharing of high-quality genomic and clinical data. Part of its responsibility is to respond to these challenges, enhancing the deployment of cancer-related data for patient benefit. We developed the Framework for Responsible Sharing of Genomic and Health-Related Data to reflect the rights of all people to benefit from scientific advances. Now translated into 12 languages, this framework provides research consortia with robust policies, tools, and adapted consent procedures that respect patient autonomy while supporting international data-sharing practices.

An additional challenge is financial: in a recent survey of more than 100 cancer-sequencing ini-

tiatives worldwide, we identified lack of affordability as the most substantial barrier to effective data sharing.[3] Annotation and validation of clinical and epidemiologic data remain expensive and time consuming, and until the community commits more resources to these efforts, clinical utility will be hampered.

Despite these challenges, a number of national initiatives

have made tangible progress (see table), which suggests that long-term solutions can be achieved. The National Cancer Institute Genomic Data Commons[4] has made available more than 4 petabytes of genomics data — a valuable resource for comprehensive data mining that could unmask previously hidden associations between genomic variations and cancer. The U.S.-based DiscoverEHR cohort study, which linked exon sequences from 50,000 people to EHR data (with patients' consent), revealed that 3.5% of patients harbor actionable mutations in 1 of 76 genes related to cancer or cardiovascular disease.[5] The 100,000 Genomes Project is generating data from whole-genome sequencing to inform clinical decision making regarding rare diseases and cancer in the United Kingdom; similar projects are under way in France and the Netherlands.

But more progress is needed to stimulate global advances, particularly to facilitate data sharing between countries. We believe an ideal model would be one in which data were stored by the originating hospitals or research institutions, rather than in a centralized repository, but could be retrieved and analyzed by members of the global cancer-research community using widely applica-

An ideal model would be one in which data were stored by the originating institutions, rather than in a centralized repository, but could be retrieved and analyzed by members of the global research community using widely applicable but secure methods.

ble but secure methods. Such a system would provide assurance of data privacy and security for individual institutions or consortia while improving access for outside researchers.

Key enablers of this federated data-ecosystem model have been so-called container technologies, such as Docker, which allow users to package genomics-analysis tools and pipelines in a portable format in order to perform "onsite" data analysis on diverse computing platforms. Interoperability is further enhanced by the Global Alliance's application programming interface, which facilitates exchange of genomic information from next-generation sequencing reads among multiple organizations and on multiple platforms.

Warehousing and retrieving federated data using cloud-based solutions is increasingly possible, through both commercial entities and initiatives such as the Na-

Selected Cross-Institutional Data-Sharing Initiatives with a Cancer Component or Focus.*	
Initiative	Description
National	
National Cancer Institute Genomic Data Commons	A knowledge system to store, analyze, and distribute cancer genomics research data
CancerLinQ	An American Society of Clinical Oncology (ASCO)–led data informatics system that collects clinical data from EHRs; ASCO plans to incorporate genomic data in the future
Geisinger MyCode Community Health Initiative	A precision medicine program that uses next-generation sequencing to identify potentially actionable mutations in human disease (including cancer)
Cancer Genome Collaboratory	A Canadian initiative to make data available for cloud computing in a community cloud infrastructure
Electronic Medical Records and Genomics (eMERGE)	A national network combining DNA biorepositories with EHR systems for high-throughput genetic research
Oncology Research Information Exchange Network (ORIEN)	A coalition of North American cancer centers that combines genomic, clinical, and epidemiologic information to inform disease diagnosis, prognosis, and treatment
Precision Medicine Exchange Consortium (PMEC)	A consortium that is aggregating genomic and clinical data to develop precision medicine approaches to human disease
100,000 Genomes	A national whole-genome-sequencing initiative that reports results back to the National Health Service
International	
International Cancer Genome Consortium for Medicine (ICGCMed)	A follow-up to the International Cancer Genome Consortium, linking genomic sequencing data to longitudinal clinical data
BRCA Exchange	A curated catalogue of *BRCA1* and *BRCA2* variants and corresponding clinical data
Genomics Evidence Neoplasia Information Exchange (GENIE)	An American Association for Cancer Research–enabled transatlantic initiative to integrate genomic profiles and longitudinal clinical data and release them to the cancer-research community
Cancer Core Europe (CCE)	A pan-European consortium sharing a common translational genomic platform to conduct next-generation clinical trials

* EHR denotes electronic health record.

tional Cancer Institute Cancer Genomics Cloud Pilots and the European Helix Nebula project. Cloud-enabled solutions are generally compatible with country- or region-specific legal frameworks and also deliver economic value, with user costs about 1/10 those of academic-based high-performance computing. The recent launch of the European Open Science Cloud provides a potential framework for open-source data sharing, and the Innovative Medicines Initiative, through its Big Data for Better Outcomes program, has prompted companies with oncology drug-development programs to share their data in the early

stages of the research-and-development process.

Several important cross-border projects have been developed by the cancer-research community in response to the international data challenge. BRCA Exchange has mobilized researchers, clinicians, and patient advocates worldwide and created a curated Web portal for sharing data on the pathogenicity of expert-reviewed *BRCA1* and *BRCA2* mutations, allowing clinicians, patients, and testing laboratories to understand the significance of any individual variant. Although this approach has successfully linked high-quality genomic and clinical data, its

scope is limited to information at the individual-gene level; a long-term goal is to extend this model to other genes related to cancer.

Cancer Core Europe involves six premier cancer centers in various European countries and is based on the model of a virtual cancer institute. The centers share a harmonized data infrastructure and conduct cooperative biomarker-driven clinical trials in a catchment population of more than 60,000 patients — a model that supports clinical utility.

The Genomics Evidence Neoplasia Information Exchange project, supported by the American Association for Cancer Research,

takes this concept one step further, with a transatlantic data-sharing cooperative. During the project's first year, clinical-grade genomic data and baseline clinical information from about 19,000 patients at eight major cancer centers in the United States, Canada, and Europe were harmonized using a common data dictionary for recording tumor subtypes; the data were then made publicly available. Longitudinal data from subgroups of these patients are being collected in order to establish genotype-specific disease registries for use in clinical care. This federated approach to longitudinal clinical data collection overcomes institutional concerns related to sharing medical record data.

The Cancer Moonshot has captured the imagination of researchers, patients, and the public. Putting its ideas into practice will be possible only through the use of an interoperable, scalable data framework in which the quality of data is maintained. Despite the efforts outlined above, this work is still in the early stages.

Even with substantial enthusiasm for data sharing and an ever-expanding volume of genomic data sets, the inability to routinely correlate longitudinal clinical information with precise genomic data within a secure and acceptable framework continues to hamper the development of innovative data-driven care pathways for patients with cancer — as does uncertainty regarding the financial sustainability of international data sharing. We believe creating a global informatics ecosystem in which precision oncology seamlessly transitions from cancer diagnosis to molecular discovery to patient recovery must be our common goal.

The Clinical Cancer Genome Task Team of the Global Alliance for Genomics and Health includes Mark Lawler, Ph.D., Centre for Cancer Research and Cell Biology, Queen's University Belfast, Belfast, United Kingdom; David Haussler, Ph.D., University of California, Santa Cruz Genomics Institute, Santa Cruz; Lillian L. Siu, M.D., Princess Margaret Cancer Centre, University of Toronto, Toronto; Melissa A. Haendel, Ph.D., and Julie A. McMurry, Ph.D., Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University, Portland; Bartha M. Knoppers, Ph.D., Centre for Genomics and Policy, McGill University, Montreal; Stephen J. Chanock, M.D., Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; Fabien Calvo, M.D., Ph.D., Cancer Core Europe and Institute Gustave Roussy Cancer Campus, Grand Paris, Villejuif, France; Bin T. The, M.D., National Cancer Center, Singapore; Guneet Walia, Ph.D., Bonnie J. Addario Lung Cancer Foundation, San Francisco; Ian Banks, M.D., European Cancer Organization Patient Advocacy Committee, Brussels; Peter P. Yu, M.D., Hartford HealthCare Cancer Institute, Hartford, CT; Louis M. Staudt, M.D., Ph.D., Center for Cancer Research, National Cancer Institute, Bethesda, MD; and Charles L. Sawyers, M.D., Memorial Sloan Kettering Cancer Center, New York.

1. Casali PG. Risks of the new EU Data Protection Regulation: an ESMO position paper endorsed by the European oncology community. Ann Oncol 2014;25:1458-61.
2. Dove ES, Thompson B, Knoppers BM. A step forward for data protection and biomedical research. Lancet 2016;387:1374-5.
3. Vis DJ, Lewin J, Liao RG, et al. Towards a global cancer knowledge network: dissecting the current international cancer genomic sequencing landscape. Ann Oncol 2017;28:1145-51.
4. Grossman RL, Heath AP, Ferretti V, et al. Toward a shared vision for cancer genomic data. N Engl J Med 2016;375:1109-12.
5. Dewey FE, Murray MF, Overton JD, et al. Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. Science 2016;354:aaf6814.