In the world of bioinformatics, the rush is on to extract gold from a data mine.

The amount of data that health care providers and scientists collect from patients and research participants is growing explosively. This information ranges from the genetic to laboratory tests and imaging exams, to medical histories and information about treatment and outcomes - and in some cases to survey data on large populations.

This deluge of data and the bioinformatics capabilities necessary to take advantage of it were the focus of this year's daylong UCSF School of Medicine leadership retreat on January 20. In his welcome address, Dean Sam Hawgood, MBBS, outlined his goal for the day - to engage campus leaders in the question of how to optimally develop, organize and integrate clinical-outcome data, research data, business intelligence, and population data so that information is accessible and usable to empower research and improve medical practice.

Making Greater Use of Available Data to Improve Care

Some have compared efforts to take advantage of the data to trying to drink from a fire hose. Consider DNA as an example. Our genetic variations contain clues to disease risk, disease prognosis and treatment response. The identification of such clues by scientists and their translation into medical practice is a major enterprise. Soon, expense will no longer be a major limitation to obtaining a readout of an individual's entire genetic makeup. "A complete human genome assay will be an assay like any other, at least in terms of cost and time," Hawgood said.

However, as vivid as the water hose analogy may be, it might be more apt to say that in many organizations data of many types dwells in various unconnected dammed reservoirs, with little of it flowing to potential users.

Much of the day's discussions centered not only on identifying data to collect and on ways to use this data, but also on how best to "unlock the data" that already exists.

In introducing the day's theme, Hawgood said that in his view, UCSF, despite the depth and range of clinical and research data being collected, must develop ways to make much greater use of available data in the day-to-day workflow to improve research and patient care.

UCSF's leaders in informatics can learn from what others have already done. "We want to take enough time to make sure that we're not repeating other people's mistakes," Hawgood said.

Already this year UCSF is completing implementation of a new electronic medical records system, called APeX, tailored to UCSF by EPIC Systems Corporation of Madison, Wisconsin. The system features a single, comprehensive record for each patient.

Apart from being a boon to physicians and other care providers, information within the new clinical record can be "de-identified" to protect privacy and then made available to researchers.

The experience gained from the implementation and the system itself may be a good jumping off point for using data to advance research and education. But the true potential of APeX as a platform to support collaborative research is still being investigated. Taking full advantage of UCSF medical records for research - and enabling ways for new research findings to be used to better guide patient care - will require new innovations.

UCSF already has convened a task force, soliciting input from an external team of international leaders
in the field, to explore strategies to bolster bioinformatics on campus, including the establishment of new academic programs and infrastructure through which computer sciences faculty and other bioinformatics experts could be recruited, new experts trained, and novel research collaborations launched. In addition, Hawgood noted, thanks to UCSF's proximity to Silicon Valley, "We are in a spectacular region for partnerships.”

UCSF is exploring how to accomplish its bioinformatics goals within UCSF and its affiliates, as well as ways in which to access and use data in research networks that span institutions. In addition, there is a need to share information with other providers to provide the best care to patients who also obtain health care outside of UCSF, a theme discussed at last year's School of Medicine retreat as well.

**Using Hospital Systems as Living Laboratories**

In his keynote address at this year's retreat, "Aligning the Academic Health Care Enterprise for Acceleration of Precision Medicine," Isaac Kohane, MD, PhD, a renowned bioinformatics expert and a professor of pediatrics and health sciences and technology at Harvard Medical School, said that the expense of working with data from clinical records has historically been much more expensive than working with genetic and molecular laboratory data.

Kohane, who co-directs the Harvard Medical School Center for Biomedical Informatics, has led efforts to develop computer systems to allow cheaper use of clinical records data from multiple hospital data systems in the study of genes and disease, while maintaining privacy.

Kohane talked about the potential for "apps" to capture useful information related to health outside of the hospital or clinic - for instance a tool that tallies nutritional information on purchases from the cash register.

In his own research Kohane now combines clinical and genomic data to learn more about cancer and autism, but he also presented research showing that such systems - had they been in place earlier - could have called attention to serious side effects of Vioxx and other drugs much sooner.

Kohane's described workflows and systems that can better "unlock" clinical data to speed research discovery and its application in medical practice, while lowering the costs of using clinical data.

**Imagining UCSF’s Future in the Digital Age**

Speakers at a morning panel titled "Imagining UCSF's Future in the Age of Information Technology," included Opinder Bawa, chief technology officer for the School of Medicine; Michael Blum, MD, medical director of information technology for the UCSF Medical Center; Catherine Lucey, MD, vice dean for education; Joe DeRisi, PhD, co-chair of the Department of Biochemistry and Biophysics; and moderator Clay Johnston, MD, PhD, director of the Clinical and Translational Science Institute at UCSF and vice chancellor for research.

Not all bioinformatics applications are orchestrated institution-wide from the top down. Panelsists and commenters from the audience highlighted a role for applications developed by smaller groups. For instance, as an educational tool, UCSF faculty from the Department of Emergency Medicine have spearheaded implementation of software used by physicians-in-training to respond to simulated clinical scenarios unfolding in real time.

The data collected during these exercises can be used to better understand how long it is likely to take for emergency room physicians to take critical actions - in the management of chest pain or in the ordering of pain medication, for example. In essence, the data can be used to learn more about how we learn, knowledge that can be incorporated into successive generations of teaching tools.

In the coming years, physicians will be trained to become increasingly comfortable using improved, data-driven, decision-making software tools, according to Lucey. "We will be able to teach students how to learn for themselves for 30 to 40 years, and we will free faculty up to teach in person what needs to be taught in person."
graduate school applicants is being used to identify factors associated with future success. In addition, he described how extensive data collection is being used to log graduate students' progress and decision-making throughout their careers - another way of identifying early career paths that bode well for future success. DeRisi also talked about an information-technology partnership that allows lab notebook entries made on pad devices to be immediately incorporated into a computerized database.

Developing More Innovative Research, Clinical Protocols

An afternoon panel - "The Future Is Now" - moderated by Robert Hiatt, MD, PhD, co-chair of the Department of Epidemiology and Biostatistics and deputy director of the UCSF Helen Diller Family Comprehensive Cancer Center, focused on two large-scale collaborative research programs co-led by UCSF researchers in which molecular, clinical and demographic data already are being put to work to develop more innovative research and clinical protocols.

Laura van't Veer, PhD, leader, and Laura Esserman, MD, co-leader of the cancer center's breast oncology program, described the ATHENA Breast Health Project, which unites UC academic medical centers in a state-wide collaboration. The project will initially involve 150,000 women throughout California who will be screened for breast cancer and followed for decades.

ATHENA project leaders aim to create common systems to integrate clinical research and care across the UC campuses to advance the science of prevention, screening, diagnosis, and treatment of breast cancer. The collaborators are creating a biospecimen repository that has broad racial and ethnic representation. A major goal is to marshal molecular and clinical data to better personalize breast care, tailoring treatment to the patient and avoiding overtreatment, and to use the information gained to drive innovation in prevention, diagnosis and treatment.

Neil Risch, PhD, co-chair of the Department of Epidemiology and Biostatistics at UCSF and director of the UCSF Institute for Human Genetics, along with Catherine Schaefber, PhD, director of the Kaiser Permanente Research Program on Genes Environment and Health, described progress to date in building the largest data base of its kind to focus on genetic variation and environmental exposures in an older population.

The average age of the hundreds of thousands of individuals whose genetic information will be genotyped for the project is 65. The project's foundation is Kaiser's electronic health record, which for many Kaiser Permanente members has information spanning decades - including information on clinical diagnosis and treatment as well as lab-test results and prescription information. UCSF expertise has allowed extraordinarily fast genotyping, as well as uniquely large-scale analysis of telomeres to quickly grow the molecular component of the data resource.

The afternoon panel provided a useful point of reference for break-out groups that met afterward, charged with identifying institutional priorities, problems and potential solutions in advancing the use of bioinformatics in research, clinical care and education.

For instance, the ATHENA collaborators have standardized protocols used at the different medical centers, including protocols for mammography screening. To make clinical data more useful for research, some breakout session panelists advocated more extensive standardization of clinical imaging and clinical lab protocols and reporting throughout UCSF clinical practices.

The Kaiser-UCSF collaboration highlights ways to combine strengths across organizations. While Kaiser is famous as a health maintenance organization, UCSF is perhaps best known as a tertiary care center. Some breakout session panelists raised the question - also raised at last year's retreat - of whether or not the focus of UCSF research should more closely reflect the patient population seen at UCSF and its affiliated medical centers. UCSF specialists routinely gather extensive information on large numbers of patients with serious acute and chronic conditions, including many of the most difficult - to-treat cases. This
extensive data is a potential gold mine for research aimed at identifying factors related to disease risk, prognosis and treatment outcomes.

**Moving Toward A New Taxonomy of Disease**

The retreat followed on the heels of a similarly themed report by the National Academy of Sciences (NAS), "Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease" [PDF] by a committee co-led by UCSF Chancellor Susan Desmond-Hellmann, MD, MPH. The NAS committee advocated the creation of a "knowledge network" that could link researchers in collaborations that span the nation and globe.

The NAS panel envisioned a future in which there is much greater use of genomic and other molecular data to improve and refine the classification of diseases, but also recognized an opportunity to improve research by more extensively taking into account information about how patients fare in the clinic or hospital.

Instead of the current state of affairs, through which biological, pre-clinical and clinical research eventually lead to advances in medical practice, the new paradigm will be a virtuous cycle through which - with appropriate privacy protections - patient data also will feed back into research. In patient care increasing amounts of laboratory information will become available and interpretable more quickly to help teams of caregivers make more accurate and effective decisions and choices in diagnosis, prognosis and treatment for each individual patient.

Provided by University of California, San Francisco


*This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.*