Bringing patient data into the open

By Joanne Kotz, Senior Editor

Most translational science efforts focus on integrating the work of academic and industry scientists, and they leave out a critical set of participants: patients. This is problematic because the speed and quality of the feedback between the bench and the bedside is a critical rate-limiting factor for medical progress.

Two groups have come to the conclusion that breakthroughs in translational medicine require collecting data on patients, including outcomes, on a scale never previously attempted and making those data available with appropriate privacy protections to translational researchers.

In both cases, the key challenges will be getting industry and academic buy-in to data sharing and setting up the legal and logistical infrastructures.

Sage Bionetworks launched a portal this month through which users can contribute their own health and genomic data for research using a newly developed legal framework. Separately, a report from the U.S. National Academy of Sciences (NAS) is calling for the creation of a national infrastructure for accessing and analyzing open-source patient data.¹

Pooling patient data to enhance the value of treatment is not a new concept. Indeed, a handful of pharmas and payers already are collaborating to share patient information to improve clinical trial design and better evaluate a drug's benefit to patients.

For example, in February 2011, AstraZeneca plc announced a collaboration with the HealthCore Inc. outcomes research unit of WellPoint Inc. to analyze the effectiveness of marketed therapies to identify gaps in which new medicines are needed. In June 2011, Sanofi partnered with Medco Health Solutions Inc. to identify subpopulations of patients with the highest medical need to inform patient selection during clinical testing.²

The Sage and NAS efforts are seeking to move the concept to an open-access model in which patient data are collected with appropriate legal and privacy safeguards and pooled anonymously in an open database to aid biomedical research, drug discovery and ultimately clinical care (see Figure 1, "Open-access model for patient data sharing").

"Having come back to academia from industry three years ago, I have been surprised, actually, how little overlap there is between the world of research and the world of clinical care. I think that's to the detriment of patients and what we are trying to accomplish in translational research. With advances like electronic health records and low-cost, high throughput DNA sequencing, there is an opportunity to take advantage of routine episodes of clinical care for biomedical and clinical research. Yet most of the data are not collected, not pooled or never connected to the explosion of molecular data," said Susan Desmond-Hellmann, chancellor of the University of California,

San Francisco and co-chair of the committee that wrote the NAS report.

Desmond-Hellmann was previously president of product development at Genentech Inc., which is now part of Roche.

Although the NAS report is nominally about the need to modernize the taxonomy of disease, its real focus is creating a learning healthcare system. This would be achieved by building a national informatics infrastructure over the course of years to decades to make health data and molecular information obtained from individual patients during routine office visits openly available to inform both biomedical research and patient care.

Patient health information would come from electronic health records. Molecular information that could be collected for patients includes genome, epigenome, metabolome and microbiome data.

As for Sage, director John Wilbanks said the purpose of the not-for-profit's proposed project is to put "a lot of data together, get it clean and let the researchers of the world start running on it to make new connections."

These two efforts "represent a great step forward with considerable potential to be highly informative and impactful. By having a resource available that amalgamates genomic, biosensor, phenotypic and other critical data for a very large population of individuals, the whole biomedical research process will be markedly accelerated," Eric Topol, director of the Scripps Translational Science Institute and chief academic officer of Scripps Health, told SciBX.

Ultimately, said Desmond-Hellmann, the drivers behind these initiatives are "a need to innovate, to lower the cost of health and to challenge ourselves to tap into the explosion of how people use data to completely change how we think about doing R&D in the life sciences."

Getting to the point

According to Desmond-Hellmann, collecting point-of-care patient information is a key aspect of the proposal and is distinct from information currently collected in clinical trials.

Collecting point-of-care information would provide patient data that are more reflective of "real life, including all of us in the community—not a contrived situation," she said. "Even in large clinical studies currently conducted, the numbers of patients are relatively small and miss the power of having huge numbers of patients."

Pfizer Inc.'s David Cox, a member of the committee that wrote the NAS report, told SciBX that longitudinal data with outcomes are critical to advance translational research. "My view is that right now, without access

"With advances like electronic health records and low-cost, high throughput DNA sequencing, there is an opportunity to take advantage of routine episodes of clinical care for biomedical and clinical research. Yet most of the data are not collected, not pooled or never connected to the explosion of molecular data."

**—Susan Desmond-Hellmann,
University of California,
San Francisco**

to this kind of information, the entire pharma industry is severely limited in its ability to make new medicines. When you look at what has driven the ability to make novel therapies, in almost every case it is longitudinal clinical outcome data in clinical samples. Clinical trial data, which is more of a snapshot, is critical, but not sufficient.”

Cox is SVP and CSO of Pfizer’s Applied Quantitative Genotherapeutics Unit.

Once the molecular and health information is assembled, the NAS committee envisions integrating analytical and visualization tools to improve disease classification, facilitate more personalized clinical care and catalyze biomedical research.

The report cited an ongoing UCSF–**Kaiser Permanente** study as proof of concept for collecting additional molecular patient data during the course of routine care as an opt in that is paid for via research funding.

Under the project, which is funded by an NIH grant of about \$25 million, Kaiser patients can elect to have their DNA genotyped. To date, the partners have genotype information for about 200,000 patients. The genetic information is then integrated with self-reported health information, electronic health records and California environmental data and held anonymously in a database—separately from patients’ health records—for analysis by UCSF researchers.

The integrated information enables researchers to look at the natural history of disease. “We can go backwards in time and look at people who have acquired diseases and look at their genetic information, at what medications they were on, at risk behaviors, etc.,” said Desmond-Hellmann.

Two new pilot studies are proposed in the NAS report: “The Million American Genomes Initiative” and “Metabolomic Profiles in Type 2 Diabetes.” No specific steps have been taken to implement them, and how such efforts would be financed has not been decided, said Desmond-Hellmann.

The virtue of patients

Meanwhile, Sage is launching a study in which individual users can anonymously share their health and genomic data for biomedical research.

The first step was creating a legal framework, dubbed the Portable Legal Consent (PLC), which allows users to give broad rights for the use of data on themselves for research purposes. Only a few restrictions apply that create some protections for participants against discrimination and that require open access to publications resulting from the research.

Figure 1. Open-access model for patient data sharing. Two groups are hoping to make patient health and outcome data—and associated genomic and molecular information—openly available to facilitate translational research.

Sage Bionetworks launched a portal through which users can contribute their own health and genomic data, and a report from the U.S. **National Academy of Sciences** is calling for the creation of a national infrastructure for accessing and analyzing open-source patient data.

In these models, health and molecular data is collected from clinical trials and routine health visits [a] using a legal framework that protects privacy and ensures appropriate downstream uses [b]. The data from individuals [c] are then de-identified [d] and made available for academic and industry research, as well as ultimately to inform clinical care [e].

Next up is testing the concept. In a study called “Portable Legal Consent for Common Genomics Research (PLC-CGR),” individuals will be able to deposit data on themselves on the trial’s website, including electronic health records and genomic and lifestyle information. Sage received IRB approval for the study in April and began enrolling participants this month.

Wilbanks told *SciBX* that patients also could submit data from sites such as **PatientsLikeMe** or directly transfer information from clinical trials in which they are enrolled.

He hopes to have 25,000 participants enrolled within a year, a number roughly equal to the size of the Framingham Heart study, and ultimately to have data on a million individuals.

Wilbanks added that some pharmas and biotechs already are considering using PLC as a consent protocol in pilot trials. In this scenario, a company would run a clinical trial as usual but include an option for participants to upload their individual data to the PLC-CGR site.

“PLC-CGR is making a bold assertion here—that informed consumers can provide portable consent, allowing them to assign consent to researchers rather than for consent be taken from them in individual studies,” said Paul Wicks, R&D director at PatientsLikeMe.

PatientsLikeMe previously ran a study exploring why patients with amyotrophic lateral sclerosis (ALS) chose not to participate in a biomarker study that sought to understand the cause of their disease.³

“Many patients felt they didn’t need to donate blood because during their diagnostic workup they had already had blood taken several times, and these patients assumed that their blood would also be used for research to find a cure. In fact when we told these patients that, no, we would need to take new samples and get separate consent for each new study, they were dissatisfied with the inefficiency of the system,” Wicks said.

Opening up

The primary challenges to the NAS and Sage efforts will be obtaining enough data submissions to enable meaningful analysis of compatible data and convincing academics, payers, health maintenance organizations and biotech and pharma companies to openly share their data.

There are data analysis approaches that “simply can’t work until there’s vast amounts of data. To be blunt, neither Google nor Facebook would make a change to an advertising algorithm with a sample set as small as that used in a Phase III clinical trial. Sage’s goal is to get the sample sizes for clinical data closer to those we use in consumer systems,” said Wilbanks.

He also said completeness of the data will be paramount. For example, in Sage’s study, “if we get a million people enrolled but each only uploads one kind of data—a genotype here, a medical record there—it probably won’t be as useful as getting 25,000 people to upload their genotype, a personal health record and ongoing lifestyle data,” he said.

Ultimately, the success of these efforts will depend heavily on sharing, said Desmond-Hellmann, “and I think it’s hard for academia and it’s hard for industry.

Topol agreed. “The barriers will not likely be most individuals who are asked to provide consent but rather the researchers who are used to keeping the data in their own domain.”

“I think pharma is also in a fantastic position to benefit from” the efforts, noted Desmond-Hellmann. “I would point to what Vertex did in cystic fibrosis by partnering with the **Cystic Fibrosis Foundation**. Can you imagine scouring the earth for 4% of patients with a rare disease? Impossible. That link to the kind of databases that the CF foundation has built up over the years was essential.”

Vertex Pharmaceuticals Inc.’s Kalydeco, which was approved in January, targets a mutation found in about 4% of patients with CF. Vertex took the compound from IND to market in just over five years. The CF Foundation’s registry of patients, which contains genetic information and medical histories, and the foundation’s clinical trial network were key factors in Kalydeco’s rapid development path.

“If we get a million people enrolled but each only uploads one kind of data—a genotype here, a medical record there—it probably won’t be as useful as getting 25,000 people to upload their genotype, a personal health record and ongoing lifestyle data.”

—John Wilbanks, Sage Bionetworks

“What I think is really key is increasingly thinking about intellectual property and authorship as being driven by intellectual contribution, not by who owns the database,” said Desmond-Hellmann. “I think that patients and patient advocacy groups can drive some of these collaborative behaviors.”

She added: “A lot of different disease-oriented patient advocacy groups, who are increasingly getting into venture philanthropy, are really pushing that if you’d like funding, for example, from the **Multiple Myeloma Research Foundation** (MMRF), you need to share your data.”

In July 2011, the MMRF launched CoMMpass, a 5-year study of 1,000 newly diagnosed MM patients that will connect multiple types of genomic profiling with longitudinal clinical data. The study is supported in part by a precompetitive consortium of industry partners, and the initial data will be made openly available when the portal is launched near the end of this year. Following the portal launch, data will be made available first to members of the precompetitive consortium for five months, then to study investigators for one month and then openly to the clinical and research community every six months throughout the study.

Louise Perkins, CSO of MMRF, told *SciBX* the foundation is building a research portal with tools and interfaces to allow clinicians and researchers to “interact with the data and explore clinical and biological hypotheses.”

One difficulty the MMRF has encountered is the issue of data compatibility. Even genomic information collected by different research teams is not necessarily readily compatible. “Some groups are analyzing 100 genes, some are sequencing exomes, some are only sequencing parts of the exome,” said Perkins.

Perkins said bringing together multiple kinds of molecular information for multiple diseases will not be straightforward. “When one starts talking about collecting data across diseases it becomes very challenging to see the light at the end of the tunnel,” she said.

Kotz, J. *SciBX* 5(25); doi:10.1038/scibx.2012.644

Published online June 21, 2012

REFERENCES

1. Committee on a Framework for Developing a New Taxonomy of Disease. [Toward precision medicine: building a knowledge network for biomedical research and a new taxonomy of disease](#). (2011)
2. Bouchie, A. *BioCentury* 19(32), A9–A10; Aug. 1, 2011
3. Bedlack, R.S. *et al. Amyotroph. Lateral Scler.* 11, 502–507 (2010)

COMPANIES AND INSTITUTIONS MENTIONED

AstraZeneca plc (LSE:AZN; NYSE:AZN), London, U.K.
Cystic Fibrosis Foundation, Bethesda, Md.
Genentech Inc., South San Francisco, Calif.
Kaiser Permanente, Oakland, Calif.
Medco Health Solutions Inc. (NYSE:MHS), Franklin Lakes, N.J.
Multiple Myeloma Research Foundation, Norwalk, Conn.
National Academy of Sciences, Washington, D.C.
PatientsLikeMe, Cambridge, Mass.
Pfizer Inc. (NYSE:PFE), New York, N.Y.
Roche (SIX:ROG; OTCGX:RHHBY), Basel, Switzerland
Sage Bionetworks, Seattle, Wash.
Sanofi (Euronext:SAN; NYSE:SNY), Paris, France
Scripps Health, San Diego, Calif.
Scripps Translational Science Institute, La Jolla, Calif.
University of California, San Francisco, Calif.
Vertex Pharmaceuticals Inc. (NASDAQ:VRTX), Cambridge, Mass.
WellPoint Inc. (NYSE:WLP), Indianapolis, Ind.