Contemporary Clinical Trials 31 (2010) 394–404

Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

Discussion

Creating a global rare disease patient registry linked to a rare diseases biorepository database: Rare Disease-HUB (RD-HUB) ☆

Yaffa R. Rubinstein [a,*,1], Stephen C. Groft [a,2,3], Ronald Bartek [b], Kyle Brown [c], Ronald A. Christensen [d], Elaine Collier [e], Amy Farber [f], Jennifer Farmer [b], John H. Ferguson [a], Christopher B. Forrest [g,h], Nicole C. Lockhart [i], Kate R. McCurdy [j], Helen Moore [i], Geraldine B. Pollen [a], Rachel Richesson [k], Vanessa Rangel Miller [l], Sara Hull [m], Jim Vaught [i]

[a] Office of Rare Diseases Research National Institutes of Health, Bethesda, MD, United States
[b] Friedreich's Ataxia Research Alliance (FARA), Springfield, VA, United States
[c] Innolyst, Inc, San Mateo, CA, United States
[d] REGISTRAT-MAPI, Scottsdale, AZ, United States
[e] National Center for Research Resources, National Institutes of Health, Bethesda, MD, United States
[f] LAM Treatment Alliance, Cambridge, MA, United States
[g] Children's Hospital of Philadelphia, Philadelphia, PA, United States
[h] Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA, United States
[i] Office of Biorepositories and Biospecimen Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, United States
[j] Barth Syndrome Foundation, Inc, Larchmont, NY, United States
[k] Department of Pediatrics, Division of Bioinformatics and Biostatistics, University of South Florida College of Medicine, Tampa, FL, United States
[l] Department of Human Genetics, Emory University, Atlanta, GA, United States
[m] Office of the Clinical Director, National Human Genome Research Institute, Department of Bioethics, Clinical Center, National Institutes of Health, Bethesda, MD, United States

ARTICLE INFO

Article history:
Received 25 June 2010
Accepted 28 June 2010

Keywords:
Rare diseases
Patient registry
Disease registry
Rare disease
Patient advocacy
Biospecimen
Biospecimen repositories
Clinical data
Electronic health record

ABSTRACT

A movement to create a global patient registry for as many as 7,000 rare diseases was launched at a workshop, "Advancing Rare Disease Research: The Intersection of Patient Registries, Biospecimen Repositories, and Clinical Data." http://rarediseases.info.nih.gov/PATIENT_REGISTRIES_WORKSHOP/. The workshop was sponsored by the Office of Rare Diseases Research (ORDR). The focus was the building of an infrastructure for an internet-based global registry linking to biorepositories. Such a registry would serve the patients, investigators, and drug companies. To aid researchers the participants suggested the creation of a centralized database of biorepositories for rare biospecimens (RD-HUB)http://biospecimens.ordr.info.nih.gov/ that could be linked to the registry. Over two days of presentations and breakout sessions, several hundred attendees discussed government rules and regulations concerning privacy and patients' rights and the nature and scope of data to be entered into a central registry as well as concerns about how to validate patient and clinician-entered data to ensure data accuracy. Mechanisms for aggregating data from existing registries were also discussed. The attendees identified registry best practices, model coding systems, international systems for recruiting patients into clinical trials and novel ways of using the internet directly to invite participation in research. They also speculated about who would bear ultimate responsibility for the informatics in the registry and who would have access to the information. Hurdles associated with biospecimen

☆ Authors listed 3rd to18th alphabetically contributed equally.

* Corresponding author. Office of Rare Diseases Research National Institutes of Health, 6100 Executive Boulevard, Room 3A07, MSC-751, Bethesda, MD 20892, United States. Tel.: +1 301 402 4338; fax: +1 301 480 9655.
E-mail addresses: rubinsty@mail.nih.gov (Y.R. Rubinstein), gropts2@od.nih.gov (S.C. Groft).

[1] Director of Patient Resources for Clinical and Translational Research, Office of Rare Dieseases Reseach.

[2] Director of the Office of Rare Diseases Research, National Institutes of Health, 6100 Executive Boulevard, Room 3A07, MSC-751, Bethesda, MD 20892, United States. Tel.: +1 301 435 6041.

[3] Co-first author.

1551-7144/$ – see front matter. Published by Elsevier Inc.
doi:10.1016/j.cct.2010.06.007

collection and how to overcome them were detailed. The development of the recommendations was, in itself, an indication of the commitment of the rare disease community as never before.

Published by Elsevier Inc.

Meeting report

Executive summary

A movement to create a federated global patient registry containing core data and using a standardized vocabulary for as many as 7000 rare diseases was launched at a workshop, "*Advancing Rare Disease Research: The Intersection of Patient Registries, Biospecimen Repositories, and Clinical Data.*" http://rarediseases.info.nih.gov/PATIENT_REGISTRIES_WORKSHOP/.

The workshop, which was held in Bethesda, MD, on January 11–12, was sponsored by the Office of Rare Diseases Research (ORDR), the National Eye Institute, and the National Center for Research Resources of the National Institutes of Health (NIH), as well as patient advocacy groups and the private sector. The focus was the building of an infrastructure for an internet-based federated global registry and the linking of the registry to biorepositories. Such a registry would serve rare disease patients and their advocacy groups seeking help and information, investigators conducting research, clinicians treating patients, epidemiologists analyzing disease data, and drug companies exploring new markets. To aid researchers the participants suggested the creation of a centralized database of biorepositories for rare biospecimens (RD-HUB) that could be linked to the registry.

Over two days of presentations and breakout sessions, several hundred attendees who represented advocacy groups, researchers, clinicians, information technology (IT) experts, government and the private sector discussed government rules and regulations concerning privacy and patients' rights and the nature and scope of data to be entered into a central registry as well as concerns about how to validate patient and clinician-entered data to ensure data accuracy and timeliness. Mechanisms for aggregating data from existing registries were also discussed. The attendees identified registry best practices, model coding systems, international systems for recruiting patients into clinical trials and novel ways of using the internet directly to invite participation in research. They also speculated about who would bear ultimate responsibility for the informatics in the registry and who would have access to the information. Hurdles associated with biospecimen collection and how to overcome them were detailed, as well as success stories about gene discoveries and biomarkers for rare diseases.

IT experts explained to the workshop attendees that however complex the demographic information and clinical data to be entered into a common resource may be, including the use of multiple languages for global access, the infrastructure problems would not be an impediment. However, the establishment of rare disease HUB (RD-HUB), http://biospecimens.ordr.info.nih.gov/, for biospecimen repositories as well as for common diseases, poses formidable challenges, not only in questions of control, access, bioethics, and privacy, but also technically, in terms of developing standardized protocols to ensure optimal specimen collection and preservation.

For many, the workshop was an exhilarating experience of interaction and exchange of information across groups that rarely meet. What they shared and learned from each other has already fostered collaborations and ongoing activities. Next steps will involve arriving at a consensus on what core data elements should be collected in the central registry, how to harmonize information across different datasets, and how to resolve issues of control and access. Development of recommendations and next steps by workshop members was, in itself, an indication of the commitment and enthusiasm that is uniting the rare disease community as never before. The enthusiasm of the attendees over the course of the meeting was unflagging. There was a sense that now is the time to embark on such an ambitious agenda, for good reasons.

Workshop proceedings

For two days in January several hundred attendees at a government-sponsored meeting in Bethesda, Maryland, addressed the tantalizing possibility of creating a centralized registry of rare diseases—a kind of giant umbrella registry whose spokes would represent individual patient registries for any one of several thousand rare diseases. Beyond economies of administration and scale, such a centralized system would serve patients and families seeking accurate information about their own and related conditions; investigators conducting rare disease research (who might serendipitously discover links among diseases); clinicians treating patients; epidemiologists gathering demographic data; and, interestingly, the drug and device industry seeking new markets. Also on the agenda, primarily to support research, would be the creation of a centralized database of biorepositories for biospecimens from rare disease patients, linked to the registry that might also shed light on etiology and pathogenesis of rare diseases.

But the challenges to achieving such resources are many. Who would be in control of a centralized database? What kinds of data would be stored and how would accuracy be assured and updates provided? Who would have access? What guarantees would be in place to protect patient privacy and deal with sundry other bioethical concerns? The establishment of biospecimen repositories poses equally formidable challenges, not only in questions of control, access, bioethics, and privacy, but also technically, in terms of developing standardizing protocols to ensure optimal specimen collection and preservation.

- The power of the Internet. Diverse health and disease sites have burgeoned on the Internet, but *caveat emptor*: they vary in accuracy, reliability, and timeliness. Rare disease foundations and advocacy groups, as well as government agencies are also well represented. Many are already experienced in generating, maintaining, and controlling

access to patient databases and could model what might be possible for a unified database.

- The "ome" revolutions. Beginning with the mapping and sequencing of the human and other genomes in the 1990s, and adding the tools and techniques that are generating genome-wide association studies and libraries of metabalomes and proteomes, the prospects of discovering the gene/molecular underpinnings of rare diseases are more promising than ever. Further, as the cost of sequencing individual genomes goes down, scientists envision an era of "personalized medicine" when more and more people will obtain their own genetic maps, enabling them to identify their status as carriers or at risk for selected diseases. These advances in science have sparked the formation of many a biotech company and emboldened rare disease advocacy groups who first and foremost seek safe and effective diagnoses, treatments, and ways of prevention. A centralized rare diseases database—as well as biospecimen repositories—would be invaluable resources for research and help resolve the debilitating, often life-threatening health problems that beset rare disease patients and agonize their families.
- The pharmaceutical industry. Traditionally, drug companies have spurned development of drugs for rare diseases, considering the market too small to be profitable. Indeed, their lack of enthusiasm led to the passage of the Orphan Drug Act in 1983, which offered tax incentives for clinical trials of orphan products and a 7-year exclusive right to market any product designed to treat a rare (or orphan) disease, defined as one affecting fewer than 200,000 patients in America. But the pharmaceutical industry is changing, seeing potential profits in expensive drugs for rare diseases, with the expectation that the drugs might require lifetime use. The long and costly process of drug discovery and testing might well be accelerated and costs reduced if drug companies were given access to biospecimens and a central rare diseases database, which could be a source of volunteers for clinical trials.

These developments in science, information technology, and industry, as well as the continued growth of advocacy groups, have altered the playing field for rare disease research, leading the Office of Rare Diseases Research (ORDR) at the National Institutes of Health to organize the meeting, *Advancing Rare Disease Research: The Intersection of Patient Registries, Biospecimen Repositories, and Clinical Data*, in Bethesda on January 11–12, 2010. The objectives of the workshop were to discuss how to build the infrastructure for an Internet-based centralized rare diseases registry and establish biospecimen repositories, gain acceptance from advocacy groups, researchers and other stakeholders in attendance, and make recommendations for follow-up and next steps. ORDR worked with several NIH and non-governmental partners in planning the program, which was co-sponsored by the NIH National Center for Research Resources and the National Eye Institute, several rare disease groups, academic organizations, and information technology firms.[3] Day One was devoted to plenary sessions exploring the issues entailed in creating centralized resources. Day Two used breakout sessions to develop recommendations and next steps.

Following welcoming remarks outlining the goals for the meeting and encouraging broad audience participation, Day One began with presentations by two rare disease advocates. Dr. Amy Farber, representing the LAM Treatment Alliance, spoke movingly of the death of a friend's mother from a rare cancer. Questions about the mother's care could have been answered had there been access to clinical data that was "out there," but inaccessible. Quoting President Obama's comments on the would-be Christmas Day bomber, she saw this as "a failure to collect and understand the intelligence we have." She compared this to the situation of patients with lymphangioleiomyomatosis (LAM), a rare multi-system disorder that fatally compromises the lungs, leading to a suffocating death. Here too information was "out there"—as many as 26 patient databases—but it was uncoordinated and inaccessible. Vanessa Rangel Miller was equally emphatic on the need to improve communication of the intelligence we have, a need that could be met through a common rare disease registry. Uniting rare diseases would promote collaboration and information sharing, facilitate research, and enable participating groups to learn from each other, she said. As a representative of DuchenneConnect, a coalition of partners that includes patients with either Duchenne or Becker muscular dystrophy, she saw a common rare disease registry as complementing, but not replacing the registries individual advocacy groups maintain.

There was a question from the audience following the talks if a common disease registry should be national or international. The registry will need to begin locally, ORDR Director Steve Groft said, but he and others agreed that eventually it should be a global effort, since diseases "know no borders" and also because of increasing collaborations between U.S. researchers and international partners. One speaker added that the idea of a common rare disease registry seemed to be "on every health minister's mind."

While the initial talks addressed the attributes and utility of a common registry in broad terms, the second round of speakers explored the structure and organization of a common registry in more detail.

Pediatrician Christopher Forrest from Children's Hospital of Philadelphia described two models for a central rare disease registry. One would be a system in which categorical disease groups would feed core registry data into an Internet-based site with governance and access distributed across the individual players. Alternatively, there could be a centralized model in which individual groups would provide de-identified data[4] into what could become a global hub. Again, usage would be governed by the data providers. With approval, external users such as life sciences companies and government agencies could obtain access. Forrest agreed that the time was ripe for creating unified systems. Networking has

[3] Specifically, the National Organization for Rare Disorders, the Muscular Dystrophy Association, eyeGENE, The Transverse Myelitis Association, Emory Genetics Laboratory, Innolyst, Genetic Alliance, Maya V. Nathan Breath of Life Foundation, American College of Medical genetics, SWF, and REGISTRAT-MAPI.

[4] The issue of de-identifying data is non-trivial. It has arisen in published research, primarily on clinical trials, in which authors are requested to have available the dataset underpinning the findings, so that readers can verify results. It appears that even an attempt to remove obvious identifiers may not be enough to protect the privacy of individuals. One proposal is define a dataset as that "containing the minimum level of detail necessary to reproduce all numbers reported in a paper."

become *the* means of knowledge sharing and communication, he noted, and he saw the move to unified systems as consonant with the newer paradigms of science as holistic and integrative rather than reductionist.

But there are hurdles to be overcome. Initially, advocacy groups need to be persuaded that "there is something in it for me," Forrest said. While attendees at the meeting were generally enthusiastic, there are many fledgling advocacy groups who operate on shoestring budgets and lack a patient registry, as well as patients who are not yet organized into advocacy groups. When it comes to creating a *centralized* registry, questions inevitably arise as to what elements should be included and how they can be standardized so the data from different disease registries are compatible.

Forrest illustrated the difficulties with respect to standards by citing a one-day record from Children's Hospital in Philadelphia in which clinicians entering data into electronic medical records (EMRs) used 278 ways to describe fever for 465 patients, 123 different ways to express ear pain in 213 patients, and 99 different ways to describe red ears. Thus, while EMRs (or EHRs—electronic *health* records) are touted as the next big thing in health care, much remains to be done in terms of establishing common data-entry codes and vocabularies as well as ensuring interoperability across the different electronic record systems that have been developed.

Information technology (IT) experts in the U.S. and abroad have been addressing these and other issues in the course of developing national models for the communication of health information on the Internet. Dr. Dan Russler, Vice President for Clinical Informatics at Oracle Health Sciences, discussed work in progress in the U.K., Canada, and the U.S. The U.K. uses a centralized system called "Spine" that stores a Personal Care Record for enrollees in the National Health Service. The record includes personal/demographic data (e.g., name and address, age, sex) and clinical information, such as medications used and allergies. The system also stores de-identified patient information for research analysis and planning. Individuals use a card and a personal identification number (PIN) to access the Spine Directory Service on the Internet, which, via an Access Control Framework, then determines what information the user can see. Canada uses a province-by-province system, which also stores demographic and clinical information for provincial residents enrolled in the National Health Service, a decentralization made necessary in part because privacy legislation in Canada is provincially rather than nationally based.

In contrast to the U.K. and Canada, the currently evolving U.S. National Health Information Network (NHIN) is a completely decentralized system with all personal and clinical information on patients held by individual Health Information Organizations, (HIOs). An HIO can be a federal agency, a health community, a pharmacy network, or other health-related entity, whether for-profit or not-for-profit. If an HIO contracts to join the network, it agrees to a set of standards, specifications, protocols, legal procedures, and services that will enable it to exchange secure health information with other NHIN members over the Internet. Network members control access to their own data and require participants using their data to address all applicable legal requirements before disclosing any data. Security measures were built into the design of the system at the start, using encryption, audits, and a series of authentication and authorization steps to prevent misuse. Patients can also indicate their preferences concerning the sharing of any identifiable medical data.

Russler cited several instances of how NHIN is currently being used. For example, the Centers for Disease and Control and Prevention (CDC) uses special software to connect to NHIN to collect summarized biosurveillance data (information that may indicate naturally occurring or intentional disease outbreaks) from State Health Departments. He concluded with suggestions for how a centralized rare disease registry might be adapted to any of the national health communication models he discussed. In the U.K., permission might be sought to use patient information from a centralized rare disease registry through the Ethics and Confidentiality Committee of the National Information Governance Board for Health and Social Care; in Canada, via negotiation with each province for inclusion in the system; and in the U.S., by establishing the rare disease central registry as an HIO member of NHIN.

Comments from the audience at this point agreed on the complexity of issues and the need for standards, but also questioned how desirable it was to share data across unrelated registries and how to get clinicians and researchers on board, either because of time constraints or competitive attitudes with regard to sharing data.

Standards, informatics, and technology

Creating a mega database of rare diseases requires sophisticated IT architecture and security measures as well as agreements on the elements to be included and how they will be standardized. The panel of speakers which addressed these issues was reassuring that technology would not prove to be a stumbling block—IT is capable of developing the hard- and software to handle massive databases with appropriate levels of security. But the business of determining standards for data elements is fraught with so many choices—even in the ways simple demographic data like names and addresses are recorded—that Kyle Brown, CEO of the IT company Innolyst, (and a partner in DuchenneConnect) said it may take "a seismic shift" to get individual groups to commit to the effort and agree on standardized terminology.

Brown went on to provide some perspective on what a centralized rare disease database might encompass. He observed that currently there are:

- 7000 rare diseases recorded by ORDR
- 30 million people affected
- 4200 patients per rare disease on average
- 1300 patients per rare disease registry (assuming 30% register)

Thus the IT challenge is how to build a common infrastructure to handle the unique needs of 7000 rare diseases and 9 million registrants and do this in a multi-language format. His experience working with patient groups has taught him that they first have to be reassured that the cost of developing a registry will not be a deterrent. "The technology is easy, getting organizations to agree on standards is hard." he says. As a further incentive, if groups with limited resources could make use of a standardized format for data entry, they might more readily commit to creating patient registries and

contributing to a centralized registry. This in itself would enrich the store of information available for stakeholders. In turn, Brown thought the central system could be further developed to provide de-identified information aggregated into a searchable clearinghouse of rare disease data. The sticking point is getting to that standardized format.

Two NIH experts, Dr. James Cimino from the Clinical Center's Laboratory for Informatics Development and Technology and Dr. Clement J. McDonald from the National Library of Medicine further elaborated on the issue of standards. Cimino saw the growing amounts of clinical data now being recorded and updated periodically on electronic health records as a gold mine for clinical researchers. But he echoed Forrest's warnings about the multiple ways clinicians enter data as well as the variations in the criteria they use in making diagnoses or judging treatment outcomes. The resulting EHRs, from the researcher's point of view, may be misleading, incomplete, or in error.

To illustrate the lack of standard nomenclature for many disease entities Cimino observed that one clinician might diagnose a patient's condition as celiac disease while others might use the terms nontropical sprue, gluten enteropathy, idiopathic steatorrhea, Gee disease or Gee–Herter disease. Even entering straightforward laboratory findings can vary depending on the way the information is requested. Thus, if the form asks for blood type, the entry could be “A positive” but if the form requests “ABO Panel” the response will be “major A, Rh+.” Cimino concluded that the good news is that there is a trove of clinical information now being collected, updated, and stored electronically. His hope is that with federal incentives to use electronic records, combined with the adoption of data standards, the information collected will be that much more valuable for all users.

Dr. Clement McDonald from the National Library of Medicine provided a kind of tutorial on how to create a registry by outlining the step-by-step decisions needed in designing any sort of health questionnaire or survey instrument, beginning with the choice of data to collect. Next come decisions on whether responses are going to be numeric (in which case the measurement units and ranges need to be specified) or coded (in which case an explicit list of responses—“yes,” “No,” “missing,” and so on—must be given). But there is no need to start from scratch. He urged the audience to search the literature for already existing survey instruments that may be relevant to particular diseases as well as reviewing government questionnaires. Especially useful are those that have been refined and validated over the years, such as the forms developed for the National Health and Nutrition Examination Survey, which the CDC's National Center for Health Statistics uses in its periodic assessments of the nation's health.

Newer instruments relevant to the interests of rare disease groups are PROMIS (Patient-Reported Outcomes Measurement Information Systems), which captures quality-of-life information from clinical studies, based on a patient's report of pain, fatigue, emotional distress, physical functioning, and social role participation. Also of interest is PhenX, a project funded by the National Human Genome Research Institute to facilitate the integration of epidemiologic and genetics research in large genetic studies, such as genome-wide association studies. Essentially the project aims at building a consensus among experts on how to measure environmental variables (such as diet and nutrition) known to interact with genetic factors in giving rise to complex diseases. The idea is to standardize the environmental measures so that data can be compared or aggregated across studies.

With regard to the quest for a lingua franca that could be used in referring to diseases and disorders, signs and symptoms, and other clinical data, McDonald cited a number of coding systems that NLM supports, each with its own acronym, successive versions, and subcategories. LOINC (Logical Observation Identifiers Names and Code) is a coding system recommended by the U.S. and some other governments for use in diagnostic reports, survey instruments, lab tests, and clinical measurements. SNOMED CT (Systematized Nomenclature of Medicine—Clinical terms) represents the merger and expansion of a system used by the College of American Pathologists and the Clinical Terms codes used by the U.K. National Health Service. SNOMED can assign codes for organisms, anatomic parts, specimens, diagnoses, and symptoms, and appears to be the most comprehensive clinical vocabulary available in any language. RxNorm is a more specialized vocabulary that has been developed by NLM to provide a standardized nomenclature for prescription drugs and drug delivery devices. Clearly, there is no dearth of systems that could translate rare disease data into a standardized form for a centralized registry, but coming to consensus on which might work best is another question.

In contrast to the alternative systems available for coding disease terms, McDonald spoke highly of the progress made in standardizing technical data such as blood or urine analyses, radiological findings, and other clinical tests. Standardized codes currently enable the computer transmission of laboratory results to clinics and hospitals throughout the world in spite of wide variation in computer systems used. In this regard McDonald proclaimed, “HL7 version 2.x is king.” The acronym stands for Health Level 7, the 2.x clinical practice version. HL7 is a volunteer not-for-profit international organization of experts who develop global standards and provide the framework for the exchange, integration, sharing and retrieval of electronic health information.

The session on standards and IT closed with presentations of new and ongoing projects that specifically address the need and value of standards in the development of rare disease registries. With funding from the American Recovery and Reinvestment Act, Dr. Rachel Richesson from the University of South Florida described PRISM—the Patient Registry Item Specifications and Metadata project, which aims to establish standardized libraries of data elements and registry questions that can be used by advocacy groups in developing new registries or revising existing ones. The intent is to encode the questions, answers and definitions using data standards that will permit consistent data collection and data sharing.

Dr. Christophe Béroud, representing the French research organization INSERM, described TREAT-NMD (Translational Research in Europe—Assessment and Treatment of Neuromuscular Diseases), a project initially supported by the European Union in 2007, but which has now expanded to include more than 30 countries worldwide. TREAT-NMD was designed to improve care for patients with neuromuscular disease primarily by facilitating the conduct of clinical trials—often a problem in rare disease research because of the small

numbers of patients available for recruitment. TREAT-NMD project directors initially decide on the content of what will comprise a supranational database (or registry) for selected neuromuscular diseases. (Duchenne muscular dystrophy and spinal muscular atrophy were the first diseases chosen.) They then enlist national patient organizations, clinicians, geneticists, researchers, and others as partners to provide the data needed to build the international base, with all due care for the legal and ethical issues involved. National "curators" are trained to collect and validate patient data to create a national database to feed into the supranational base. (A TREAT-NMD toolkit is available as an aid, but countries can use their own systems to build national registries if they choose.) Only encrypted data is used in the supranational database and patients can only be contacted at national levels. Since the advent of the system, Béroud was pleased to report that inquiries have come in from pharmaceutical companies interested in running clinical trials, proving that the system is working.

Biospecimens/biorepositories

Would that could be said for the situation with regard to establishing national, much less supranational rare disease biospecimen repositories. The accounts of the panel of speakers addressing repository issues amounted to a litany of problems ranging from poor collection and preservation techniques, to "silo" attitudes of investigators and institutions unwilling to share samples, to the burden of well-intentioned state and federal legislation to protect patient privacy and access, but which in practice can severely hinder the research enterprise.

Adding to these obstacles Dr. Christopher Moskaluk, a pathologist at the University of Virginia, reminded the audience that the vary rarity of rare diseases means that tissue samples available for study are few and far between. He urged rare disease groups to be pro-active in contacting potential donors and clinicians and he saw the move to create a centralized rare disease registry as a significant step that could facilitate creating national repositories. But such repositories would have to resolve today's serious obstacles. Among them, specimens are usually derived from "leftovers" from surgery or autopsy and may be severely degraded. Even if an institution archives specimens, the preservation technique may not be informative, for example, with regard to biomolecules. Clinical data for the sample, including the stage of disease (early, late) are time-consuming to obtain and may be hard to recover or blocked by regulatory issues.

Dr. Carolyn Compton, who directs the National Cancer Institute's Office of Biospecimens and Biospecimens Research (OBBR), similarly commented on the lack of rigor in tissue collection, not only for rare diseases but for *all* diseases, pointing to variations in tissue processing, data annotation, patient consent forms, access policies, and materials transfer agreements, resulting in wide variation in specimen quality and data. These negatives were among the lessons learned from two NCI projects, The Cancer Genome Atlas (TCGA) and the Clinical Trials Network of Cooperative Groups. Compounding the problem, she also spoke of the lack of tissue sharing among investigators, citing a study in which over half the researchers responding said they got tissue from their own patients or colleagues in the same institution and rarely from outside sources. To resolve the issues, she described caHUB (the cancer Human BioBank), NCI's new project to establish a unique, centralized, non-profit public resource that would develop evidence-based strategies for high-quality tissue collection and well-annotated biospecimens. Importantly, well-characterized normal human tissue will be available for controls. She suggested that NCI could partner with ORDR, and make available the protocols and operating principles of caHUB for use in creating rare disease biospecimen repositories.

The ideal repository, Dr. Benjamin Greenberg, from the Transverse Myelitis and Neuromyelitis Optics Program at the University of Texas Southwest Medical Center, would be one in which we capture tissue samples from a patient over the course of a lifetime—from birth through various environmental exposures to symptoms, diagnosis, treatments, and outcomes. His point was that we don't know the questions that are going to be asked *tomorrow* and we need to take into consideration that disease is the sum of genetics + environment + timing. The difficulties of rare disease repositories are that there are too few numbers, too few entry points, and too little infrastructure for longitudinal studies. To remedy the situation, he proposed finding common ground with larger related diseases, saying that in his case much could be learned by considering the rare demyelenating diseases he studies as outliers in registries and repositories associated with not-so-rare multiple sclerosis.

Greenberg also recommended using patient registries as sources for donors to build tissue repositories—an idea further developed by Jeffrey Thomas, Director of Donor Services for the National Disease Research Interchange (NDRI). The Interchange is a national network of over 100 tissue procurement centers, including eye banks, tissue banks, surgery centers, hospitals, and organ procurement organizations, along with contracted professional recovery specialists. Thomas' point was that with appropriate education and consent forms in place, rare disease patient registries could become major resources for tissue donations, including the kinds of serial donations over time that Greenberg would like. At present NDRI has two special projects involving rare diseases. In collaboration with the Von Hippel-Landau Family Alliance, NDRI is processing and storing buffy coats, plasma and DNA, and with funding from the National Heart, Lung and Blood Institute, the Interchange is acquiring biospecimens from LAM patients during transplant and post mortem.

The Interchange has formed a National Rare Disease Voluntary Health Organization Partnership whose 22 members are building donor registries to facilitate tissue collections. NDRI is also working informally with some 50 other rare disease groups.

Dr. Marsha A. Moses from Harvard Medical School and Children's Hospital Boston, observed that there was an added value in collecting rare disease biospecimens because their study can improve understanding of more common diseases. Over the decade her laboratory has established a human urine bank using samples from patients with rare diseases, including progeria, LAM, chronic pelvic pain syndrome, and vascular anomalies, as well as common diseases, such as cancer, all with age- and sex-matched controls. The urine analyses and relevant clinical data are electronically archived in a searchable database. Her laboratory has shown increased expression of urinary matrix metalloproteins (uMMPs) in

parallel with the extent and activity of vascular anomalies. Indeed, her research indicates that many of the pathological changes seen in rare diseases mirror those that occur in more common diseases. She has used these findings to develop and validate a number of biomarkers for several of the diseases her lab has studied. Some may indicate disease progression, for example; others can be used to measure risk for breast cancer. Experience in the course of building the tissue bank has enabled the laboratory to develop "best practices" for the collection, shipment, and storage of samples. She concluded that the study of rare disease pathology continues to shed light on more common diseases—and vice versa.

A member of the audience asked whether protocols for new biobanks were publicly available, including multiple normal samples from cadavers. Compton said yes, they have gotten permission to make them public and are using new funding to develop the protocols. Others asked about reimbursements for tissue and worried that hospitals remain unwilling to give up samples but rather reserve them for their pathology departments.

Over lunch, keynote speaker University of Pennsylvania bioethicist Dr. Jonathan Moreno, reviewed the growth of bioethics as a discipline, from a time when practitioners were regarded as authority figures imbued with the traditional ethical principles and practices they learned in medical school, to today's emphasis on patients as participants in an ongoing dialogue with their care providers. The result has been a greater exchange of information and truth telling to patients and the advent of informed consent forms and other mandates to protect patients' rights. The strength of rare disease registries lies in their potential to serve as resources for subjects in clinical research, he said. But care must be taken to ensure patient privacy. He saw this issue as especially relevant in the case of rare diseases because the close relationships that often exist among investigators, sponsors, and patients, can blur the lines of communication and the roles that each plays. He cautioned that the risk of patient identifiability will grow with continued mining of genetic information and some of the legal safeguards in place might inhibit researchers' access to human subjects.

Clinical research, patient care, and disease management

With Moreno's talk to set the stage, the afternoon plenary sessions began with a discussion of issues with regard to rare disease clinical research from the point of view of representatives of industry, government agencies, and academia. Dr. Ronald Christensen, co-CEO of the U.S.–French company REGISTRAT-MAPI, explained that the firm provides consultant services to pharmaceutical companies and other clients in the design and statistical analysis of late phase clinical trials, post-marketing studies, epidemiologic, and other clinical research. He noted that passage of the Orphan Drug Act in 1983 has resulted in:

- 339 approved orphan drugs on the market
- About 14 new orphan drug applications approved by FDA every year
- 139 new drugs approved over the decade from 2000 through October 2009

Clearly, the act has had an impact, but Christensen believes that more impressive gains are on the horizon because the pharmaceutical industry is changing and entering the market. Some companies are expanding orphan drug research and also acquiring smaller biotech firms that have developed orphan products. The move toward personalized medicine and increased genomics research is also motivating investment. Other promising indicators are NIH Challenge Grants for rare diseases and registries, and a Therapeutics for Rare & Neglected Disease Program, funded at $24 million in the 2009 NIH budget. All these developments argue for the establishment and maintenance of rare disease patient registries, Christensen said, which not only can serve as a resource for recruitment for new clinical studies, but also as the source for subregistries following particular subjects' responses to other therapeutics. Similar registries are maintained by industry and government agencies, he noted, such as ones used to monitor the post-marketing experience of patients exposed to new drugs or devices, or registries collecting observational data on patients defined by a particular disease or environmental exposure.

The Food and Drug Administration is changing, too, Theresa Toigo, Director of the agency's Office of Special Health Issues declared. To raise awareness of the new and expanded powers of FDA she pointed the audience to the more detailed information available on the agency's re-vamped web site. The critical changes are outlined in the 2007 amendments to the Food, Drug & Cosmetic Act, she said. These authorize FDA to require post-market studies, make safety-labeling changes, and develop "Risk Evaluation and Mitigation Strategies" (essentially guides to medications and their safe use). The amendments also empower FDA to expand its involvement in clinical trial design and ensure that the new, more extensive information about drugs and device trials and trial results are posted on government web sites. Some examples of registries reflecting FDA's new post-marketing and monitoring vigilance include a registry of patients for whom particular blood-enhancing drugs are ineffective because the patients have developed antibodies to the products, and registries of pregnant women exposed to various medications and diseases.

Toigo also described a Sentinel Initiative launched in 2008 to create a national electronic system for monitoring drug safety. It will be a private–public partnership and make use of electronic record systems, insurance claims databases, and other sources to collect product safety information. She ended by describing FDA's role in administering the Orphan Drug Act through its Office of Orphan Products, again inviting the audience to search the web site, which includes plans for two workshops in 2010 to provide guidance in applying for orphan drug designation.

As a representative from the Agency for Healthcare Research and Quality (AHRQ), Jean Slutsky described yet another valuable contribution rare disease registries could make to the scientific enterprise—by playing a role in *effectiveness research* (ER). The term, often mentioned in the debates on healthcare reform, refers to studies that aim to determine what clinical therapies work in the real world—in whom, when, under what circumstances and dosages, and with what risks, costs, and benefits. The "real world" is meant to distinguish actual medical practice in the community, rather than in the case of the carefully selected and monitored patients in a double blind randomly controlled clinical trial. Slutsky's office, AHRQ's Center for Outcomes and Evidence,

has developed a detailed handbook explaining how to create registries for evaluating patient outcomes. A rare disease advocacy group wanting to develop such a registry would need to recruit providers as well as patients, ascertain the quality of data collected at the outset and over the course of a study, record any adverse events, and finally analyze and interpret the data to evaluate the outcomes. AHRQ's investment in ER research now includes regional evidence-based practice centers, education and therapeutic research centers, and other sites. The agency also supports research through a grant program which includes an initiative called Clinical and Health Outcomes in Comparative Effectiveness (CHOICE) and one called Prospective Outcome Systems using Patient-Specific Electronic Data to Compare Tests and Therapies (PROSPECT).

The value of a registry is only as good as the accuracy of the data collected. This was the point emphasized by the last speaker in the session on clinical research, Andy Faucett, from Emory University School of Medicine. His message was that it is critical—and advocacy groups need to make this clear to patients and their families—that the diagnosis and nomenclature used by a physician in assessing a patient reflect a consensus of experts in the field and that any laboratory tests in support of the diagnosis be of the highest standard. Faucett is Program Coordinator for Collaboration Education and Test Translation (CETT), an organization that promotes collaboration among researchers, clinicians, advocacy groups, and clinical laboratories to facilitate the development and translation of tests used to diagnose rare genetic diseases. He explained that research investigators may discover a candidate gene or genes implicated in a rare disease, but the work needs to be replicated and a process developed for translating the findings into a valid and reliable test that commercial clinical laboratories can use to establish a diagnosis. In CETT's programs researchers stand ready to advise clinical labs when new gene variants for a disease are discovered and also help develop and update educational materials. He suggested that the collaborative model for establishing reliable clinical tests is one that rare disease advocacy groups should consider to assure the accuracy of data in their patient registries.

A question raised at the end of this session would surface again in several breakout groups: how can we be sure that registries are truly representative of the patient community and not simply a better-educated and affluent group. Slutsky replied that this is a real challenge and it may mean developing more than one type of registry. Another pondered whether new government-initiated requirements such as post-marketing studies would have advocacy groups collaborating with drug companies.

Patient participation and outreach activities/patient advocacy

The final two plenary sessions of the workshop focused on rare disease advocacy and outreach, and issues concerning privacy protection and human subject research, beginning with presentations by two organizations representing multiple advocacy groups. NORD, the National Organization for Rare Diseases, Inc., was founded over 25 years ago by the mother of children diagnosed with Tourette syndrome. Speaking for the organization, Dr. Sukirti Bagal explained that NORD acts as a clearinghouse for rare disease information and that from the outset has been committed to facilitating drug development and helping patients obtain treatments. (NORD was a major force behind the Orphan Drug Act.) The organization now has a database of over 300,000 individuals affected by rare diseases and is in the process of compiling registries for specific diseases. Looking ahead, Bagal said that NORD envisions a more activist role for the rare disease community and strongly endorses the goals of the ORDR workshop to build the infrastructure for a centralized rare disease registry and biospecimen repositories, actions that can fuel progress in finding the causes and cures for rare diseases. NORD especially encourages collaboration and communication among advocacy groups to create a strong and unified community to bring about change.

Collaboration and unified systems were also major themes in the presentation by Sharon Terry who, as President and CEO of Genetic Alliance and PXE International, can speak to experience working with multiple groups as well as having co-founded a disease-specific advocacy organization. PXE International provides education and support worldwide for patients and families affected by pseudoxanthoma elasticum, a genetic connective tissue disorder. The organization also supports research and has established a 33-lab consortium.

In her talk Terry stressed the need for culture change, both among advocacy groups and the research community-away from competition for funding and the quest for recognition, and towards an open environment exemplified by sharing, dynamic networks, and permeable boundaries between organizations and systems. Proof that it is possible to overcome the silo culture comes from the Genetic Alliance itself, an organization of disparate groups brought together in 1986, formed a cooperative called the Genetic Alliance BioBank in 2003. The Bank's disease organization members collect, store and distribute tissue samples (now numbering over 10,000) according to the specifications of each disease organization's Advisory Board and with the approval of the BioBank's Institutional Review Board. The resulting research has already generated important gene discoveries and diagnostics. Terry proposed the BioBank as a model that could further the goals of the workshop by supplying the training, mentoring, tools, and templates advocacy groups could use to recruit patients into a registries with shared infrastructure and provide options for the collection, processing, archiving, and distribution of biospecimens

While organizations like NORD, the Genetic Alliance, and other groups addressing the workshop celebrated the gains they have made as collectives or cooperatives, impressive results can also be achieved by a single dedicated rare disease advocacy group working independently, albeit one that reaches out to many collaborators. Dr. Leslie Gordon provided just such a telling example in the story of progeria. This rare genetic disease of premature aging affects some 200 children worldwide with perhaps a dozen youngsters in the U.S., most of whom will die from strokes or heart disease around age 13. In 1999, the year after her son was diagnosed, Gordon formed The Progeria Research Foundation with a mission to find the cause, treatment, and cure. At the time there was literally no research (it was not even known whether the disease was genetic), no central source of clinical information, and certainly no treatment. Gordon explained that the Foundation created an infrastructure that enabled it to find patients and

clinicians worldwide, establish an international registry, create a tissue bank, fund research, and hold scientific meetings. In turn, the Foundation developed a second registry of medical information and test data and linked it to its tissue bank. Throughout the decade these activities have resulted in gene and biomarker discoveries and the first-ever progeria treatment trial has been lunched.

In building the Foundation, Gordon emphasized creating programs that inspire trust and serve the mission and find and involve patients, families, clinicians, and researchers as collaborators. It is also important to balance basic and clinical research, she said, and make it clear that studying rare disease is relevant to understanding common conditions.

The link between rare and common diseases, as articulated by Gordon and earlier by Marsha Moses in her research on uMMPs, was the theme of the talk by Dr. Susan Love, founder of the Dr. Susan Love Research Foundation, dedicated to studies of breast cancer. Dr. Love's point was that while it may be scientifically satisfying to study a rare disease gene or risk factor in isolation, the real world is messier and diseases are basically complex. Breast cancer is considered common, but may in fact represent at least five different subtypes. She argued for the creation of an infrastructure that will "allow us to study the commonalities of the rare diseases as well as the rare subgroups of the common ones." One way to do this would be to recruit "unspecified" people who are willing to volunteer for research and then ask them later to self identify as needed for whatever research is proposed. This is the tack she has taken in partnering with the Avon Foundation for Women to form the Love/Avon Army of Women (AOW). The goal is to recruit 1 million women of all ages and ethnicity, with or without breast cancer, who sign up online to take part in breast cancer research. Researchers submit proposals for approval and potential recruits are alerted through email to determine their interest and qualifications. Since 2008, 325,000 women have joined the "Army" and recruitment has now expanded to studies of ovarian cancer. A new Health of Women study has also begun, a collaboration of the Army with NCI and Philadelphia's City of Hope Hospital, which will explore less common diseases.

Online recruitment may also be the answer to finding individuals who belong to no advocacy group, but who may be at risk or in the early stages of a rare or not-so-rare disease. Dr. David Goldstein from the National Institute of Neurological Disorders and Stroke described how such an approach is proceeding in the case of Parkinson disease (PD). Like many neurodegenerative diseases by the time individuals are diagnosed, they have already suffered serious neuronal losses. Using an online recruitment strategy to find potential research subjects who are pre-symptomatic or in the very early stages might enable investigators to elucidate the pathogenetic factors and develop neuroprotective agents.

NINDS is exploring this approach in a web site called Biomarkers of Risk of Parkinson Disease (PD). Interested parties can register, receive an identification number, sign a consent form and check a list of risk factors. As appropriate, subjects can agree to biomarker testing to determine if there is any loss of neurons using neurotransmitters associated with PD and be followed long term to see if they actually develop PD. The beauty of using the Internet, Goldstein commented, is that it is cost-effective and all but guarantees outreach to a vast audience. He added that tabulating the figures on risk factors checked off by registrants is already providing useful research information.

Discussion after the session raised concerns on getting leadership on board in an organization and what to do if a group splinters as a result of discovery of subtypes. Love said that she did not see that as an issue; some subgroups are forming but not breaking off. Also the discovery of other diseases a patient may experience, say a breast cancer patient who develops Parkinson disease, can invite future collaborations. Another attendee commented that patients do not need organizing around a diagnosis, but around their needs. How can they be empowered? What can a family do in a crisis?

Human subjects: bioethical and legal issues for clinical studies

The final Day One plenary session dealt with regulatory issues governing human subject research and concerns for privacy and confidentiality as they relate to research associated with patient registries and biospecimen repositories. Dr. Julie Kaneshiro from the Department of Health and Human Services' Office of Human Protections, and Dr. P. Pearl O'Rourke, Director, Human Research Affairs for Partners HealthCare System in Boston, presented their reviews in tandem. Together they addressed Title 45 of the Code of Federal Regulations Part 46, known as the "common rule," and the Health Insurance Portability and Accountability Act (HIPAA). 45 CFR 46 embodies the federal regulations for the protection of human subjects in research, states the laws, stipulates exemptions, and defines Institutional Review Boards and their formulation. HIPAA was designed to protect the health insurance coverage of workers when they change jobs, but includes a title addressing the security and privacy of health data (the Administration Simplification provisions). These provisions require establishing national standards for electronic health care transactions and national identifiers for providers, health insurance plans, and employers. Kaneshiro and O'Rourke presented examples of where the rules and regulations would come into play in the course of creating and using centralized resources. Thus, if registry data were to be used for a research study by an institution, IRB approval and informed consent or a waiver would be necessary. Suffice it to say that whoever "owns" the resource must be responsible for its rules of operation and address all the regulatory, ethical and business issues that pertain. In turn, whoever is the recipient of resource material must also comply with the relevant rules.

And those rules and regulations do not stop with the federal provisions. The final speaker in the session, Jack Schwartz, Visiting Professor at the University of Maryland School of Law, made it clear that state law can add additional protections, answer questions left unanswered by federal law, and also establish responsibilities not covered by federal law. Among points to consider was whether state laws add anything to HIPAA or whether state property and gift law affects relationships among donors, researchers and research institutions. His take-home message: Don't assume that federal compliance is all that matters!

While Day Two of the workshop was largely devoted to breakout sessions and the reporting out of recommendations, it began with a keynote speech by Dr. Joe Selby from

the Division of Research of Kaiser Permanente (KP) of Northern California. As the largest and one of the oldest health maintenance organization in America (3.2 million enrollees in Northern California; 8.2 million in the U.S. overall) he explained that his presence at a rare disease meeting was because KP uses information technology and conducts research germane to the aims of the workshop. Specifically, KP's "EPIC"-based electronic health record system covers all enrollees and has great potential for data expansion and standardization. In addition, KP has a corps of researchers well versed in building registries and repositories, as well as conducting patient surveys and clinical trials. With regard to rare diseases, he remarked that the company employs clinical geneticists and counselors at specialized clinics in its largest centers where rare disease patients are seen and tracked.

However, the main point of his presentation was not about KP's day-to-day care for rare disease patients, but about how a major HMO can contribute to improving health care by conducting clinical research using KP's huge patient populations. Specifically, KP has in place a Health Maintenance Organization Research Network (HMORN), which connects KP regional centers with other collaborating HMOs representing 12 million patients overall, all covered by EHRs; many using the EPIC system. The network supports a range of federally funded collaborative research programs and uses a standardized database of clinical information that can be fed into a "virtual data warehouse" for access by research teams, thus modeling the kind of centralized registry envisioned by the workshop. Selby also spoke of a particular Northern California KP research program to build the most comprehensive resource for research on genes and environmental influences on health, anticipating enrolling at least 500,00 participants whose survey and genetic information will be linked to their EHRs.

While KP research focuses on common diseases, based on criteria that emphasize the magnitude of the problem, Selby concluded his remarks by noting that the EPIC record system uses a diagnostic coding system more granular than others in common use and so can identify patients with rare diseases. Thus the EPIC codes provide a means of contacting patients for possible recruitment into registries. As well, he provided information on KP consent forms and IRB practices that might also serve as models that advocacy groups could adopt.

Workshop recommendations

A. Standardized vocabulary, terminology, codes and diagnoses

Recommendations

- Standardize questions.
- Find commonalities across all rare diseases.
- Provide guidance to advocacy groups.
- Establish a centralized store of questions.
- Develop a "minimal common registry model."
- Strive for Electronic Health Records standardization.

B. Technology and informatics

Recommendations

- Develop a central wiki/website for resources, best practices.
- Develop an open-source software/hosted registry solution.
- Develop a common repository of questions, answers and data elements.
- Establish a standardized way to share data.
- Develop a scalable method of curation and validation.
- Ensure that technology solutions include multi-lingual capability.

C. Biorepositories/biospecimens

Recommendations

- Establish national rare disease biospecimen repositories using patient registries as sources for donors to build tissue repositories and managing some aspects of informed consent.
- Continue ORDR's existing partnership with the Office of Biorepository and Biospecimen Research (OBBR) at the National Cancer Institute and the National Institute of Diabetes and Digestive and Kidney Diseases to integrate rare disease specimens into the Specimen Resource Locator.
- Have ORDR identify and vet standard operating procedures for specific biospecimen types so as not to duplicate efforts. ORDR should partner with OBBR, the International Society for Biological and Environmental Repositories (ISBER) and other biorepository resources in developing evidence-based protocols.
- Recommend that ORDR sponsor additional workshops and focus groups to investigate issues relating to the communication of results of specimen-based research to donors.

D. Clinical research, patient care and disease management

Recommendations

- Proceed in an incremental and stepwise fashion to develop a centralized registry.
- Use existing applications (for example, IT).
- Evaluate existing models.

E. Patient participation, outreach activities and patient advocacy

Recommendations

- Establish community trust, and ensure follow-up over time.
- Develop inventory of registries outlining their scope, any resources that exist that could be shared, and some quality assessment.
- Find ways to provide incentives to clinicians and reduce costs of clinical data entry into registries.
- Establish a one-stop source listing resources, best practices, and the dissemination of information and guidance.
- Develop a "Registry-building for Dummies" handbook.

D. Bioethical and legal issues

Recommendations

- Decide upon the registry/repository structure that is desired, and then bring in ethical/regulatory expertise to help guide what needs to be done, including from the Office of Human Research Protections.

- Recommend that ORDR develop and make available FAQs and materials to help clarify the relevant provision of 45 CFR 46 (the Common rule), HIPAA, Certificates of Confidentiality, state laws, etc., as related to registries.
- Explore the creation of a centralized IRB and the development of reliance agreements (by which institutions agree to rely on the IRB review conducted at a single institution).
- Develop and formulate model consent forms
  - Consider consent approaches that allow for maximum use of samples and data while making sure that it reflects what individuals want done with their samples and data.
  - Develop standardized language, templates, and approaches to informed consent.
- Communicate aggregated results to participants regularly out of respect for their contribution and to maintain contact and commitment to the research effort.

Acknowledgments

We wish to thank the following organizations for their workshop support:

American College of Medical Geneticist
Emory Genetics Laboratory
Innolyst
Life Fondation.Inc
MDinteractive
Muscular Dystrophy Association
National Center for Research Resources, NIH
National Organization of Rare Diseases
National Eye Institute, NIH
Genetic Alliance
REGISTRAT-MAPI
Sturge-Weber Foundation
Trans Myelities Association
And for Miss Joan Wilentz for preparing workshop transcript