ISSUE BRIEF

Conference on Clinical Cancer Research

October 2010

PANEL 4

Using Patient-Initiated Study Participation in the Development of Evidence for Personalized Cancer Therapy

Stephen Friend, Sage Bionetworks (Co-chair)
Richard L. Schilsky, University of Chicago (Co-chair)
Laurie Fenton Ambrose, Lung Cancer Alliance
Ken Buetow, National Cancer Institute
Jamie Freedman, GlaxoSmithKline
Sue-Jane Wang, US Food and Drug Administration
**Additional contributors appear at the end of the article*

Predicting Response or Non-Response to Approved Oncology Therapies

Approval of new cancer drugs by the U.S. Food and Drug Administration (FDA) relies upon safety and efficacy data from population-based trials. To date, such trials have typically employed tumor classification systems that do not fully account for the growing body of genomic knowledge regarding tumor diversity.[1] When drugs evaluated in these trials are approved and become standard of care, the implications of failing to account for tumor diversity become apparent. When used to combat cancer, standard of care therapies may benefit only one in four patients, leaving perhaps upwards of 75 percent of patients without effective initial therapies and at risk of experiencing only toxic effects.[2]

The goal of personalized cancer therapy can be achieved through the development of new therapies or the selective use of existing therapies in patients more likely to benefit. Designing new targeted therapies requires a clear molecular understanding of the tumor biology and how it varies in the patient population. In cancers for which this understanding is still developing, an alternative approach is to study variations in response to available treatments in search of biomarkers that predict favorable outcomes. In cases where adequate evidence can be developed, a primary goal would be to modify the label of a marketed drug to specify the subgroups most likely to benefit or those unlikely to benefit.

Such a post-approval labeling change happened recently in the case of cetuximab, a member of the class of drugs known as epidermal growth factor receptor (EGFR) inhibitors, which is used to treat certain colorectal, and head and neck cancers. In the years following FDA's February 2004 approval of cetuximab, researchers conducted retrospective analyses of patient samples collected in clinical trials of patients with colorectal cancer and correlated this with the existing clinical trial data, demonstrating a strong connection between *KRAS* mutation and clinical outcomes in this disease setting.[3] In July 2009, FDA responded with labeling changes to

[1] Friend SH, Golub T, Radich J, Sawyers C, Schilsky R. Identification of Non-Responders to Approved Cancer Drugs through Patient-Oriented Sample and Data Collection: Strategic Summary and AML Pilot. Overview for NCI Meeting, September 23, 2010.
[2] Friend SH, Schilsky RL. Sage Project: Finding Non-Responders to Approved Drugs through patient oriented sample/data collection and hosted in the public domain. March 2010.
[3] NCI. Colorectal Cancer Drugs Require Careful Patient Selection. 2008 [last update]. 17 September 2010. <http://www.cancer.gov/clinicaltrials/results/summary/2008/kras0608>

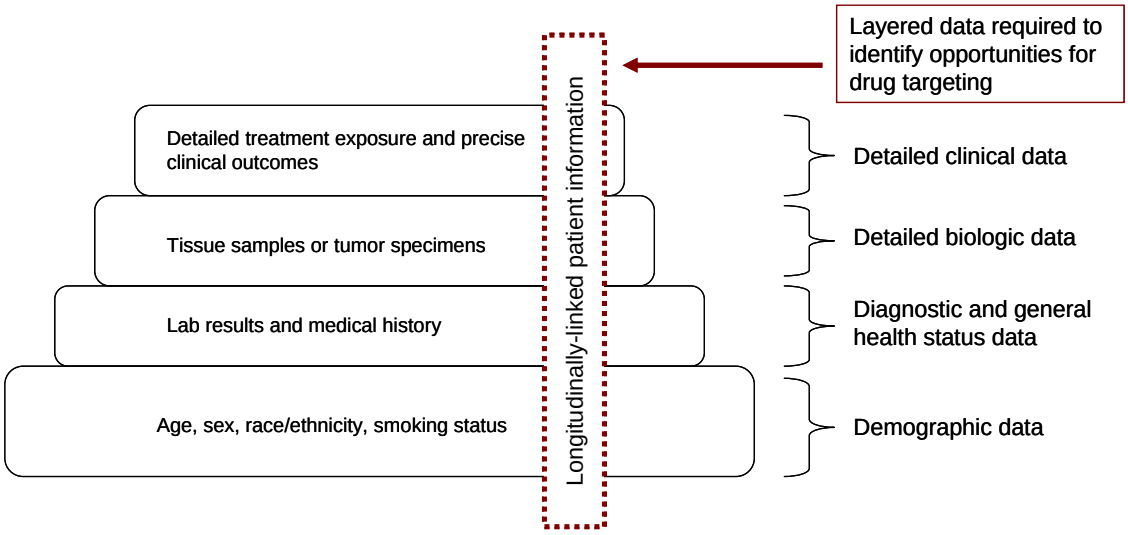
cetuximab that narrowed the indication to use in advanced colorectal cancer patients with EGFR-expressing tumors and without *KRAS* gene mutations.[4] While this process illustrates one potential route to successful labeling changes for newer drugs, for older products, existing data may not contain the needed genomic information to identify markers of response or non-response. New efforts to collect comprehensive data regarding use of these treatments may be required to identify subsets of likely responders and non-responders.

One potentially promising avenue for developing such evidence rapidly lies in engaging patients directly to contribute detailed information about their tumors, treatments, and clinical outcomes. This discussion document explores the types of evidence that might be required to inform labeling changes and clinical decision-making, as well as the feasibility of developing that evidence through patient-initiated study participation. For the purposes of this issue brief, we define patient-initiated study participation as a model in which patients are engaged and recruited directly by the sponsor of an IRB-approved study, and patients in turn drive the participation of their physicians and other health care providers to facilitate collection of required data and/or tissue samples.

Data Required to Identify Patient Subsets

Developing evidence to support targeting available treatments to a subgroup of patients requires collecting detailed and high-quality data. This data can be thought of in layers of comprehensive, longitudinally-linked information so that treatments can be tracked over time and within subgroups of patients. In addition to basic information like demographics, lab results, and medical history, needed layers will likely include normal tissue samples, tumor and other biological specimens, detailed information on treatment exposure, adverse events and clinical outcomes (see Figure 1 below). Few data sources presently have the breadth and depth of information necessary to support analyses with sufficient statistical power.

Figure 1. Layered Data to Enable Drug Targeting



[4] Cetuximab (Erbitux) and Panitumumab (Vectibix). 2010 [last update] 23 Sept 2010. <http://www.fda.gov/AboutFDA/CentersOffices/CDER/ucm172905.htm>

Patient Engagement in Generating Samples and Data

Cancer clinical trials are vital to expanding our knowledge regarding new and existing therapies. However, in the traditional physician investigator-based patient recruitment model, these clinical trials are plagued by low enrollment and even lower compliance for sample collection. Difficulties with low enrollment persist despite high patient interest in being part of the research process. A 2005 survey of cancer patients found that while 65% would be receptive to enrolling in a cancer clinical trial, only three to five percent actually participate.[5] A potentially more efficient model of conducting cancer research is to start by engaging the patient rather than the physician investigator. Obtaining information and/or samples by directly engaging patients to participate in a study is a promising approach to creating comprehensive layered data. Indeed, many advocacy groups have begun to mobilize their networks to participate in a variety of ways, and several have begun creating their own datasets as a result of patient-initiated participation. As described below, these efforts are positive signs that patient-initiated study participation might be feasible.

Love/Avon Army of Women and Health of Women Initiatives

The Dr. Susan Love Research Foundation sponsors two initiatives that underscore new ways to engage patients. The Love/Avon Army of Women Initiative, co-sponsored in 2008 by the Avon Foundation for Women, is attempting to recruit one million healthy women (including breast cancer survivors and women at risk for breast cancer) to participate in breast-cancer related studies. The Army of Women Initiative maintains a database with basic demographic information (e.g., name, e-mail, year of birth, zip code) regarding volunteers. To date, over 337,000 women have registered online and 34 studies have been launched after successful matches were made between interested women and researchers. To initiate the process, researchers fill out a standardized online application that an outside board reviews in coordination with an institutional review board (IRB). If approved, Army of Women sends an 'e-blast' to everyone in the database, and women choose in which studies they would like to participate.[6]

The second initiative, the Health of Women Study, launched December 2009 in collaboration with the cancer Biomedical Informatics Grid (caBIG) and City of Hope. At its launch, the Health of Women Study invited all volunteers registered with the Army of Women study to participate. The Health of Women Study is the first large longitudinal cohort study for breast cancer done entirely online – a significant break from traditional studies that are often lengthy, paper-based, and not tailored to individual participants. Every month, volunteers will receive a questionnaire and Army of Women will collect and analyze the data after it has been de-identified by City of Hope. By moving the process online, the questionnaire can be adapted to ask the right questions for a given participant (e.g., omitting pregnancy-related questions for post-menopausal women), which can ease the often-arduous data cleaning process by asking the correct questions at the beginning. In addition, monthly online contact helps to engage patients in the research process in a consistent manner. The initiative is planning a pilot to collect biological samples (e.g., tissues, blood) over the next 12-18 months, and to enable patients to upload medical records within the next two years.[7]

---

[5] R. L. Comis, D. D. Colaizzi and J. D. Miller. Barriers to cancer clinical trials (CCT) participation: "We have met the enemy and he is us." Journal of Clinical Oncology, 2007 ASCO Annual Meeting Proceedings (Post-Meeting Edition). Vol 25, No 18S (June 20 Supplement), 2007: 6567

[6] Love / Avon Army of Women. Web. 23 Sept. 2010. <http://www.armyofwomen.org/>

[7] Health of Women Study. Web. 23 Sept. 2010. <http://www.armyofwomen.org/HOW_Study>

Lung Cancer Alliance's Give a Scan Program

The Give a Scan program is intended to accelerate imaging research and software development so that imaging can become an accurate quantitative "biomarker" and thus expedite improvements in the early diagnosis and treatment of lung cancer. Leveraging the motivation and commitment of patients and their families, the Lung Cancer Alliance (LCA) sponsors a website (www.givascan.org) which allows individuals to submit information to a public database of patient-reported clinical data and images. The LCA collects, de-identifies, and provides public access to the data.

For initial data collection, patients sign informed consent and authorization for release of medical information forms, and request their health care providers to provide a copy of their electronic image data (e.g., computed tomography scans). After this information is submitted to the Give a Scan website, patients are asked for additional information about their condition, stage at diagnosis, smoking history (if any), and familial history of lung cancer.[8]

The Clearity Foundation

The Clearity Foundation is a not-for-profit foundation that was founded in 2007 by an ovarian cancer survivor and scientist who was unable to have her tumor profiled when diagnosed. The foundation sponsors molecular profiling for ovarian cancer patients, maintains a database of results, and provides doctors and patients with clinical trial options informed by individual tumor biology. After a patient fills out informed consent and authorization for release of medical information forms, The Clearity Foundation works with physicians to order necessary genetic tests and obtain their results. The Clearity Foundation maintains an ovarian cancer-specific database of de-identified patient reports, which allows them to superimpose individual patient reports onto larger populations for comparative, longitudinal analyses.[9]

Multiple Myeloma Research Consortium

Established in 2004, the Multiple Myeloma Research Consortium (MMRC) comprises thirteen research institutions which contribute patient-donated tissue samples (with corresponding genomic and clinical data) to a central MMRC Tissue Bank.[10] With an overall goal of accelerating "the development of novel and combination treatments for patients with multiple myeloma," the MMRC to date has facilitated 19 Phase I and II clinical trials.[11] In order to donate samples, patients at one of the thirteen participating institutions can elect to give additional bone marrow aspirate and blood during their regularly scheduled procedures. Patients who are not associated with one of the thirteen institutions can donate by scheduling their next bone marrow aspirate at one of the participating institutions.

Although these initiatives have been developed to achieve a variety of research objectives, they speak to the motivation and commitment of patients and their families to advancing cancer research in general, and personalized cancer care in particular. In order to ensure that their efforts result in actionable information, what is needed now is a clearer understanding of how such data can be most effectively collected and used to inform the decisions of doctors, patients, FDA, and payers.

---

[8] Lung Cancer Alliance. What is Give a Scan? Web. 23 Sept. 2010. <http://www.lungcanceralliance.org/giveascan/>
[9] The Clearity Foundation. Web. 23 Sept. 2010. <http://www.clearityfoundation.org/index.html>
[10] The MMRC Model. Web. 23 Sept. 2010. <http://www.themmrc.org/model_mmrc.php>
[11] About the MMRC. Web. 23 Sept. 2010. <http://www.themmrc.org/about_mmrc.php>

Designing a Study to Identify Markers of Response or Non-Response Through Patient-Initiated Study Participation

Changing FDA-approved labels and recommended standards of care requires robust evidence built on high quality data and an acceptable study design. We envision collection of these data through post-approval studies, which could be prospective studies of marketed drugs, in which genomic data from biospecimens are used to identify biomarkers predictive of clinical outcomes. The objective of such studies is to identify biomarkers that reliably predict patients either likely or unlikely to benefit, and to use such data to support revision of the approved drug label and a change in standard of care. The following general principles should guide the development of such evidence.

1. Study Objective(s) Should be Clearly Laid Out

The study objective(s) should address a clinically meaningful improvement in patient outcomes.

2. *Study Design Should Have the Ability to Address the Primary Study Objective*

Examples of patient-driven post-marketing studies for approved drugs could include single arm cohort studies with clinical outcome measures, randomized controlled trials with placebo or active comparators, or other designs. Patients will provide genetic data from normal and/or diseased tissues (including tumor and/or blood, etc.) to test biomarker hypotheses of clinical response or non-response. Ideally, the response biomarkers will be evaluated in a prospective randomized controlled trial such that the results can be used to modify the drug label. Alternatively, such hypotheses may not exist *a priori*, so trials can also be designed where retrospective correlation between clinical outcome and genetic markers can be used to generate new hypotheses for subsequent prospective testing. In this scenario, the FDA has recommended high sample collection rates (>95%) to provide sufficient power for results that support the scientific hypotheses being tested.

3. Biomarker Data Used to Modify the Label for Marketed Drugs

The data will need to clearly demonstrate that a biomarker(s) predicts response or lack of response to a given therapy with sufficient statistical power. The number of studies required to achieve label revision, if a patient subpopulation, e.g., responder or non-responder, exists, will depend on the strength of the clinical effect and the patient population studied.

There are several factors that can impact the ability to generate useful data from a genomic biomarker trial including: 1) the prevalence of the target biomarker in the population; 2) the prognostic impact of the biomarker to distinguish clinical outcomes in the population; 3) concordance between primary and metastatic tumor tissue; 4) qualification and validation (analytical and clinical) of the biomarker assay; 5) availability of tissue specimens containing the biomarker; and 6) quality and quantity of the tissue samples for biomarker analysis. Each of these factors will need to be addressed in any study of this nature.

Hypothetical Example

To help illustrate the issues, challenges and potential solutions in using data generated through patient-initiated study participation for the purpose of informing labeling changes for existing cancer therapies, we examine study design considerations within the context of treatment for non-small cell lung cancer (NSCLC). First-line treatment for NSCLC typically consists of chemotherapy with a two-drug regimen containing either cisplatin or carboplatin and another agent, which is typically vinorelbine, paclitaxel, docetaxel, gemcitabine, or pemetrexed. However, experience with these regimens indicates that only approximately 30% of patients respond favorably.[12]

In order to identify molecular signatures that explain variation in treatment response, several initiatives, including the Sage Bionetworks Non-Responder Project, are working to design studies that identify predictive markers of non-response. The Non-Responder Project has chosen several candidate tumors to study, including NSCLC, with an initial pilot study in AML[13] based on four "first principles" for tumor selection:[14]

1. The treatment under investigation should have substantial response and non-response rates (>20 percent in either group).
2. The disease must have clear, robust definitions of response and non-response that are clinically important. (A non-response biomarker should have the potential to change clinical practice.)
3. Routine clinical management of the disease guarantees access to high quality tissue specimens. (Use of archival tissue from diagnostic samples introduces risk when assessing treatments given at relapse.)
4. The non-response group should ideally be defined as patients refractory to treatment rather than those who respond then relapse early. (If early relapse is caused by a resistant subpopulation at diagnosis, genomic analysis of tissue at diagnosis may or may not be informative, depending on the size of the resistant pool.)

Informed by the Non-Responder Project and other projects such as the Texas Clinical Trials Network (CTNet) organized by MD Anderson, we illustrate the potential use of patient-initiated study participation to identify predictors of non-response within the setting of NSCLC. The objective of the proposed study is to identify one or more molecular markers of non-response to first-line platinum-containing therapies for metastatic NSCLC, with the goal of supporting the revision of FDA-approved labels and recommended standard of care for these drugs.

Study Design

We propose to develop a single-arm prospective registry, in which enrollment and sample/data collection can be initiated by patients and/or physicians. Among patients recently diagnosed with NSCLC and to be treated with front-line chemotherapy containing a platinum agent, the registry will prospectively collect normal and tumor tissue samples prior to initiation of treatment, patient demographics, treatment details, and clinical outcomes. These data will then be analyzed to identify molecular signatures that predict treatment outcomes.

[12] G.V. Scagliotti, F. De Marinis, M. Rinaldi, L. Crinò, C. Gridelli, S. Ricci, E. Matano, C. Boni, M. Marangolo, G. Failla, G. Altavilla, V. Adamo, A. Ceribelli, M. Clerici, F. Di Costanzo, L. Frontini, M. Tonato. Phase III Randomized Trial Comparing Three Platinum-Based Doublets in Advanced Non–Small-Cell Lung Cancer. Journal of Clinical Oncology, Vol 20, Issue 21 (November), 2002: 4285-4291

[13] Friend SH, Golub T, Radich J, Sawyers C, Schilsky R. Identification of Non-Responders to Approved Cancer Drugs through Patient-Oriented Sample and Data Collection: Strategic Summary and AML Pilot. Overview for NCI Meeting, September 23, 2010.

[14] Ibid.

Definition of Treatment Response/Non-Response

The standard definition of objective response in NSCLC uses RECIST criteria and tumor measurements. A unidimensional reduction in size of target lesions of at least 30% is considered a response to therapy, while progression of disease is characterized by an increase of 20% or more.[15] Tumor changes between these thresholds are considered stable disease. About 30% of patients exhibit an objective response to first-line chemotherapy regimens containing platinum, and another 20% will achieve stable disease. Thus, approximately 50% of patients will be "non-responders" to these regimens.

Patient Population and Sample Size

Based on the assumption that 50% of patients are expected to be non-responders, and that 20% of biopsies will not provide enough tissue for analysis, we suggest enrolling at least 360 metastatic NSCLC patients (to obtain 150 responders and 150 non-responders with evaluable specimens). Enrollment and sample collection will occur after diagnosis but before treatment is initiated. Patients are to be treated according to standard of care with first-line platinum-containing chemotherapy, based on their physicians' recommendations. To the extent that there are multiple doublet combinations (e.g., cisplatin-docetaxel, carboplatin-paclitaxel), the study would need to be powered sufficiently to detect biological signatures unique to the smallest group assuming that drug-specific signatures exist.

Collection, Storage and Analysis of Tissue Samples

Upon consenting to participate, the patient must provide samples of tumor and normal tissue. The required tumor biospecimens may be collected as part of routine clinical care (e.g., for testing EGFR status) or may be protocol-specified to test for other biomarkers. In some cases, tumor biopsies will be performed for research purposes only following informed consent.

For the current exploratory study, fresh frozen tumor samples are needed. As technologies progress there will soon be methods to generate robust RNA and DNA data from archived samples embedded in paraffin blocks. After initial exploratory studies have identified a key set of alterations at the DNA and RNA level that predict response or non-response, diagnostic assays used for general patient screening purposes will be performed on routine paraffin sections. Since this study does not pertain to a hematologic malignancy, normal tissue will be collected in the form of peripheral blood mononuclear cells. In similar studies that do pertain to hematologic malignancies, skin biopsies or hair follicles can be a source of normal DNA.

Tumor samples will consist of two core needle biopsies, which are required to obtain the necessary amount of RNA (2-3 micrograms) for planned analyses. The site of the biopsy will be recorded after determination by the patient's care team of the optimal site based on likely yield and safety of the required procedure. We recommend that institutions participating in this initiative engage a team of specialists (e.g., oncologists, pathologists and surgeons/interventional radiologists) to enable accurate and adequate sample collections. After each procedure, samples should be fresh frozen. Specific SOPs for sample acquisition, handling, transport and storage will be provided in the study protocol. All tumor specimens will be analyzed in a CLIA-certified laboratory for known and clinically-actionable genetic variants

[15] National Cancer Institute. Imaging Response Criteria. Web. 4 October 2010. <http://imaging.cancer.gov/clinicaltrials/imaging/>

such as EGFR mutations or ALK translocations, and the results will be returned to patients and their physicians within 7 days.

Given the limited quantity of RNA available from these samples, research analyses will be prioritized to include whole genome sequencing, mRNA gene expression arrays, and micro RNA profiling. These assays will be performed in research laboratories under the direction of qualified collaborating investigators to be determined by the study directors. If clinically actionable results are identified from the research analyses, they will be returned to the patient and treating oncologist, who have the option of requesting more complete information on their test results—whether clinically actionable or not.

Collection of Other Data

Following the collection of biospecimens at a designated research center, the patient returns to the care of their oncologist (with additional care and services provided by pathologists, surgeons/interventional radiologists, and others as medically necessary). The patient and oncologist then collaborate to collect and report additional data needed for the study. Required elements include complete information on treatment of each patient, including patient demographics as well as the following types of clinical information and patient outcomes:

- Tumor stage and histology
- Chemotherapy start and stop dates, drugs and dosages administered
- Tumor measurements pre-, during, and post-treatment (including dates of measurements) which form the basis for determining response
- Grade 3-5 treatment-related toxicities by treatment cycle
- Date of tumor progression
- Date of death

Of these outcomes, the primary clinical endpoint for this study is response vs. non-response based on tumor measurements. In order for these data to be viewed as valid, information on this endpoint must be provided by more than patient self-report of response or non-response. In this patient population, it is expected that the primary clinical endpoint will clearly segregate non-responders from responders within 6 months of study entry. Dates of tumor progression and patient death will also be collected.

Statistical Analysis

All data analyses will be pre-specified in the IRB-approved study protocol. Included in this study design would be the classical statistical analysis along with the network-biology modeling done to identify not just isolated markers for non-responder populations, but also sets of genes or “gene signatures” capable of identifying non-responders. The goal for lung cancer might be to identify patients who have greater than an 85% chance of not responding with a certainty of this outcome of 90%. This certainty around the likelihood that a patient may not respond would need to be set at predetermined level of stringency in order to enable clinicians to use this information to determine whether to forego the original approved therapy and instead provide the patient an opportunity to receive an investigational regimen. These standards for foregoing standards of care would be tumor and regimen specific and would need to be agreed to up front with regulators and physicians before the study was started.

The Feasibility of Patient-Initiated Study Participation

Enrollment takes place when a patient nearing treatment decisions becomes aware of the opportunity to participate in the study via a website description or other form of outreach. While patient-initiated study participation offers promising opportunities for more efficient and dynamic clinical trial enrollment, a number of feasibility issues must be considered during data collection so that resulting data are relevant for regulatory decision-making. The following section outlines a suggested set of principles for patient outreach and data collection that can best anticipate current and future needs of researchers and regulators.

Patient Engagement

Patient-initiated study participation begins with raising awareness of donation opportunities, achieving patient/family engagement, and supporting patients/families through the process of enrolling in the study. Patient advocacy groups like the Research Advocacy Network produce and publish a range of accessible educational materials that can help to provide patients with an understanding of how tissue samples fit within the context of clinical research, an explanation of the science behind genomics, and overall background on the clinical research process.[16] In order to ensure optimal patient participation, we recommend that patient-initiated study participation efforts do not impose any sort of fees on patients. Instead, the organizer of such efforts should absorb any associated costs or such costs should be incurred as part of routine cancer care.

Role of Health Care Providers

Health care providers have an important role in data/sample acquisition. Health care providers can help to streamline the process of donation, making it as easy as possible for patients to participate. They can encourage patients about the importance of providing tissue samples, even when the collection of such tissue samples poses health risks to the patients. The importance of the biomarker data generated from these sorts of trials needs to be clearly communicated. In order to enlist the support of providers, those leading patient-initiated study participation efforts should consider identifying supportive providers and providing these providers with detailed information about the initiative and what's expected in terms of provider involvement.

Sample Collection and Storage

When necessary, collection of biologic samples must address specific challenges. In general, normal tissue (e.g., blood, skin, hair follicles) is easier to collect than tumor specimens. However, even these samples may require more complex sample collection schemes (e.g., peripheral blood mononuclear cells from whole blood) that require specialized collection methods and expertise at the clinical sites. Tumor samples are generally a lot more difficult to collect because they require invasive procedures and the quality of the specimens may not be adequate. Certain anatomical sites are more amenable than others for collection of tumor specimens. For example, the skin, lymph nodes, bone marrow, and liver are usually straightforward collection sites. Primary lung cancer specimens are very hard to collect because of location. If there is an assay for archival tissue from the original surgically obtained tumor specimen, then that may allow the highest yield.

[16] Research Advocacy Network. Web. 6 October 2010. <http://www.researchadvocacy.org/publications/posters.php#TISSUE>

After samples and other data are obtained from patients, processes must be developed to efficiently compile and integrate large amounts of patient-donated data/samples. Efforts that rely on patients to directly transfer data (e.g., CT scans) to the organizer will be more direct and simple to accomplish. Obtaining biospecimens from patients may be more challenging because patients are typically not the "owners" of these samples and coordination must occur with health care providers. Such coordination may be more possible if sample collection occurs at designated collection facilities and if relationships have been previously established with a core set of providers. As patient data/samples are collected, they should ideally be stored in a way that preserves the ability to link to other sources of electronic clinical data information (e.g., from electronic health records). This is critical to creating the type of layered data necessary to identify markers of response and non-response.

Data Access

To fully realize the goal of patient-initiated study participation, we recommend that data are compiled and made available in a standardized electronic format to all qualified researchers, rather than restricting access to a particular investigator or team. All qualified researchers should have access to compiled data without incurring any fee. In other words, initiatives to collect data/samples through patient-initiated study participation should not be in the business of selling these data to interested researchers. We believe that these principles will help enable the widest possible access to patient data, and therefore the greatest possibility for important discoveries.

Patient Privacy and Data Security

Ethical use of the data and samples requires review to ensure protection of human subjects, as well as assurance of patient privacy and data security. Up front, efforts to collect data/samples through patient-initiated study participation should be aware that IRB approval for compilation of data and working out the legal issues, preparing patient consent forms and the EULA for the site requires considerable time and money. Patients should be clearly advised that their donated data would be openly accessible to researchers and that the product of the research may be commercialized. The level of identification risk associated with donating their data must be transparently communicated to the participating patients and informed consent obtained.

Patient privacy should be protected by removal of all HIPAA "identifiers" and by agreements not to seek re-identifying information except for research covered by the informed consent. Double de-identification may provide further privacy protection, by employing two levels of coding between HIPAA "identifiers" and information relevant for research purposes (e.g., health outcomes, genetic/genomic test results). Using this approach increases the stringency of privacy protection, while retaining the potential for future analyses building upon the collected data. Other approaches, such as total anonymization, would likely not be favorable because these approaches involve permanently breaking the links that allow identification of research participants for collection of further data.

Patient-initiated study participation efforts should employ controlled access to patient-level data through which researchers seeking such data would have to make appropriate commitments including: 1) use only for approved research; 2) no sharing of data/samples with others without such sharing having been referenced in the consent form; 3) no effort to re-identify; 4) and the repository would be obligated to confirm that the proposed research is consistent with the scope of the consent forms.

Governance

Governance policies are required to establish oversight of data collection and use. This is essential to maintain aspects of compliance, privacy, and access to data and models within the project. Existing projects involving clinical/genomic data set generation by structures such as caBIG, The Cancer Genome Atlas (TCGA), and trials such as the BATTLE trials and the I-SPY trial network provide precedents for executing on these governance rules and processes. One potentially important issue related to governance involves the types of entities that might bring data forward to a regulatory agency as a result of patient-initiated study participation. Given that such efforts may be spearheaded by nonprofit as well as commercial organizations, it is possible that a nonprofit organization, not affiliated with a commercial product sponsor, might develop and submit data on molecular markers associated with response/non-response to an approved drug for review by a regulatory agency. If these data led to a labeling change, such a change might occur independently of the product sponsor and possibly without their agreement. Arguably, these changes would likely be in the interest of the patient community and society in general, but they might not always be in the interest of product sponsors.

Principles for Effective Management of Patient-Initiated Data Collection

After patient data has been collected in accordance with the feasibility principles outlined above, it needs to be standardized and stored in such a way that researchers can use and share it for a variety of diverse purposes – all while maintaining the confidentiality of individual patients. In order to arrive at that point, there are a number of data management issues that need to be addressed in order to ensure data can be effectively pooled and linked across sources. The following section discusses these data management issues.

As with any clinical research, it is essential that data be of a standard form when used in analysis. In a traditional research setting, standardization of the collection of data from multiple sources is accomplished through the use of common data collection forms and adherence to common practices in completing the forms. Years of practice in the oncology community has resulted in the generation of a large library of these standard forms, composed of common data elements which use terminologies and ontologies that are national and international standards.

In a partnership among academia, industry, and the FDA, these elements and ontologies have been used to create a common information model that serves to support electronic regulatory submission. The NCI maintains tools and infrastructure that share these resources in an unrestricted manner with the entire biomedical community. Wherever possible, data collection should leverage these and other standard information representations.

Data generated through patient-initiated study participation is unlikely to arise solely from the clinical research arena. Instead, the data will arise from healthcare encounters that utilize different information representation standards. Further complicating this data capture is that such encounters occur in multiple disparate locations, with each location using variant dialects. Lastly, most data captured in a clinical setting is not captured in structured form (elements with controlled values), but instead in narrative form. This diversity and lack of structured information raises both challenges and opportunities.

First, where only narrative data exists it is possible to extract structured information using the same tools that support clinical research. It is important that these tools be independent of any

specific vendor platform and, ideally, accessible via the web. As indicated above, in partnership with the Army of Women, the NCI's caBIG program has created web-based consumer tools that utilize the information representations commonly used in clinical research.

There has also been significant progress made in bridging clinical research information with clinical care information representation. First, the clinical care community has worked with standards-developing organizations, such as HL7, to create a common representation of care information to which alternative dialects can be mapped. Second, a common information model has been generated that permits the clinical research and care representations to be joined. It is therefore possible to convert alternative dialects to a common information representation that meets both care and research needs. These mappings are being used to support a new, adaptive-design clinical trial, the I-SPY2 trial. In this trial, patient information from standard care encounters is transferred to the clinical trial setting using the caBIG Integration Hub.

In oncology, the American Society of Clinical Oncology and the cancer community have worked with the NCI to create the "Clinical Oncology Requirements for Electronic Health Record" (CORE) specification that leverages this common information representation. A subset of the CORE specification can be used to create an ultra-light electronic health record for oncology that captures critical information on demographics, diagnosis, intervention, and outcomes. This specification represents information necessary to support recent HITECH statutory obligations to provide information to patients in electronic form within 24 hours of request. The NCI, in partnership with the community, has generated tools that facilitate this information being provided to patient-controlled, or personal health records. This capability permits the patients to directly choose to participate in clinical research and to share their information with the study. The portals hosting such patient-controlled records are normally maintained by advocacy organizations.

In a project such as the non-responder project, clinical information represents only a single dimension among the multiple diverse types of data that must be managed and interconnected. Similar considerations exist for biospecimens, imaging data, and the molecular data that will be used to characterize the individual participants. This information must have common representation across the diverse organizations acquiring and sharing the data. The caBIG community has created such representations and a collection of tools, accessible as web-tools, that utilize them. However, similar to clinical information, the caBIG Integration Hub permits disparate types of information to be cross-mapped to a common representation. Researcher-generated data can then be collected in a standardized manner and captured in infrastructure that can support reuse by other investigators as authorized by patients.

Aggregation and analysis of the complex, multidimensional data also requires novel infrastructure. The caBIG community has created data mart/data warehouse tools that facilitate the collection and effective use of the multidimensional clinical and molecular data through its caIntegrator capabilities. These tools effectively manage the large volume and complexity of data for projects such as The Cancer Genome Atlas.

Controls that protect patient privacy and assure that only authorized individuals have access to data are essential to projects such as the non-responder project. To this end, it is necessary to establish a "trust fabric" that grants access only where appropriate and only to data components that have been authorized (HIPAA) or consented (OHRP) by the patient. As HIPAA assigns responsibility for protections to local groups that hold patient information, this trust fabric needs to recognize the need for local control of data release. However, once patients have access to their data and control its redistribution, the regulatory framework changes. Much of this policy

framework is still developing at the national level. Honest brokers, such as advocacy groups who are acting as gatekeepers for data, may have a less complicated regulatory framework in which to operate.

Patient control of data has raised concern over issues of potential discrimination. More specifically, with patient control of information one issue is whether disclosure is required to insurers, employers, or others of information previously not accessible to patients. This is especially true of research data. Recent legislation -- such as the Genetic Information Non-discrimination Act and the Patient Protection and Affordable Care Act -- reduces some concerns, but as above, much of the policy framework is still under development. Given this, patient-initiated study participation efforts should employ governance policies that ensure collected data and samples are not disclosed to insurers, employers, or other groups, in order to minimize the risk of discrimination.

Conclusion

Patient-initiated study participation is a potentially promising way of rapidly generating evidence to support better targeting of previously approved cancer therapies. Cancer patients, caregivers, and their advocates have demonstrated strong enthusiasm for improving the efficiency of clinical research. We have proposed a model that leverages the motivation and commitment of cancer patients to overcome some of the challenges in the collection of data and biospecimens that can be used to identify biomarkers predictive of non-response to previously approved chemotherapeutic agents. The model is based on several important principles for study design, patient engagement, and data/biospecimen collection and management. Care should be taken to ensure that such studies are designed with broad-based input from all stakeholders, so that patients are informed appropriately, the correct types of data and biospecimens are collected, information is compiled and managed efficiently, the resulting database is made available to researchers with appropriate protections and security features in place, and that the data are analyzed in a way that yields evidence of sufficient quality to inform regulatory decisions and clinical practice. It is especially important that patient advocates, regulators, clinical oncologists, and research methodologists be participants in the design of such studies.

Acknowledgement of Contributors:

- Joshua S. Benner, Engelberg Center for Health Care Reform, The Brookings Institution
- Roy Herbst, The University of Texas MD Anderson Cancer Center
- Erin Karnes, Engelberg Center for Health Care Reform, The Brookings Institution
- Elaine Khoong, Medical Student at Washington University School of Medicine
- Annie Martin, GlaxoSmithKline
- Ed Walters, Engelberg Center for Health Care Reform, The Brookings Institution
- Ignacio I. Wistuba, The University of Texas MD Anderson Cancer Center