PANEL 1
Data Submissions Standards and Evidence Requirements

Richard Schilsky, MD, University of Chicago Medical Center
Jeffrey Abrams, MD, National Cancer Institute
Janet Woodcock, MD, Food and Drug Administration
Gwen Fyfe, MD, Genentech
Robert Irwin, Marti Nelson Cancer Foundation

The need for data collection and reporting standards

The design of any clinical trial involves a variety of tradeoffs. One set of issues involves balancing the opportunity to collect large amounts of data on each subject with other important priorities, such as the ability to verify the quality of the data, the number of subjects that can be studied in the trial, the length and intensity of follow-up, and the cost of the trial. In some cases, the most important factor is randomizing a sufficient number of subjects with only a few data points per subject, the large simple trial – for example, an outcome study of an approved drug. In many other cases, such as trials intended to support initial regulatory approval, extensive amounts of data are often collected on each subject. More data are sometimes collected than is necessary to ensure that cancer treatments are safe and effective, increasing the cost and duration of clinical trials. A risk also exists that the magnitude of data collection may compromise overall data quality by creating an enormous burden on investigators and clinical study sites.

The U.S. Food and Drug Administration (FDA) has issued Guidance on data collection in registration trials for cancer therapies and general Guidance on how to conduct safety reviews.[1,2,3] However, these Guidances are not prescriptive and sponsors often err on the side of caution by collecting a great deal of data on each subject. Variability exists across development programs regarding the nature and extent of data collection. Some variability is expected, given the diversity of drugs, diseases, phases of clinical investigation, and ultimate intended use of the product. However, better agreement and understanding about the essential data elements that should be collected for various types of oncology trials intended for regulatory submission would be beneficial. It is appropriate to ask what should be collected, why, and whether greater efficiency can be achieved – not only in developing a standardized approach, but also in minimizing certain aspects of data collection without jeopardizing development of the database essential to evaluate the safety and efficacy of the agent. Common data collection and reporting standards can improve the efficiency of clinical cancer research.

In cancer, different kinds of primary approvals have led to confusion and inconsistency in the approach to safety data collection for a supplemental approval. What is required in supplemental approvals depends very much on how “well-characterized” the safety profile of the agent is, and on the intended use of the product. This confusion and subsequent lack of consistency in regulatory approach sometimes leads to detailed and prolonged negotiations between the FDA and the sponsor on a protocol-specific basis, as well as inconsistent approaches to data collection.

These issues further burden a clinical trials system that is already struggling. In contrast to other countries where 20 percent or even greater percentages of adult cancer patients may enroll in trials, less than five percent of U.S. adult patients participate. In part this deficiency is caused by the high costs and low reimbursement offered to sites participating in clinical trials, as well as the regulatory risks associated with participation in complicated trials that have detailed data collection requirements. Given the broad commercial availability of oncology agents following initial regulatory approval, clinical trial participation in the U.S. is increasingly undertaken only by trial sites that are

deeply committed to the assessment of investigational agents in the medical milieu of America's health care system.

The Brookings Conference on Clinical Cancer Research provides a unique opportunity to discuss data collection elements for therapeutic cancer trials. This paper will serve as background for Panel 1 by laying out the benefits of data collection standards, proposing a framework for data collection, and noting areas where further dialogue on standards is necessary.

Benefits of data standards

A variety of data elements and formats for collection exist. As studies of adverse event (AE) data collection have demonstrated, the method of data collection can profoundly influence how a trial's results are interpreted. One study found that rates of reported AEs are dependent upon how information is gathered; patients who received a checklist of adverse events to complete reported significantly more adverse events than those who were asked open-ended questions.[4] The study authors observed that comparing treatments for the same indication will be uninformative if studies of those treatments use different methods for collecting AE data.[5] Other variables that may impact reporting of adverse events include the frequency of follow-up visits and how trial paperwork is completed.[6]

Uniform data collection standards could improve the interpretation of trial results by delineating consistent standards for content and methodology. Consistency of approach is likely to improve site performance. Conversely, the variety of approaches currently in use may hinder good safety data collection. A recent review of 22 trials sponsored by the National Cancer Institute (NCI) discovered that AEs for some trials were reported differently in published articles versus the NCI's Clinical Data Update System. The authors noted that published reports inadequately described low-grade, high-grade, or recurring AEs.[7] Another study of safety data reporting in randomized drug trials found that none of the 192 trials reviewed exhibited "safety reporting [that] can be deemed satisfactory."[8]

Faster and less costly trials

Pharmaceutical company sponsors and academic investigators/sponsors often gather more data than are necessary. The extent to which the adverse event data collected during trials – especially trials that are intended to support a subsequent cancer indication following full approval and a well-characterized safety profile – are useful for gauging treatment risks is not currently known. Recent post hoc simulations based on data gathered during the Avastin Non-Small Cell Lung Cancer (AVAIL) trial demonstrated that toxicity data collection could be streamlined without significantly increasing the risk of missing an important adverse event. If toxicity data on Grade 1 and 2 events had been collected in a subset of 200 patients per arm rather than in every trial participant (n=~650), approximately 2,500 fewer adverse events would have been collected; extrapolations to estimate monitoring and transcribing efforts suggest that this could result in a time savings of at least 2,500 hours.[9] When researchers simulated a scenario in which Grade 3 and 4 events were collected in a subset of patients (instead of all patients) during a large trial, adverse events occurring at least five percent more frequently in the study drug arm were almost always observed in the smaller subset,

although those at least two percent more frequent were missed almost half the time.[10] This simulation includes serious adverse events reported by the parallel expedited adverse event regulatory reporting system for "serious" events, which ensures that all toxicities leading to death or hospitalization are captured. Whether such abbreviated data capture is viewed as adequate depends in part on the size of the pre-existing database and the degree to which regulators and health care professionals find that this new information is likely to provide novel insights.[11]

Impact on regulatory review and approval

A shared understanding between regulators and sponsors regarding the quality and quantity of data that are necessary to adequately weigh a treatment's risks and benefits will benefit all. Clinical trials could be designed and conducted more efficiently and the regulatory review process could be more uniform and rapid if a set of data collection and reporting standards were consistently applied to clinical trials conducted by industry, academia, and the NCI's Cooperative Groups. The challenge, however, is in balancing the need for efficiency in the conduct of trials with the need for developing adequate evidence to demonstrate safety and efficacy and for ensuring adequate labeling to inform clinical use. In addition, standards for individual data elements, as well as standard data collection approaches, need to be developed.

A proposed framework for data collection standards

Ideally, cancer clinical trial data collection standards would be based on what is already known about the treatment under investigation, the objectives of the study of the treatment, the study population and the intended use of the agent. Moreover, the standards would be flexible enough to deal with specific subsets and risk groups of interest.

When little is known about the therapeutic agent, as would usually be the case at the time of the primary (initial) cancer indication, the studies should include fairly extensive data collection.* This is congruent with the expectation of international regulatory bodies and general practice for pharmaceutical development. The FDA has approved cancer therapeutics with substantially less comprehensive data collection and in the absence of substantial prior existing data; however because of the great uncertainty associated with oncology drug development, this approach is risky and not recommended. Cancer therapeutics are particularly likely to fail in development compared to other therapies – one estimate shows that 60 percent of cancer drug development programs fail in the late clinical phase. On the chance that the cancer under study is incurable and the new drug offers substantial potential benefit, such as a survival advantage, patients and their physicians are generally willing to accept a greater degree of uncertainty about side effects. Too often, however, the efficacy is substantially less than anticipated. That, coupled with the inability to characterize the safety profile, may result in the need to conduct more trials.

A complete set of standards would identify a "core set" of data elements, along with recommendations about which elements to include in protocols and case report forms under various circumstances. Ultimately, such standards would be endorsed by regulatory agencies worldwide, so that evidence based on these standards could be globally applied.

All parties concur that a substantial amount of data reduction is not appropriate when little is known about the agent; usually this will be in the setting of the initial indication. Even in this setting, however, there is not universal agreement on what is essential, and a range of practices exist. Table 1 is intended to reflect what has been the range of data collection practices in the setting where the safety profile has not been established.

In contrast, for supplemental approvals or when substantial safety data are available, data reduction may be considered – for example, where the agent has previously received full approval (in contrast to accelerated approval) based on a randomized controlled trial and a substantial safety database of many hundreds of patients. Table 2 is intended to reflect areas where data reduction may be appropriate.

* Extensive data may already be available about the agent, such as from studies in other indications, or earlier studies in certain cancers that failed to demonstrate efficacy but generated substantial safety data. In such cases, it may be appropriate to reduce data collection in the new study intended to support an initial approval.

Most clinical studies intended to support a marketing claim will be the subject of an End-of-Phase II meeting with the FDA, which provides the opportune time to refine and modify the data collection elements as the situation warrants. The Brookings conference is the beginning of a process to develop a common approach to data collection for therapeutic oncology trials; identification of a common approach will, in turn, inform and improve the End-of-Phase II meetings.

The core set includes the following categories of data elements: eligibility, on-study form, medical history, lab findings, disease measurement, treatment, vital signs, non-protocol therapy, long-term follow-up, concomitant medications, and toxicity.

The overriding principle is that less data may be collected when substantial safety data are already available; ordinarily, we consider this in the context of secondary versus primary indications, although occasionally substantial data are available at the time of the initial indication. Regardless of whether the indication is primary or secondary, when the safety profile of the agent has been well-characterized, the burden on investigators, human subjects, sponsors, regulators, and patients in need of the treatment under study can be lessened without compromising the adequate evaluation of risk and benefit. It could also be argued that, by focusing investigator efforts on critical safety variables, that more accurate and complete safety reporting of the most serious events will be facilitated.

For drugs without a well-characterized toxicity profile, most participants agree that collection of all Grade 3-4 AEs, at least a subset of Grade 1-2 AEs, and all concomitant medications is appropriate. For both primary and secondary indications, investigators would always collect data on deaths, serious adverse events, and adverse events leading to the discontinuation of treatment.

For secondary indications (i.e., when data are deemed sufficient to characterize the safety of the agent), some members of the panel suggest that investigators collect data on Grade 3-4 adverse events by cycle in either a subset of sites or another subset of patients. In addition, based on the known biology and safety profile of the study drug, targeted AEs could be collected in all patients. Similarly, targeted concomitant medications could be collected in a subset of sites or patients based on the pharmacology of the drug and its known safety profile.

The suggested new approaches for concomitant medications and toxicity might lead to concern that insufficient data will be available on which to assess safety and formulate risk/benefit assessments. If less data on concomitant medications were collected, investigators might overlook possible drug-drug interactions that they might otherwise catch. Likewise, collecting less data on toxicity during trials that study secondary indications could be argued to risk underestimating the rates of a treatment's adverse effects. This might be a particular concern when the populations for initial and supplemental indications differ greatly, such as a third-line setting in advanced cancer for initial and adjuvant setting for supplement. The benefit/risk analysis may differ, and assessment of lower-grade adverse events potentially might become more important. These issues will be raised during the panel discussion.

Minimizing collection of redundant or unnecessary data in these two categories when testing secondary indications could potentially make trials more efficient without posing additional risk to patients, especially in the context of adequate post-marketing surveillance. It would be important in this context to specify what designs would make post-marketing surveillance "adequate" to detect events that could have been missed using this type of data collection. When a treatment undergoes testing for secondary indications, data on adverse events – including complete sets of information on Grades 1 through 4 adverse events, deaths, drug discontinuations, dose decreases and delays, chemotherapy modifications, and serious adverse events – should already be available. Moreover, in many circumstances, the additional information gained from the full cohort of subjects may not be more informative than that based on a subset. Empirical studies and statistical simulations have

demonstrated that while increasing the number of subjects can narrow confidence intervals, the point estimate for AE rates are relatively stable under small samples. As a result, the amount of AE data gathered should be carefully considered, particularly in instances where a treatment is undergoing testing for a secondary indication.[12]

Beyond the framework outlined here, it may also be useful to establish an additional set of standards related to post-marketing evaluation. The conduct of post-approval trials according to appropriately designed standards might facilitate the more rapid approval of certain therapies. Such standards would aim to increase precision around safety and efficacy estimates – perhaps in specific subpopulations. Under the FDA Amendments Act of 2007 (FDAAA), the FDA can require post-marketing studies to further characterize the safety profile of an approved drug or biological product. Implementation of the FDA Sentinel Initiative to conduct population-level post-marketing surveillance will be an additional step toward ensuring that gains in efficiency are not offset by risk to patients.[13]

Electronic data reporting

These recommendations identify what investigators should collect, but do not address how the standard items should be collected. We support the collection of clinical trial data through electronic means using consistent definitions and formats across trials. The connection of data collection standards and electronic data reporting standards, as developed by the Clinical Data Standards Interchange Consortium (CDISC) and Health Level 7 (HL7), can further increase clinical trial efficiency and speed up regulatory review.

CDISC and HL7 are two principal groups that have taken on the task of creating clinical-research reporting standards. CDISC has created models to standardize the submission of data to regulators, and HL7 produces standards for clinical and administrative data. In order to promote interoperability among applications using CDISC or HL7 standards, the FDA, NCI, CDISC, and HL7 created the Biomedical Research Integrated Domain Group (BRIDG).[14] The BRIDG model, released in 2007, has been adopted by the NCI's cancer Biomedical Informatics Grid (caBIG) initiative.

CaBIG is an open-source, open-access, open-development, and federated platform that facilitates the collection and sharing of standardized data across participating sites, and shows great potential for increasing the pace of innovation in cancer treatment. CaBIG offers participants tools that aid clinical trials management, integrate and analyze various types of data, and assure secure data-sharing connections among the 56 NCI cancer centers and 16 community health centers in the network. Whenever possible, investigators should utilize the CaBIG platform and adhere to accepted standards for data reporting. Of particular interest will be the forthcoming proposed rule for *Electronic Submission of Data from Studies Evaluating Human Drugs and Biologics*, expected in September 2008.

Areas for panel discussion, with a focus on scenarios where an adequate safety database exists

- Data collection on CRFs for toxicity;
- Data capture on CRFs for concurrent medications;
- Types of studies in post-marketing that can provide robust safety information; and
- A request for industry to conduct retrospective and prospective evaluations of differing amounts of toxicity or concomitant medication data collection in their clinical trials.

Conclusion

Development of standards for data collection – both qualitative and quantitative – can improve the efficiency of cancer clinical research and development. Adoption and consistent implementation of such standards throughout the development process could help facilitate the rapid development of safe and effective cancer therapies.

Table 1. Standard practices for data collection when little prior data available†

DATA TYPE	SCOPE OF COLLECTION
Eligibility	Collect major inclusion/exclusion criteria (e.g. PS, disease or treatment characteristics) as individual yes/no boxes and remaining eligibility as a single yes/no on a case report form (CRF); do not collect source data (e.g. labs, scans).
On-Study Form	Collect all relevant patient and baseline characteristics.
Medical History	Collect targeted baseline medical history in checkbox format (e.g. diabetes, hypertension requiring treatment, history of myocardial infarction).
Physical Exam	Variable: (ranges from all to none)
Lab findings	Varies: from all routine laboratory values at baseline and during treatment to subset, and via central lab vs. site laboratory.
Disease Measurement	Collect all tumor assessment measurements at all time points.
Treatment	Collect actual dose and treatment date, or reason for modification, delay, hold, or discontinuation.
Vital Signs	Collect routine vital signs. Collect weight/height or body surface area (BSA) on initial Treatment page. If a change from the initial dose, a reason (weight change, toxicity, protocol specified, etc.) must be provided.
Non-Protocol Therapy (NPT)	Collect all NPT (but not doses), including start and stop date (month/year), until first progression. (Need to clearly define what therapies are included.)
Concomitant Medications	Collect all concomitant medications at baseline by name; practices vary from all start and stop dates, to by cycle.
Toxicity	Collect deaths, Grades 3-4 toxicity, serious AEs, AEs leading to discontinuation of treatment. For Grades 1-2, practices range from collection of all grades with start and stop dates, to all grades by cycle, to collection of Grades 1-2 in a subset.
Long-Term Follow-Up	First treatment initiated after disease progression; dose and duration of treatment not needed.

† This usually applies to studies supporting initial indications; the table shows a range of practices.

Table 2. Data collection for secondary indications or where substantial data exist‡

DATA TYPE	SCOPE OF COLLECTION
Eligibility	Collect major inclusion/exclusion criteria (e.g. PS, disease or treatment characteristics) as individual yes/no boxes and remaining eligibility as a single yes/no on a case report form (CRF); do not collect source data (e.g. labs, scans).
On-Study Form	Collect all relevant patient and baseline characteristics.
Medical History	Collect targeted baseline medical history in checkbox format (e.g. diabetes, hypertension requiring treatment, history of myocardial infarction).
Physical Exam	Do not record physical exam on CRF.
Lab findings	Do not collect routine laboratory values (except in the case when they are eligibility criteria or where certain targeted laboratory data are important) at baseline or during treatment except as adverse events. However, if there is a lab-related serious adverse event (SAE), the SAE should include whether the patient's initial value was normal, prior treatment values that were abnormal and related history.
Disease Measurement	Collect all tumor assessment measurements at all time points.
Treatment	Collect actual dose and treatment date, or reason for modification, delay, hold, or discontinuation.
Vital Signs	Do not collect routinely except where certain targeted vital signs are important. Collect weight/height or body surface area (BSA) on initial Treatment page. If there is a change from the initial dose, a reason (weight change, toxicity, protocol specified, etc.) must be provided.
Non-Protocol Therapy (NPT)	Collect all NPT (but not doses), including start and stop date (month/year), until first progression. (Need to clearly define what therapies are included.)
Concomitant Medications	Needs further discussion Current proposal: Collect targeted concomitant medications by specific name based on safety profile of drug. Collect at baseline and when a SAE occurs.
Toxicity	Needs further discussion Current proposal: Collect deaths, targeted AEs, serious AEs, AEs leading to the discontinuation of treatment; collect Grades 3-4 events by cycle at subset of sites (or patients).
Long-Term Follow-Up	First treatment initiated after disease progression; dose and duration of treatment not needed.

‡ Data collection refers to data that are specifically recorded on case report forms; it is expected that all patients will receive routine evaluations (physical examination including vital signs, laboratory evaluation, etc.) as per standard of medical care.

References

[1] U.S. Food and Drug Administration. Reviewer Guidance: Conducting a Clinical Safety Review of a New Product Application and Preparing a Report on the Review [online], <http://www.fda.gov/CDER/guidance/3580fnl.pdf> (2005).

[2] U.S. Food and Drug Administration. Guidance for Industry: Premarketing Risk Assessment [online], <http://www.fda.gov/CDER/guidance/6357fnl.pdf > (2005).

[3] U.S. Food and Drug Administration. Guidance for Industry: Cancer Drug and Biological Products - Clinical Data in Marketing Applications [online], <http://www.fda.gov/cder/guidance/4332fnl.htm> (2001).

[4] Bent, S., Padula. A. and Avins, A.L. Brief Communication: Better Ways to Question Patients about Adverse Medical Events: A Randomized, Controlled Trial. *Ann. Intern Med.* 144, 257-261 (2006).

[5] Bent, S. Padula, A. and Avins, A.L. 2006.

[6] Ioannidis, J.P.A., Mulrow, C.D., and Goodman, S.N. Adverse Events: The More You Search, the More You Find. *Ann. Intern. Med.* 144, 298-300 (2006).

[7] Scharf, O. and Colevas, A.D. Adverse Event Reporting in Publications Compared With Sponsor Database for Cancer Clinical Trials," *J. of Clin. Oncol* 24, 3933-3938 (2006).

[8] Ioannidis, J.P.A. and Lau, J. Completeness of Safety Reporting in Randomized Trials: An Evaluation of 7 Medical Areas. *J. Am. Med.* Assoc. 285, 437-443 (2001).

[9] Genentech, data on file.

[10] Genentech, data on file.

[11] Genentech, data on file.

[12] Mahoney M.R. et al. Dealing with a Deluge of Data. *J. of Clin. Oncol.* 23, 9275-9281 (2005).

[13] U.S. Food and Drug Administration. FDA Sentinel Initiative: A National Strategy for Monitoring Medical Product Safety [online], <http://www.fda.gov/oc/initiatives/advance/reports/report0508.html> (2008).

[14] Richesson, R. and Krischer, J. Data Standards in Clinical Research: Gaps, Overlaps, Challenges and Future Directions. *J Am. Med. Inform. Assoc.* 14, 687-696 (2007).