Developing CT Image-Processing Tools to Accelerate Progress in Lung Cancer Drug Development

November 01, 2006
By James L. Mulshine, MD [1], Fred R. Hirsch, MD, PhD [2], and David F. Yankelevitz, MD [3]

On April 21, 2005, the Cancer Research and Prevention Foundation (CRPF), in conjunction with academic researchers, federal scientists, lung cancer advocates, and representatives of a number of pharmaceutical and diagnostic imaging companies, participated in a workshop held in Annapolis, Md, on the development of high-resolution spiral computed tomography (CT) imaging tools to assess therapeutic response in lung cancer clinical trials. In this report, we will address developments that led up to that workshop, what was discussed, and recommendations that came out of the meeting.

Lung cancer is the leading cause of cancer death throughout the world. Despite massive public health efforts, global tobacco consumption—the driver of the lung cancer mortality—continues to grow.[1] In the United States, tobacco-control efforts have had significant success, resulting in a large cohort of close to 50 million former smokers. Unlike the observed decline in cardiovascular disease risk after smoking cessation, former smokers accrue a persistent lifelong elevated risk of lung cancer. This trend may account for the recent emergence of lung cancer as the leading cause of tobacco-related death. The “War on Cancer” is being reinvigorated with the goal of significantly improving outcomes by 2015. Given the dominant impact of lung cancer mortality, it is timely to consider opportunities for strategic breakthroughs in improving outcomes in lung cancer.[2]

Progress has been modest in improving outcomes for patients with this disease. At the time of initial diagnosis, lung cancer is typically already established in regional or distant metastatic sites, meaning chemotherapy is the mainstay of treatment. Chemotherapy is rarely curative in this setting, but recent randomized reports suggest that chemotherapy added to surgical management of early-stage lung cancer is associated with a significant improvement in 5-year survival.[3-5] This finding suggests that as with breast cancer, chemotherapy in treating early lung cancer-potentially with new targeted therapies—may have greater impact than in treating late disease. Chemotherapy has been recently shown to convincingly reduce the rate of recurrence after surgical management in several major studies, including one of a well tolerated oral drug from Japan.[6]

The US Food and Drug Administration (FDA) has signaled that it would favor the evaluation of adjuvant therapy with tyrosine kinase inhibitors in conjunction with a predictive test to predict responsiveness. This opportunity is already the subject of clinical trials.[5] The purpose of this research is to find effective therapy that is also safer and potentially less toxic to improve outcomes in patients with early lung cancer, so there are promising developments in the area of lung cancer therapeutics.

Drug Development

There has been considerable interest in improving the effectiveness of drug therapy for lung cancer, but this process has been slowed by the enormous cost and long duration of drug development.[7] One approach to this problem is the use of tumor imaging as a surrogate marker of disease response, which allows for more rapid validating clinical trials than waiting for standard clinical endpoints. The precedent for this approach would be the use of cholesterol levels for evaluating lipid-lowering drugs rather than evaluating the frequency of heart attack or cardiovascular death.[8] Over the past 3 decades, the transition from surgically based management of advanced cardiovascular disease to medical management of early cardiovascular disease has been associated with a 4-year increase in average life expectancy for all Americans.[8] This achievement provides a model of success that is attractive to pursue in the setting of lung cancer, where success in treating advanced disease has been elusive.

Spiral CT

Recent promising reports have suggested that spiral CT may be a more effective tool for early detection of lung cancer than chest x-ray.[9,10] The enhanced sensitivity of this rapidly improving tool is presenting a challenge for conventional radiologic interpretation in the detection of early lung cancers. In response to that challenge, the National Cancer Institute (NCI) and others (the Lung
Image Database Consortium, or LIDC[11] have begun an innovative collaboration to collect a database of spiral CT images generated to find early lung cancer. This resource can be used as a validation matrix to assist in the development of computer-assisted detection (CAD) software. The database would allow software algorithm development and validation.

Spiral CT imaging is particularly suited for such a database resource since a standardized CT image file format (DICOM) already exists. This means that transporting and storing CT images for pooled analysis can be done readily.

The need for these tools is already pressing, as the current generation of CT scanners are capable of considerably higher resolution than is used in standard clinical practice. A growing problem with this enhanced sensitivity is the inability of the radiologist to manage and extract all useful clinical information from a high-resolution CT study. To address this problem, major CT vendors and others have been working to develop software tools to enhance workflow with the large CT image files. However, there remains a challenge in understanding the clinical significance of some of the newly identifiable features evident on high-resolution CT images. The difference between the potential information on a CT image and the ability of a reader to appreciate that information has been termed a "software gap."

Computer-Assisted Detection

Important questions concern whether CAD approaches have relevance to the assessment of drug response and whether they could improve on current approaches as codified in the Response Evaluation Criteria in Solid Tumors (RECIST) guidelines.[12] RECIST is the only validated tool for lung cancer drug evaluation in a regulatory setting. This metric involves the measurement of tumor size on CT scan in one dimension. Its utility was established by comparing unidimensional measurement to the bidimensional World Health Organization (WHO) criteria in a study of over 4,000 patients from 14 different clinical trials. While RECIST and WHO criteria generally result in comparable assessments of response status (partial or complete) and disease status (stable or progressive), the RECIST determination involves subjectivity. Therefore, RECIST is frequently associated with response classification errors.

Since the time that RECIST was validated, there have been explosive improvements in the speed and resolution of spiral CT. Current 64-detector CT systems are capable of acquiring high-resolution imaging of the thoracic cavity (0.6-mm slice thickness) in a matter of seconds.[13] While it remains to be determined how to best use this enhanced-resolution imaging capability in a clinical trial setting, recent FDA trends suggest a growing appreciation of the utility of imaging as a practical metric for drug approval.[14] In light of CAD developments in the early detection of breast cancer and lung cancer, it may be that computer-assisted image-processing approaches represent an opportunity to improve decision-making in a variety of clinical settings.

An important question is whether CAD approaches also have relevance to the assessment of drug response and could improve on the current approach to cancer image evaluation with RECIST. Like the LIDC initiative, it would be useful to have a database of characterized lung cancer CT images from individuals undergoing chemotherapy to allow the development and validation of CAD capabilities for drug response assessment in this setting.

First Workshop

Recognition of this opportunity was the impetus for the first CRPF lung cancer imaging response monitoring workshop, held April 15-16, 2004, in Washington, DC. This initial forum also brought together academic researchers, federal scientists, lung cancer advocates, and representatives of a number of pharmaceutical and diagnostic imaging companies to consider the potential of high-resolution spiral CT imaging tools for the assessment of therapeutic response in lung cancer clinical trials.

The remarkable improvement in spiral CT capabilities are in part related to technical refinements of the CT instrument with image acquisition, but the real driver has been the improved microprocessor capability allowing acquisition of profoundly larger amounts of imaging information.[10] Considerable tension surrounds the rapid evolution of spiral CT imaging capabilities, as the resolution of CT scanners has doubled every 2 years for the past decade and this rate of improvement appears to be sustainable for the foreseeable future. Although the progressively greater data acquisition with higher-resolution CT studies represents a challenge for radiologists, this data load provides a richer information source from which image-processing can characterize the precise nature of the intrathoracic structures. After appropriate validation of this capability, clinical researchers would have an objective quantitative tool with which to establish the response of cancers to the administration of therapeutic drugs, potentially allowing a more efficient process of drug development.
Challenges and Prospects

At the first workshop, participants discussed the development of image-processing tools and computer-assisted diagnosis of lung cancer to catalyze progress in lung cancer research. To explore the challenge, representatives of the FDA reviewed the regulatory implications of using changes in tumor volume as a metric for clinical drug response. An NCI investigator discussed the current standard for phase II drug evaluation (as outlined in RECIST), involving serial comparisons of a single measurement of the diameter of a tumor in one dimension. The benefits and limitations of this standard were discussed, as were the steps required to validate a new standard for imaging-based evaluation of drug response.[15]

A small number of diagnostic imaging companies dominate the development of helical CT imaging devices and computer-assisted diagnostic tools. If these powerful new capabilities are developed such that imaging studies can be compared across different commercial platforms, it would greatly facilitate the delivery of health care. It could also represent a major reduction in costs for governments, insurance providers, and the pharmaceutical industry. The prospects were discussed for enabling cross-platform comparisons of image changes from serial scan of the same individual across time points using different manufacturer's equipment. The aging of the baby boomers will be associated with an increased number of lung cancers, which comes at a time of unprecedented health-care costs. Developing tools that can accelerate the identification and validation of more successful drugs for lung cancer treatment would be a strategic breakthrough.

Since the focus of the 2004 workshop was to explore the prospects for developing effective CAD tools, there was considerable discussion about the best way to validate CAD tools. CT software development is typically commercialized under the premarket notification (510[k]) process of the FDA.[16,17] The 510(k)-type regulatory approval is the most accessible level of CAD approval and allows a vendor to sell a software-based measurement tool. For software tools to be used as a basis for clinical management requires a more rigorous level of FDA review called a premarket approval application (PMAA). A major factor in this application process is the need to convincingly prove the robustness of a device for a particular application. A very large image database is required to validate convincing software performance given the biologic range of variation that would be reasonably expected in the relevant clinical setting. In addition, the images for the database have to be acquired on the same types of device and sensitivity settings that are anticipated for the eventual clinical application. With the rapid technologic evolution of CT scanners, developing and sustaining such an expensive validation resource is thought to be beyond the resources of a single institution. Consequently, developing this resource emerges as an important area for shared development.

Consensus and Concerns

A point of consensus from workshop participants was the strategic value in developing a large image/clinical outcomes database. This was agreed to be an essential resource to facilitate the development and validation of CAD and related image-evaluation tools for drug responses with high-resolution CT scanning. A working coalition of clinical and imaging scientists emerged, committed to developing image analysis for drug response evaluation for lung cancer clinical trials. This group defined a series of concrete steps to advance the field. These steps included the definition of techniques for scanner calibration to ensure the acquisition of quality images as well as the potential design of phantoms to enable cross-platform imaging comparisons.

Representatives from the NCI reported on their ongoing collection of high-resolution images to catalyze software development for the early detection of lung cancer. In the course of this work, particular attention has been focused on the optimal development of an imaging database. A particular concern relates to the establishment of "ground truth" in regard to whether a CT image file represents a "false-positive" condition or a confirmed cancer case.

Second Workshop

Based on feedback from the initial workshop regarding strong potential for real progress, the consensus was that an ongoing workshop forum was needed for academic and federal scientists as well as medical imaging and pharmaceutical industry representatives to continue developing their future vision of image-processing technology and computer-assisted diagnosis. It was agreed to meet again to discuss these issues as they relate to lung cancer treatment and prevention applications, to facilitate improvements in lung cancer outcomes. The purpose of the second workshop is summarized in Table 1. A major impetus in sustaining this forum was that lung cancer was a particularly strategic focus for research since the disease is so lethal. Furthermore, many of the challenges in validating lung cancer CAD tools can be shared with other high-resolution imaging tools.
The second workshop was convened on April 21-22, 2005, in Annapolis, Md, with the intention of discussing interval progress. From the comments of the pharmaceutical representatives, it was clear that lung cancer was growing as a high-priority area of drug development with many potential drug targets being evaluated in clinical trials. It was also evident from the results of success with targeted therapies in particular lung cancer settings[4,18] that many disease subtypes and distinct patient populations could be of interest.

The rapid improvements with CT imaging appeared likely to continue, necessitating the continuous collection of relevant image data to sustain the development of evolving CT platforms. That is, higher-resolution imaging will continue to find earlier phases of lung cancer that have never been seen before. It is therefore critical to continuously collect appropriate forms of ground truth from large image databases to ensure rigorous CAD validation.

The concern of investigators at the workshop was whether the number of cases in the image/outcome database will be sufficient not only to capture the variance in patient conditions and lung cancer presentation, but also to reflect the expected variability in scanner acquisition parameters. The consensus was that the precision of the volumetric tool will be directly related to the amount of high-quality data available, the capability of the acquisition system used, and the sophistication of the algorithms employed. From the discussion, the ideal target size of the database ranged from 500 to over 1,000 cases of serial images.

Measuring Change
Several academic and commercial groups outlined the advances being made in measuring size and change of subcentimeter lung lesions, with the potential use of these methods to assess therapy. They pointed out, however, that significant algorithmic challenges remain in measuring change of the larger lesions present in late-stage lung cancer. It was evident from the algorithm-development community that in addition to advancing volumetric segmentation algorithms, new approaches to measuring change need to be investigated. These approaches include measuring change using image deformation and subtraction (which has been investigated in early Alzheimer's disease detection in magnetic resonance imaging exams) and approaches that factor in additional image metrics, such as lesion texture and vascular properties.[19,20]

In addition, it was recognized that significant research is needed to explore, identify, and validate the method(s) that best assess the outcome of therapy. Likewise, it will be important to establish which algorithmic techniques are resilient to variation in scanner capability, as is commonly found when recruiting sites for large clinical trials. Scanner variability exists not only from one patient to the next but also occurs within a patient's longitudinal study, as imaging may be performed at different locations with different scanners, potentially changing throughout the study. The current rapid evolution of scanner capability ensures that variation in acquisition will continue to be a challenge to algorithms for the foreseeable future.

Reducing Technical Variability

During the discussions, a number of additional technical issues emerged. For example, although the newer scanners, including those with 16 or more detectors, are capable of acquiring very-thin-slice information (1.25 mm or less), in clinical practice most radiologists are still using the much thicker 2.5- to 5-mm slice thickness. The concern expressed by the image vendor community was that the thicker images would contain less precise information for optimal CAD-based volume determination. This would be particularly important in "segmentation," which is the process of defining the borders of adjacent structures. Considerable discussion focused on the relative merits of acquiring thicker-slice images that reflect current clinical radiology practice vs images with slice thickness less than 1.25 mm, where the prospect for successful volume analysis may be more favorable.

While some investigators contended that CAD algorithms should be robust enough to handle even suboptimal clinical images, others argued that the time to impose quality standards for imaging to minimize measurement variability through time is now, ie, early in the development of the field. Since it is not clear how these tools will evolve, it was felt that a broad range of images (acquired under variable degrees of resolution) should be obtained, to address all possible scenarios. Compounding this problem is the fact that standardized scan acquisition parameters have never been established to ensure that scans obtained for volume analysis are generated in a way that reduces technical variability. Unresolved issues include standardization of radiation dose, instrument calibration approaches, or acquisition parameters for the scanner itself to normalize for patient size and geometry.

Other Technical Issues

Further discussion focused on whether volumetric measures of tumors implied merely defining new standards as to how much of a change constitutes a meaningful clinical response. Alternatively, does the enhanced resolution provide more information that links more reliably to ultimate clinical outcome? These issues led to a technical discussion on how to define imaging endpoints and acquisition parameters to allow for robust measurements. Ultimately, it was agreed that the accuracy of the new tools may vary depending on the feature (subject to location, geometry, density, etc) that was being evaluated, and so performance standards need to be established. Such qualification will require the type of well-defined database that was proposed in the first workshop. Other critical imaging issues for using these tools in clinical management relate to cross-platform compatibility among different manufacturers. Inevitably, with patients being evaluated for drug response, a different vendor's instrument will be used for the baseline and final imaging studies. In this regard, the performance of volume change or a related analysis would be subject to the conditions required for the drug evaluation, which must be defined in a regulatory setting. This situation again underscores the need for a large image database to allow the responsible development of this class of drug response evaluation tools.

Third Workshop

At the most recent workshop, held on April 19-20, 2006, in Bethesda, Md, participants explored the implications of the fast pace of improved performance of CT scanners. It was agreed that current scanners already achieve high enough resolution when looking for large changes in lung cancer volumes. If it becomes necessary to look more quickly for very small changes in tumor characteristics, then new calibration methods will be necessary. This capability would be particularly...
important in looking at serial imaging over short intervals to determine drug responsiveness. This use of imaging as a biomarker of drug response is consistent with FDA efforts to responsibly accelerate drug approval under their new Critical Path Initiative (www.fda.gov/oc/initiative/criticalpath/). Other metrics of responsiveness such as tumor nodule density may also be of value. Such new metrics require standardization and calibration of the image-acquisition protocol (Table 2). Such provisions could simplify image data analysis. Participants suggested that some level of measuring compatibility was emerging, and there are currently some standards for calibration.

Table 2
Areas of Consensus on Image-Acquisition Parameters

- Calibration requirements
- Need to address phantom issues
- Need to populate database
- Need to achieve true cross-platform functionality
- Acceptance of database as critical validation resource
- Need for ongoing attention to progress

Role of Industry and FDA
Further dialogue, especially among the imaging manufacturers, is needed to resolve many of these issues, but the clear message from the pharmaceutical industry was a desire to have access to such powerful imaging tools that could evaluate response across vendor platform types. We urgently need to establish proper methodologies in this regard, to test algorithm performance in drug assessment. A consensus emerged as to the FDA being a critical participant in this process, including divisions of the FDA responsible for clinical trial drug evaluation (Center for Drug Evaluation and Research) and device evaluation (Center for Devices and Radiological Health). A representative from the FDA outlined the agency's extensive interest in this promising yet challenging area. An important issue from a regulatory perspective is establishing the robustness of imaging information.

Phantom Development and Data Sharing
During the workshop, the suggestion was made several times for the field to consider the use of a phantom to assist in validating the reproducibility of a spiral CT scanner's sensitivity settings. Phantom studies would assist in making imaging comparisons across vendor platforms, which may be of value when comparing serial images acquired over time. Development of a standard phantom model would also have utility in CAD algorithm development and validation. The issue of sharing data from drug research trials was discussed. While some representatives of the pharmaceutical industry expressed sensitivity regarding this issue, most companies were quite comfortable with sharing anonymized data from experimental drug trials. It was noted that the development of image databases prompts Health Insurance Portability and Accountability Act (HIPAA) and institutional review board (IRB) issues that must be addressed. Further, representatives of pharmaceutical companies expressed a willingness to insist on standardization of imaging measures with new clinical trial protocols.

Conclusions
A summary of the conclusions and recommendations from the workshop appears in Table 3. A strong motivation for this forum has been the shared recognition that no single institution can bear the recurring costs of keeping a large database current. Yet the pace of progress in the field was restricted by the lack of such an imaging database. The absence of such a resource would constitute a structural bottleneck to realizing the full benefit of the rapidly improving resolution of spiral CT in managing the lethal consequences of lung cancer.
The collaboration embodied in this process involves a public-private partnership to validate imaging tools for the application of CAD in lung cancer management. It was understood by all parties that this effort is a model for developing other emerging high-resolution imaging tools for disease management so that the shared goal of improving patient outcomes can be realized much more quickly. Future plans for this effort include ongoing discussions about the refinement of strategies for optimal database design.

Additionally, it is critical that information about the existence of the database continues to be disseminated. Finally, and most critically, more contributors must be identified to provide high-quality, serial images with clinical follow-up information as available, to rapidly populate the database. These measures will greatly accelerate the maturation of this promising field.

**References:**


