

Integrating clinical, CT and PFT patient information in a database to determine a follow-up CT interval and the malignant potential of solid and ground glass pulmonary nodules

M. Khilkin, DO¹, C. Mattmann, PhD², P. Rivera, MD¹, J. Ko, MD¹, B. Phalan, BS¹, E. Eylers, RN¹, Sean Kelly, BS², D. Crichton, MS², and W.N. Rom, MD, MPH¹.
¹New York University School of Medicine, NY. ²Early Detection Research Network Project, Jet Propulsion Laboratory, CA. NIH U01CA86137, khilkm01@med.nyu.edu

BACKGROUND Currently solid pulmonary nodule size and character alone play the major and minor role respectively in assigning a follow-up CT interval. No clear guidelines exist for the interval and length of follow-up of ground glass nodules. We developed a database that integrates information about nodules with additional CT scan abnormalities, patients' demographic information, detailed medical history, PFTs and environmental exposure history to better determine the malignant potential of a nodule to either extend or eliminate the follow-up interval or lead to an earlier intervention.

METHODS The NYU Lung Cancer Biomarker Center has enrolled 1358 smokers and followed them prospectively for up to 6 years for the development of lung cancer. Participants answered a yearly supervised questionnaire that includes demographic information, ATS respiratory symptoms, occupational, family, medical, surgical and female-only history and underwent PFTs and MDCT scanning. Nodules were followed according to suspicion for lung cancer in either 3 and 12 or 6 and 12 months. Solid nodules were followed for at least 2 years and ground glass nodules were followed indefinitely yearly.

RESULTS In collaboration with NYU Radiology and EDRN/JPL we constructed an online MySQL database that allows retrospective and prospective follow-up of nodule size, 5 nodule and 6 adjacent parenchymal features, 121 thoracic and abdominal findings, 6 PFT and 700 questionnaire values. To construct the database, we are leveraging several emerging Java-based open source technologies, including: (1) Apache Tapestry for building a compact, easily maintainable web-based user interface to use for data input and for free-text and forms-based search, and (2) an object-relational persistence technology called Hibernate that affords us the ability to rapidly effect database design changes to our backend CT scan data model and collected data. Using these technologies, we are able to easily provide a means for persisting our collected data into MySQL, an open source relational database management system for further analysis.

CONCLUSIONS An online database that tracks the evolution of a nodule and integrates this information with clinical data can better define the follow-up interval in specific situations and identify earlier nodules with a greater malignant potential.