

## **COLLECTION OF POPULATION LUNG CANCER STAGE DATA BY CLASSIFYING FREE-TEXT PATHOLOGY REPORTS**

*Darren Moore, Anthony Nguyen, Iain McCowan, Mary-Jane Fry*

CSIRO e-Health Research Centre, Brisbane

### **Purpose of Study:**

Evidence-based treatment guidelines for lung cancer treatment are informed by analysis of patient outcomes, where data is first stratified into comparable cases according to the TNM (tumour, nodes, metastases) staging standard. The preferred method for staging lung cancer is through multi-disciplinary team (MDT) conferences. In Queensland, the Integrated Lung Cancer Outcomes Project (QILCOP) collects formal stage data from MDTs. However, due to the resource- and time-intensive nature of MDTs, the state-wide coverage of QILCOP stage data is approximately 50-60% of all lung cancer cases.

The purpose of this study is to develop a prototype system to automatically determine a TNM stage for lung cancer patients using free-text pathology reports stored in clinical information systems. The system uses automatic text categorisation techniques to detect individual observations within reports that are relevant to staging, and to automatically assign a stage.

Such a system could be used to obtain stage data for patients not formally staged by an MDT, allowing more comprehensive population-level analysis of lung cancer outcomes in Queensland.

### **Conclusions:**

A data set consisting of de-identified pathology reports for 710 lung cancer cases was obtained from the AUSLAB pathology information system. From this data, statistical models for automatic text classification were built and evaluated against pathological TNM stages assigned by MDTs. An accuracy of 75.6% has been observed for automatic classification of primary tumour involvement (T1-T4) and 91.6% for classification of lymph node involvement (N0-N2). This work is a partnership between the CSIRO e-Health Research Centre and the Queensland Cancer Control Analysis Team (QCCAT, Queensland Health).

### **Category:**

Population Health

### **Contact Details:**

Darren Moore, CSIRO e-Health Research Centre, Brisbane  
Phone: (07) 30241617 Email: darren.moore@csiro.au