

# International Journal of Population Data Science

Journal Website: [www.ijpds.org](http://www.ijpds.org)



Swansea University  
Prifysgol Abertawe

## Methods for identifying health state transitions from administrative data: the case of metastasis in prostate cancer

Mitsakakis, N<sup>1</sup>, Bremner, K<sup>2</sup>, Croxford, R<sup>3</sup>, Abrahamyan, L<sup>1</sup>, Ryan, W<sup>4</sup>, Carcone, S<sup>4</sup>, and Krahn, M<sup>4</sup>

<sup>1</sup>University of Toronto

<sup>2</sup>University Health Network

<sup>3</sup>Institute of Clinical Evaluative Sciences

<sup>4</sup>THETA Collaborative

### Introduction

Health administrative data are a rich source of population-based information, useful for building state transition models for medical decision making. These models require identification of health state transitions and associated times. Indirect methods are needed to predict this information, as it is rarely available in administrative data.

### Objectives and Approach

We considered a set of criteria to identify transitions to metastasis for prostate cancer patients in administrative data, utilizing dates of diagnostic and medical billing codes for secondary malignancy, palliative radiation therapy, chemotherapy and bone disorders or procedures. We evaluated the criteria using the true date of metastasis from medical charts of 195 patients linked to health care administrative data in Ontario, Canada. We also built a recursive partitioning tree to optimally combine these criteria and construct rules for identifying metastatic patients. For the evaluation, both misclassification and discrepancy between true and predicted dates for the true positives were considered.

### Results

Criteria involving chemotherapy drugs or hospital visits with secondary malignancy ICD10 diagnosis gave the best results, with high sensitivity and specificity. Criteria involving bone related problems, radiation therapy or diagnosis of metastatic cancer in physician billing data were very specific but not sensitive. The criterion involving prescriptions for narcotics was sensitive but not specific. The fitted tree was parsimonious involving only two of the criteria, while improving the accuracy over individual criteria. Most criteria gave a “delayed” prediction, with criterion based on chemotherapy giving on average the smallest delay, as well as exhibiting the least variability. Criteria involving narcotics and bone related problems predicted metastasis date very prematurely, probably triggered

by conditions other than prostate cancer.

### Conclusion/Implications

Several criteria from administrative databases satisfactorily classified prostate cancer patients with metastasis. A classification tree was built and improved the results over single criteria, demonstrating the added benefits in using advanced statistical learning methods for this task. However, “transition to metastasis” dates were predicted inaccurately, often with significant delay.

