ALGORITHMS FOR IDENTIFYING DISEASE PROGRESSION DATES FROM ADMINISTRATIVE DATA: THE CASE OF PROSTATE CANCER

Nicholas Mitsakakis, MSc, PhD, P.Stat
University of Toronto

ICHPS 2018

Disclaimer

- I have no conflict of interest to declare
- This research is funded by a grant from Prostate Cancer Canada (PI: Murray Krahn, THETA Collaborative)

Acknowledgments

- Toronto Health Economics and Technology Assessment (THETA) Collaborative
  - Karen Bremner
  - Lusine Abrahamyan
  - Steven Carcone
  - Welson Ryan
  - Murray Krahn (Director)
- Institute for Clinical Evaluative Sciences (ICES)
  - Ruth Croxford
  - Farzana Haq
  - Erin Mackdonald

Background

- Health administrative data are generated by every encounter with the health care system
- The Institute for Clinical Evaluative Sciences (ICES) holds health care databases capturing interactions for all patients in Ontario
- Their purpose is billing and administrative and not research
- In particular these data do not have the quality and level of detail for determining patient health states, e.g.
  - Diagnostic code errors
  - Missing entries
  - Diagnostic test results not included
- Prostate Cancer: no validated method for identifying metastasis

Motivating application

- A large grant from Prostate Cancer Canada to study pathways to “later” health states of PC patients
  - Factors that affect outcomes (e.g. time to metastasis, time from metastasis to death etc.)
  - Health utilization and cost
  - Transition probabilities among health states
- Being able to identify health state transitions in longitudinal admin data would enable more accurate estimation at the population level

Objectives

- To assess a number of criteria for identifying metastasis for Prostate Cancer (PC) patients
- To evaluate the usefulness of using a statistical learning method, classification recursive partitioning tree, for optimally combining the criteria

Methods

- Chart review data of 195 PC patients with known metastasis status were linked with administrative databases at ICES
- Ten separate criteria from admin data were used to “predict” state of metastasis, and evaluated by sensitivity/ specificity
- A classification recursive partition tree was used to optimally combine these criteria and improve prediction
- Discrepancy between dates of true and predicted metastasis was also measured and used as an assessment criterion

List of databases used

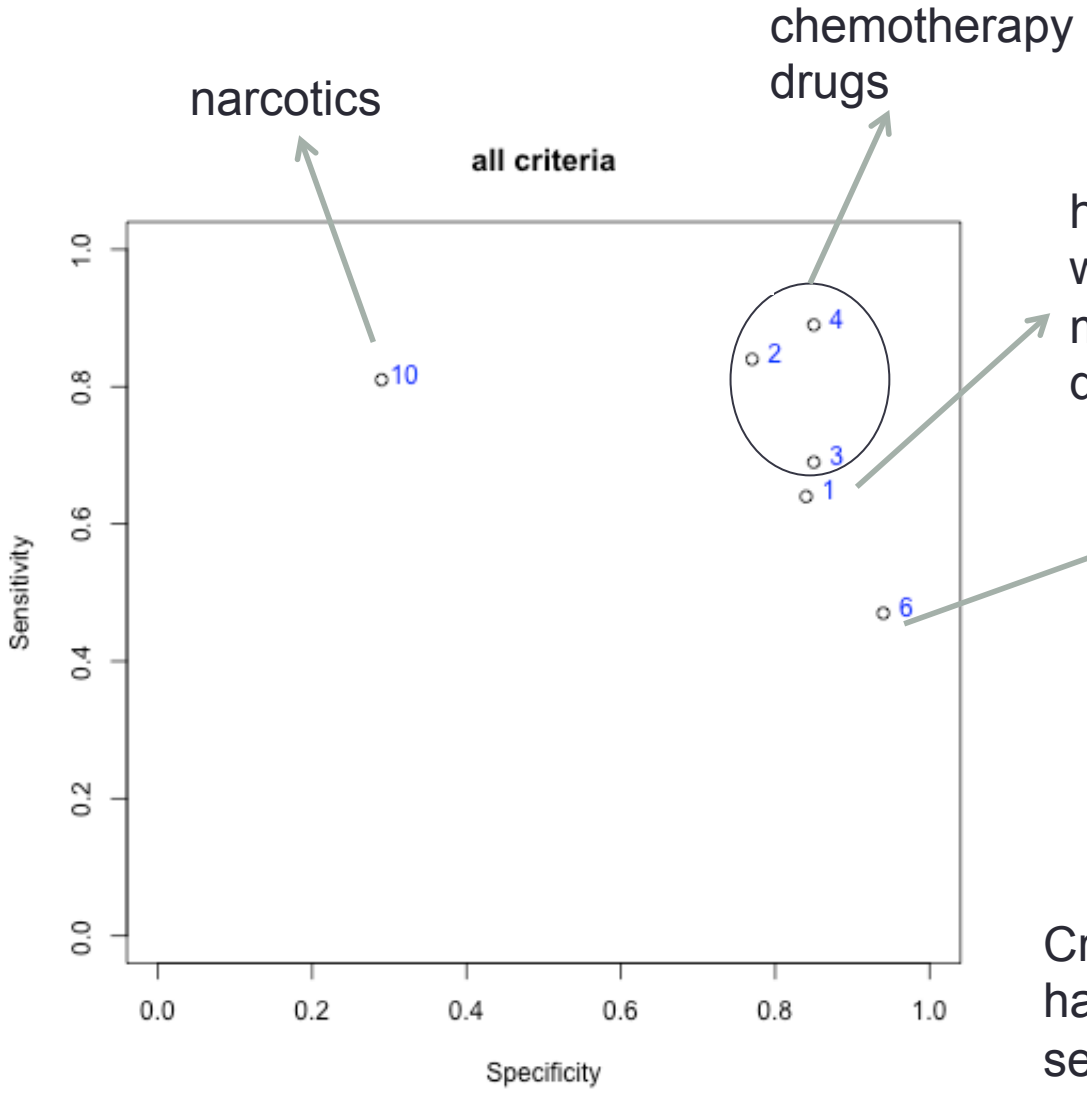
- NACRS (National Ambulatory Care Reporting System) of the Canadian Institute of Health Information (CIHI)
- DAD (Discharge Abstract Database)
- OHIP (Ontario Health Insurance Plan) Claims Database
- ALR (Cancer Activity Level Reporting)
- ODB (Ontario Drug Benefit Claims)
- NDFP (New Drug Funding Program)

List of criteria

1. First date of a hospital visit with secondary cancer, or metastatic cancer diagnosis (ICD10 codes, NACRS, DAD)
2. First date of physician billings for chemo (OHIP)
3. First date of chemotherapy administration visit (NACRS)
4. First date of a chemotherapy drug or other drug for advanced prostate cancer (ODB/NDFP)
5. First date of an OHIP diagnosis code for secondary cancer, or metastatic cancer by physician (OHIP)
6. First date of palliative radiation therapy (OHIP, ALR)
7. First date of diagnosis of pathologic fracture (DAD)
8. First date of spinal cord decompression or compression (DAD)
9. First date of bone surgery (DAD)
10. First date of a prescriptions for narcotics (ODB)

RESULTS

Classification results



narcotics

chemotherapy drugs

hospital visits with secondary malignancy diagnosis

radiation therapy

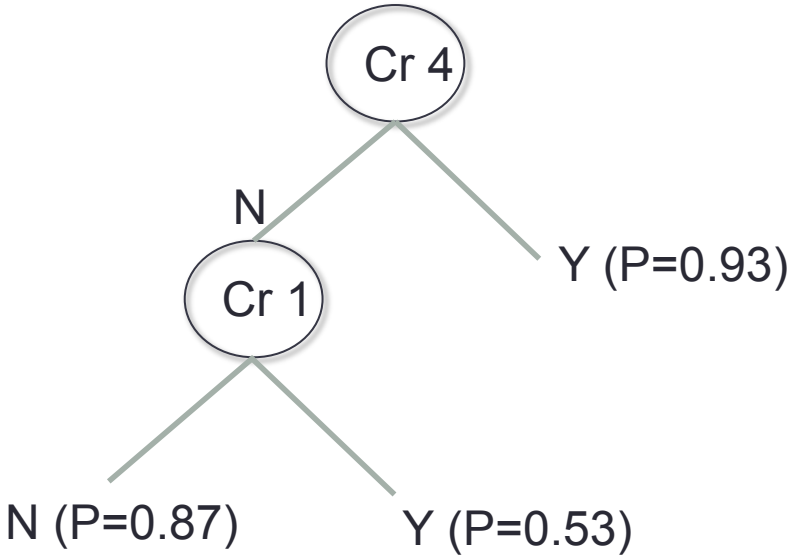
Criteria 5,7,8,9 had a very low sensitivity

Summary of results

- Criteria involving chemotherapy drugs or hospital visits with secondary malignancy ICD10 diagnosis gave the best results (high sensitivity and specificity)
- Criteria involving bone related problems, radiation therapy or metastasis diagnosis in OHIP by physician were very specific but not sensitive
- Criterion involving narcotics was sensitive but not specific

Classification tree

- `rpart` function from synonymous R package gave a *parsimonious* tree involving only criteria 4 and 1

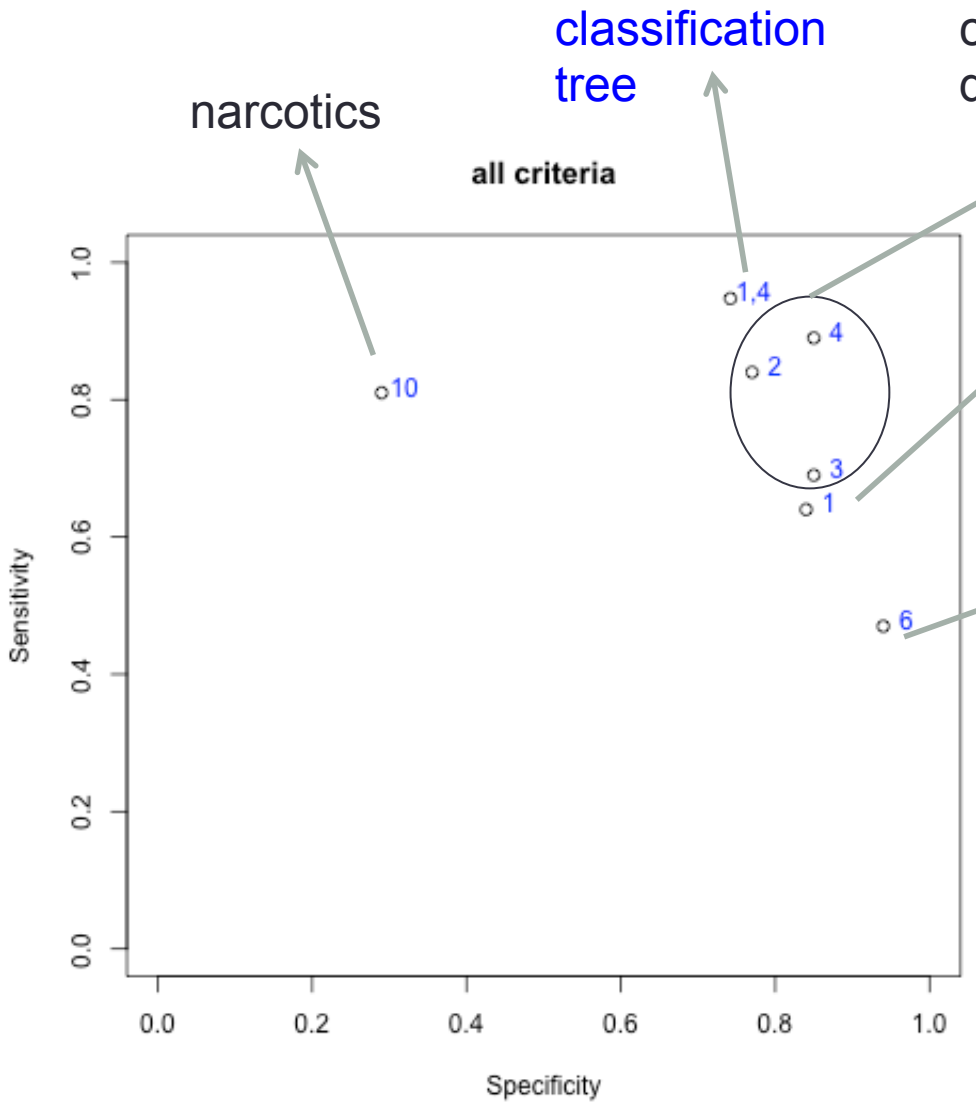


- In other words, if both Cr 1 and 4 are negative => prediction is negative
- Otherwise prediction is positive

Classification results

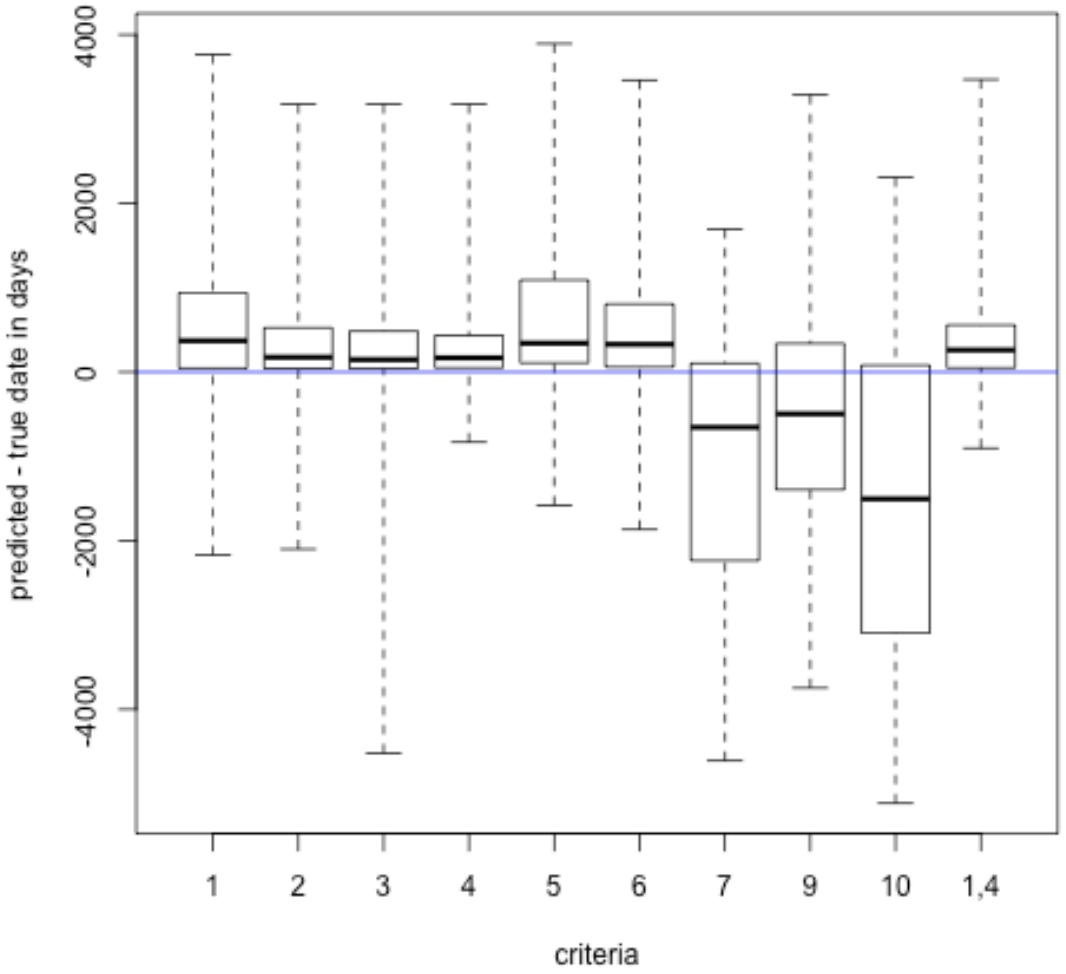
Accuracy

Cr 1 = 0.703
Cr 2 = 0.821
Cr 3 = 0.744
Cr 4 = 0.867
Tree = **0.882**



Criteria 5,7,8,9 had a very low sensitivity

Time differences



- Most criteria gave a “delayed” prediction
- Criterion 4 (based on chemo) gave the smallest bias and variance
- Opioids and bone problems related criteria predicted mets date very prematurely
- Probably triggered by conditions other than mets

DISCUSSION

Discussion

- The main goal is developing a method (algorithm, model) for *predicting* the transition to mets, using admin data
- Methods presented here are progress towards this direction
- Additionally, as a side objective, this analysis can be used for understanding properties of these databases and processes of the health case system

Discussion

- E.g. Not everyone with mets utilized narcotics (what medication did they take?)
- Not everyone used chemotherapy (or it was registered)
- Looking at criterion 5, only 23% of the those with mets are coded by physicians as such (what code did they then use?)
- Also, some non-mets patients are coded as having mets

About the tree

- It is known that tree methods suffer from instability and they can overfit the data
- Here, the resulting tree is very simple and parsimonious – no danger of overfitting
- Validation through cross-validation or Bootstrap will be performed

Conclusions

- A number of criteria from admin databases satisfactorily classified PC patients with metastasis
- A classification tree was built and improved the results
- “Transition to mets” dates were not predicted accurately, they were often significantly late

Some next steps

- Validation of the fitted tree
- A way to incorporate the dates based on different criteria as predictors in the model
- Look at alternative methods for prediction
- Further investigation around the discrepancies between admin databases and true mets state would be useful