KL2 Leadership Meeting

Alex F. Bokov, PhD

September 27, 2018

## Research Progress

### Summary

I have developed scripts for preparing and analyzing data from an i2b2 patient-set originally extracted from a cancer registry (NAACCR) and the EMRs of two health systems (UT Health Epic and UHS Sunrise). I selected and validated the variables necessary to replicate my work on data obtained directly from the Texas Cancer Registry and other NAACCR member registries. I have also confirmed, on a random subset of the data, the technical feasibility of my proposed analysis.

The next steps are fourfold and can proceed in parallel:

* Increase the number of cases available for analysis through the following approaches (Aim 2):
  + Learn the abstraction process used by NAACCR registrars so I can use the same process to abstract the needed data for kidney cancer cases which have not yet been entered into NAACCR
  + Chart review of patients with discrepancies in their records (particularly dates)
  + Re-running the analysis after the next data refresh which will cover a larger number of patients.
  + Work with CIRD to improve coverage of UHS diagnoses and procedures.
  + Work with CIRD and UHS to incorporate the UHS NAACCR into i2b2
* Now that the main predictor (Hispanic ethnicity) and outcomes (diagnosis, surgery date, recurrence, and death) are ready, extend my scripts to include covariates reflecting the patients' socioeconomic status, lifestyle, comorbidities, access to care, and stage at diagnosis.
* Determine repeatability using the following independent data sources (Aims 1 and 3):
  + Standard NAACCR Class-A reports from TCR or NAACCR
  + (2019) NAACCR and EMR data from i2b2 at Massachusetts General
* Prepare for collaboration with Massachusetts General by finishing the current version of my DataFinisher software, incorporating some of the generalizable data processing tasks I have identified since the start of this project.

I have a detailed progress report for this project at <https://rpubs.com/bokov/kidneycancer> and I update it on a regular basis. I send out a PDF version of it to my mentors whenever I reach a major milestone or roadblock and I aim to complete the next such mailing before the end of this week. I also have a change-log that I typically update several times a day at <https://github.com/bokov/kl2_kc_analysis/commits/f_content_00> (with the caveat that these are short notes for my own use and will not necessarily make sense to others).

### Timeline/Progress

| **Activity Underway and Planned** | **Sep** | **Oct** | **Nov** | **Dec** | **Jan** | **Feb** | **Mar** | **Apr** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Main variables selected and validated. Scripts for preparing them are complete (Aims 2&3) | **✔** |  |  |  |  |  |  |  |
| Update i2b2 query to also include renal mass EMR diagnosis codes and all NAACCR site codes. | **☐** | **☐** |  |  |  |  |  |  |
| Decide which class to take next semester (instead of the huge list originally proposed) |  | **☐** |  |  |  |  |  |  |
| **Revise KL2 research budget to include tuition. Insure that remaining personnel portion not committed to specific department.** |  | **☐** |  |  |  |  |  |  |
| **Ascertain whether salary calculation should be YZ or just Y and adjust KL2 salary coverage downward if it is the latter.** |  | **☐** |  |  |  |  |  |  |
| Figure out what arrangements I need to make for annual CTSA meeting. |  | **☐** |  |  |  |  |  |  |
| Look up deadlines and guidelines for CTSA/IIMS presentations. |  | **☐** |  |  |  |  |  |  |
| Follow up on Entering Mentoring, find out when next one is offered. |  | **☐** |  |  |  |  |  |  |
| Find out and register for next SOM leadership course. |  | **☐** |  |  |  |  |  |  |
| Edit [Progress Report](https://rpubs.com/bokov/kidneycancer) for readability and update the figures and tables, submit to mentors as first milestone. | **☐** | **☐** |  |  |  |  |  |  |
| Learn how stage at diagnosis is encoded in consultation with Dr. Rodriguez and other experts in Urology, write code for their analysis (Aims 2&3) |  | **☐** | **☐** |  |  |  |  |  |
| Review of NAACCR abstraction process with Mays Cancer Center Registrar (Aim 2) |  | **☐** |  |  |  |  |  |  |
| Update scripts to incorporate what is learned from NAACCR registrar |  |  | **☐** | **☐** |  |  |  |  |
| Consult with IRB about whether an additional protocol is needed beyond HSC20180345H, HSC20170563N, and HSC20150212HR |  | **☐** | **☐** |  |  |  |  |  |
| Prepare presentation/s for AMIA 2019 |  |  |  | **☐** | **☐** | **☐** | **☐** |  |
| Submit credentialing materials for Sunrise and Epic | **☐** | **☐** | **☐** |  |  |  |  |  |
| Submit request for Texas NAACCR data with the required variables I have selected. |  | **☐** | **☐** |  |  |  |  |  |
| Inquire into getting the UHS NAACCR incorporated into our i2b2 |  |  | **☐** | **☐** | **☐** | **☐** |  |  |
| Prepare and test covariates from local data (Aims 2&3) | **☐** | **☐** | **☐** | **☐** | **☐** |  |  |  |
| Make prioritized list of cases for manual chart review (Aims 2&3) |  |  |  |  | **☐** |  |  |  |
| Complete pending changes to DataFinisher including the additional transformations that I have identified in the course of my work since August (Aim 1) |  |  |  | **☐** | **☐** | **☐** | **☐** |  |
| Clean up and document DataFinisher code in preparation for presenting it to Harvard (Aim 1) |  |  |  | **☐** | **☐** | **☐** | **☐** |  |
| Make arrangements for externship at Harvard |  |  | **☐** |  |  |  |  |  |
| Update input data after next i2b2 refresh (Aims 2&3) |  |  |  |  | **☐** | **☐** |  |  |
| Updates to code for improved reliability, maintainability, and efficiency. | **✔** | **☐** | **☐** | **☐** | **☐** | **☐** | **☐** | **☐** |

## Time Commitments (Current Fiscal Year Overall)

|  |  |
| --- | --- |
| *Research/CD – 75% Effort* | *Teaching/Service – 25% Effort* |
| * KL2 project (60%) * CD (15%)   + SPARK/SOAR: 3%   + Other, including promotion packet: 12% | * IIMS Core B (informatics) 5% * IIMS Core D (workshop) 5% * IIMS Core G (TIC, IDTF) 10% * CTSA ACT 5% |

## Scholarship (since August 2018)

### Submitted (American Medical Informatics Association Summit, 2019):

* An Improved Tool for Making i2b2 Patient Sets Into Spreadsheets of Analyzable Data (Presentation)
* A (Relatively) Painless Workflow for Reproducible, Robust, and Accountable Data Analysis Projects Using Git and R (Presentation)
* Side by Side Comparison of Rockwood Index and Risk Analysis Index as Predictors of Surgical Complications at a Texas Safety Net Hospital Serving a Large Hispanic Population (Poster)

## Grants

### In Preparation

1. Bokov (PI) May 2019

NLM Career Development Award in Biomedical Informatics and Data Science (K01) PAR-16-204

*Leveraging Data Research Networks for Low-Cost High-Impact Multi-Site Studies*

Role: Principal investigator

### Current

1. Bokov (PI) 08/01/18 - 08/01/20

SOM Career Development (KL2) Award

*Novel Translational Informatics Software: an Application to Kidney Cancer*

Role: Principal investigator

Total award: $201,200

1. Clark, Shireman, Dougherty, Hargreaves (PI) 05/24/18 - 04/30/23

Clinical and Translational Science Award (UL1)

*Institute for the Integration of Medicine and Science: a Partnership to Improve Health*

Role: Contributor

Total award: $23,118,937

### Completed

1. Ramirez, Tirado-Ramos (PI) 09/01/15 - 09/01/18

Patient Centered Outcomes Research Institute (PCORI)

*Greater Plains Collaborative Clinical Data Research Network*

Role: Co-Investigator

Total award: $8,637,161

1. Musi (PI) 06/01/15 - 04/01/20

NIA/NIH P30

*San Antonio Claude D. Pepper Older Americans Independence Center*

Role: Contributor

Total award: $5,321,630