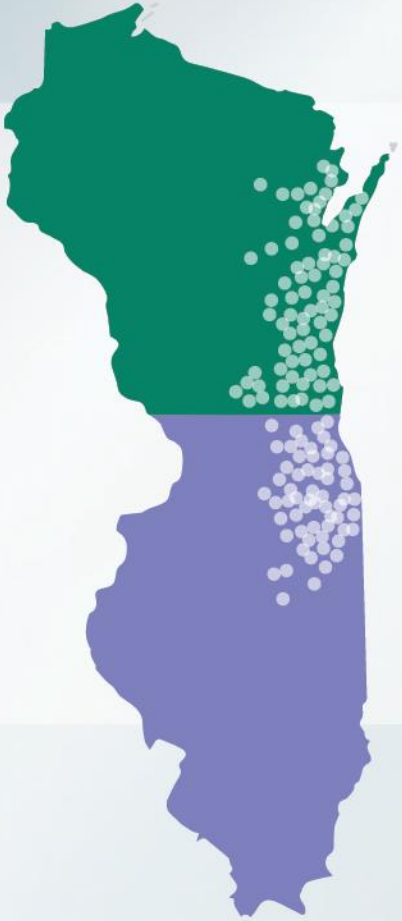


Common Data Model

for harmonizing heterogenous data

7/16/2019 | Joseph M. Plasek, PhD

Advocate Aurora Health Research Institute

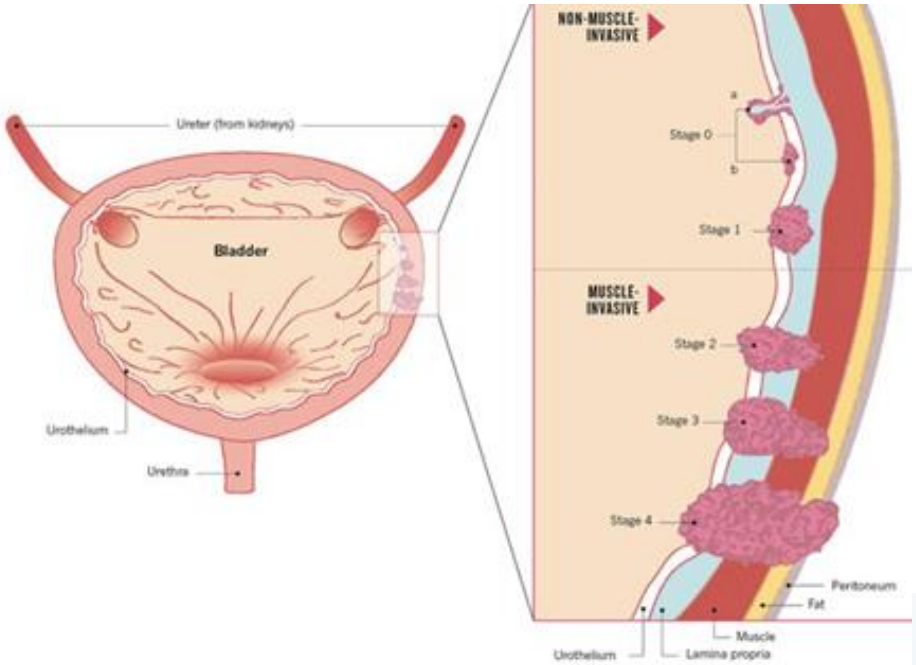
AdvocateAuroraHealth™

Advocate Health Care

Aurora Health Care®

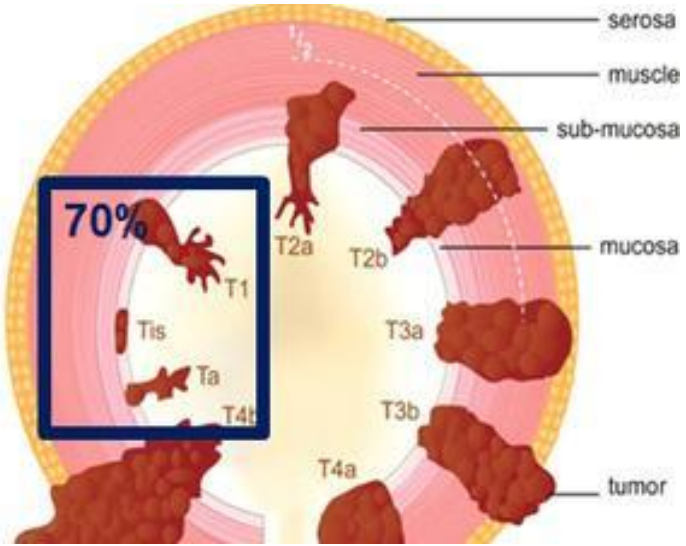
Precision Health Lab

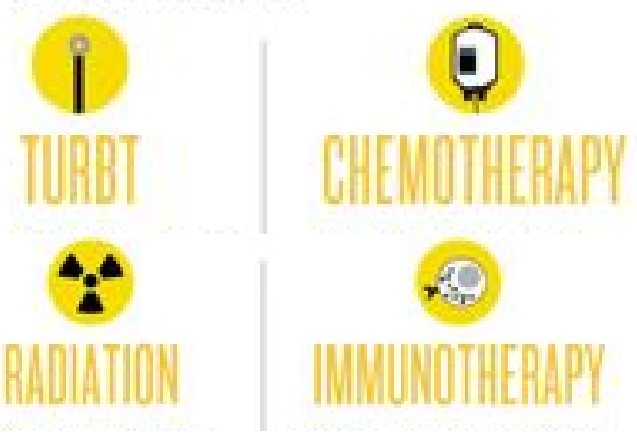
Non-Muscle Invasive Bladder Cancer



Anticoagulation Therapy

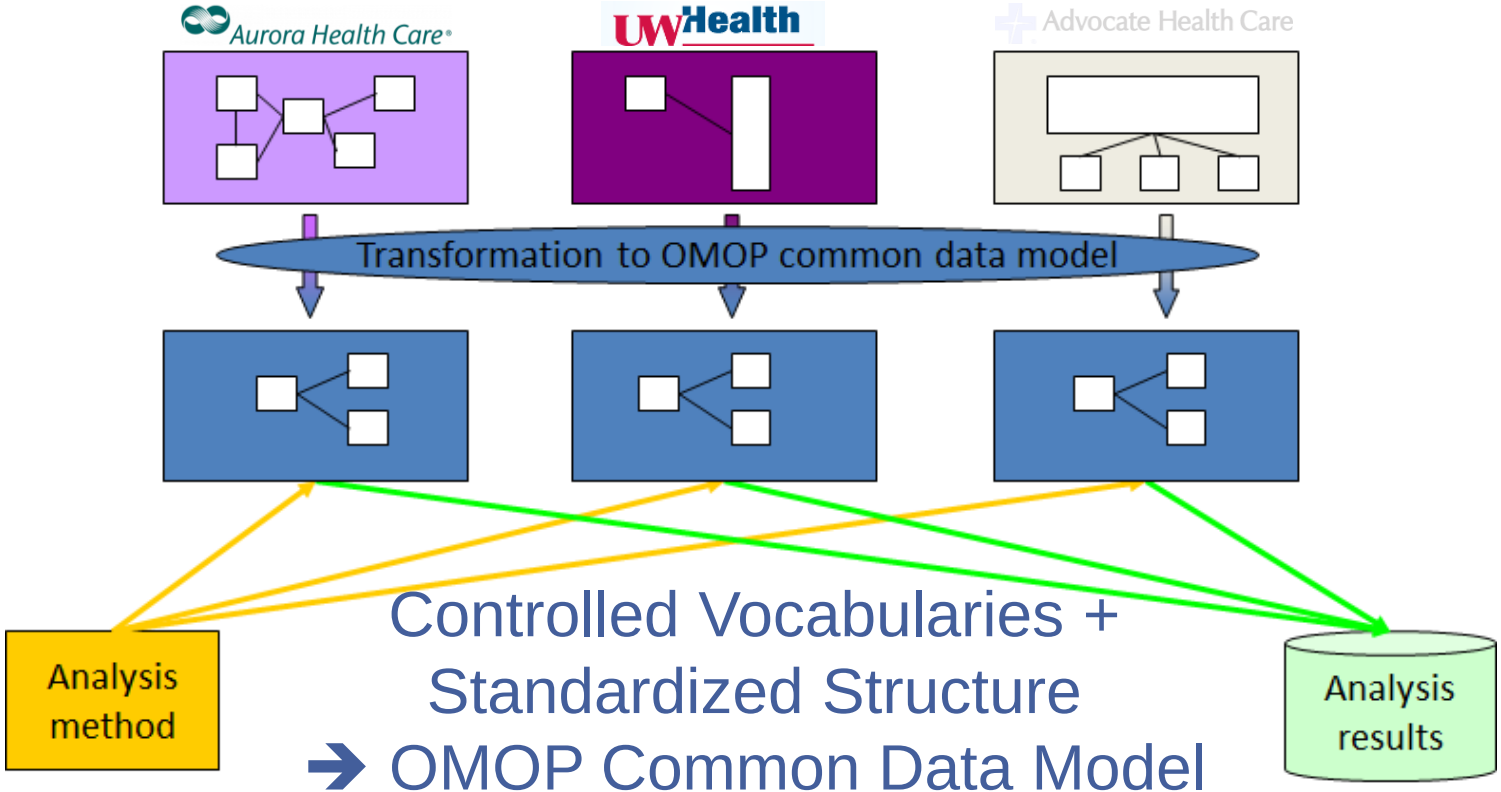
Non-Muscle Invasive Bladder Cancer





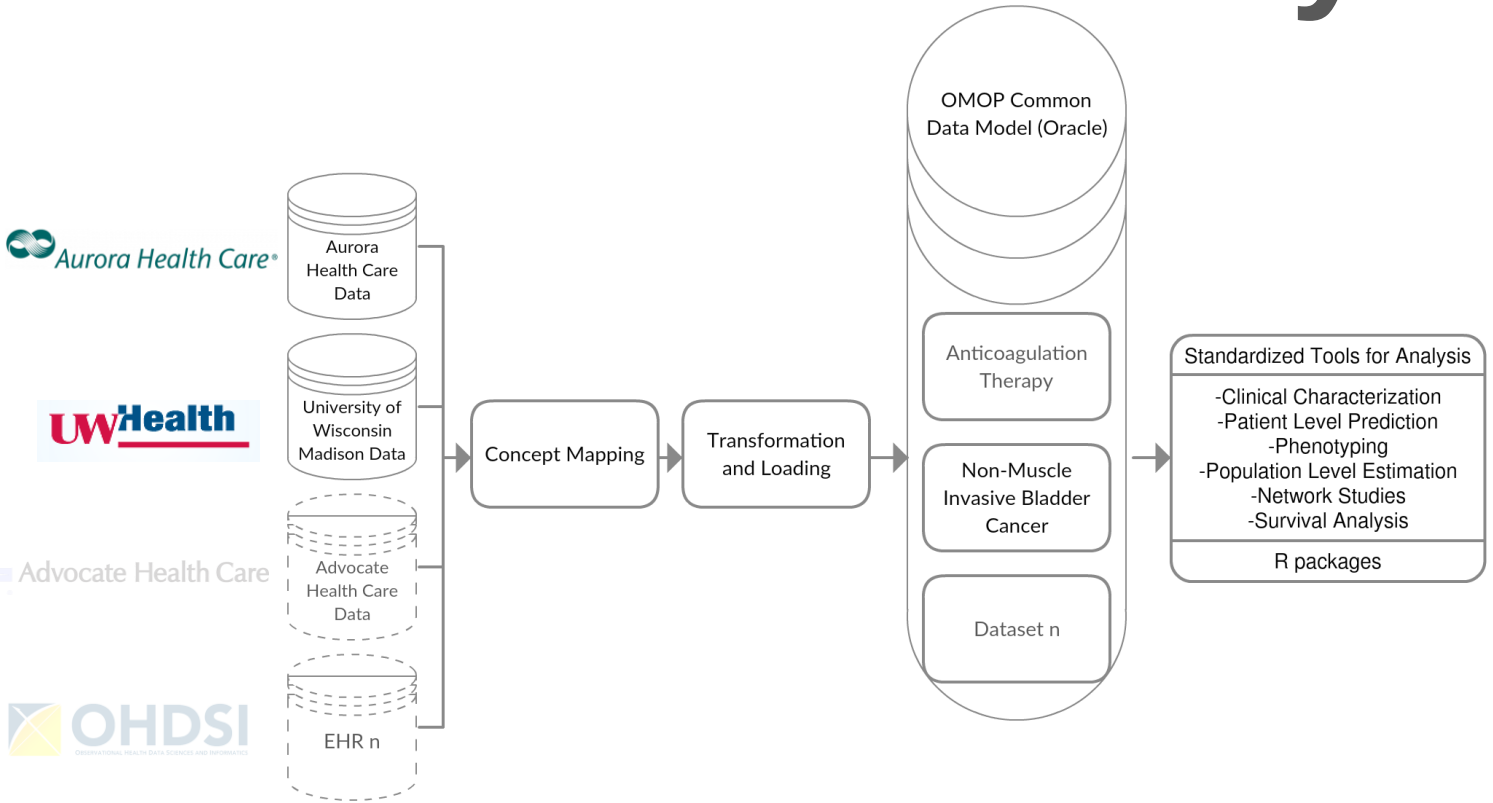
Of 1840 patients, 861 (47%) had a recurrence and 107 (6%) progressed.

Data Harmonization



Controlled Vocabularies +
Standardized Structure
➔ OMOP Common Data Model

Standardized Analytics



Non-Muscle Invasive Bladder Cancer Model

Person

- **person_id**
- gender_concept_id
- year_of_birth
- month_of_birth
- race_concept_id
- ethnicity_concept_id
- location_id
- care_site_id

8507	Male
8532	Female
8657	American Indian or Alaska Native
8515	Asian
38003563	Hispanic
8516	Black or African American
8527	White
38003579	Chinese
38003589	Pakistani
38003574	Asian Indian
8557	Native Hawaiian or Other Pacific Islander
38003564	Not Hispanic or Latino
38003563	Hispanic or Latino

Death

- person_id
- **death_date**

4086160	Insulin used
4148764	Benign
4078513	Number of tumors
4263980	Tumor location after sectioning
45771295	cytology
4029969	papillary urothelial neoplasm
4144272	Never smoked tobacco
37395605	Occasional tobacco smoker
4310250	Ex-smoker
4298794	Smoker
4276526	Cigarette smoker
4246415	Cigar smoker
37017610	User of smokeless tobacco

Condition_occurrence

- person_id
- **condition_concept_id**
- condition_start_date
- condition_end_date
- condition_source_value
- condition_status_source_value
- condition_status_concept_id

197508	Malignant tumor of urinary bladder
201826	Type 2 diabetes
433435	Carcinoma in situ
4246165	Muscularis Propria
4184220	Muscularis propria invaded
4163439	Tumor size finding
4097297	Recurrence
4168352	Tumor progression
36713631	Cystoscopy abnormal
44500580	Papillary carcinoma of urethra
4204361	Tumor grade G1
4041008	Tumor grade G2
4264742	Tumor grade G3
4054711	Tumor grade GX
4196566	Tumor grade G0
4161012	low grade histologic differentiation
4161667	high grade histologic differentiation
4241433	Ta: Noninvasive papillary carcinoma
4052407	Tumor stage T1
4195922	Tis stage
4194614	Tumor stage T2
4282467	Tumor stage TX
4129899	Penile Cancer
44501215	Signet ring cell carcinoma of bladder
201254	Type 1 diabetes
4085878	Recurrent hematuria
4299336	Lymphovascular invasion
36713631	Cystitis

Observation

- person_id
- **observation_concept_id**
- observation_date
- value_as_number
- value_as_string
- observation_source_value

4305077	TURBT - Transurethral resection of bladder tumor
4200033	Instillation of BCG into the bladder
4029571	Cystectomy
4085429	Cystoscopy
937619	Blue light cysto imag agent
4293744	Fulguration
4302745	Bladder irrigation
4022807	Nephroureterectomy
4057754	Palliative course of radiotherapy
4010266	Partial cystectomy
4273866	Radical cystectomy
4002243	Urinary diversion

Measurement

- person_id
- **measurement_concept_id**
- measurement_date
- operator_concept_id
- value_as_number
- unit_concept_id
- measurement_source_value
- unit_source_value
- value_source_value

2212285	Creatinine
4150621	Urine creatinine measurement
2212296	Creatinine clearance
4150342	Urine microalbumin / creatinine ratio
4353852	Urine Protein / creatinine ratio
4155367	Fluid sample creatinine measurement
44806420	Estimation of glomerular filtration rate
3000963	Hemoglobin
2212392	Hemoglobin A1c level
4254663	Lymphocyte count
4194332	Monocyte count
4148615	Neutrophil count
4267147	Platelet count
3045325	Tuberculosis status
4298431	White blood cell count

Procedure_occurrence

- person_id
- **procedure_concept_id**
- procedure_date
- modifier_concept_id
- procedure_source_value
- modifier_source_value

Drug_exposure

- person_id
- **drug_concept_id**
- drug_exposure_start_date
- drug_exposure_end_date
- sig
- route_concept_id
- drug_source_value
- drug_source_concept_id
- route_source_value
- dose_unit_source_value

Location

- **location_id**
- city
- state
- zip
- county

0	University of Wisconsin Madison
1	St. Luke's Medical Center
2	West Allis Memorial Hospital
3	Aurora Sinai Medical Center
4	Aurora Medical Center -Hartford (WC)
5	Aurora BayCare Medical Center
6	Aurora Medical Center Kenosha
7	Aurora Lakeland Medical Center
8	Aurora Memorial Hospital of Burlington
9	Aurora Medical Center Oshkosh
10	Aurora Medical Center Manitowoc County
11	Aurora Sheboygan Memorial Med Ctr
12	Aurora Medical Center Summit/Wilk
13	Aurora Advanced Health Care
14	Aurora Mid Market Clinics
15	Aurora South Region Clinics
16	Aurora Central Region Clinics
17	Aurora North Region Clinics
18	Aurora Medical Center Grafton

Care_site

- **care_site_id**
- care_site_name
- location_id

42903970	BCG	1750500	clarithromycin
1503297	Metformin	45892419	gentamicin
1389036	Mitomycin	1790868	Amikacin
1560171	Glipizide	902722	Tobramycin
1559684	Glyburide	1784749	Kanamycin
1502826	Nateglinide	915981	neomycin
1525215	Pioglitazone	1738521	doxycycline
1580747	Sitagliptin	1149380	Fluticasone
1314924	Gemcitabine	905233	Mometasone
1380068	Intron A	905151	Alclometasone
19137385	Thiotepa	917205	Desonide
19012543	Valrubicin	42629079	atezolizumab
1397599	Cisplatin	19026972	lenalidomide
1307046	Metoprolol	1336539	sunitinib
978236	DFMO	981691	imiquimod
970250	Spironolactone	1777417	Rifabutin
1317967	Aliskiren	1146810	piroxicam
955632	fluorouracil	1315027	cranberry
1305058	Methotrexate	920458	betamethasone
19008264	vinblastine	998415	clobetasol
1338512	doxorubicin	960988	diflorasone
1338512	Adriamycin	955252	Fluocinonide
1344905	Carboplatin	949759	Halobetasol
1378382	paclitaxel	930747	Amcinonide
1315942	docetaxel	917336	Desoximetasone
45775965	pembrolizumab	960988	Diflorasone
1782521	isoniazid	19011097	propionate
1763204	rifampin	903963	triamcinolone
1749301	ethambutol	996541	fluocinolone
1112807	Aspirin	975125	hydrocortisone
1177480	Ibuprofen	918906	oxybutynin
1115008	Naproxen	1344354	epirubicin
1118084	Celecoxib	923081	ofloxacin
1124300	Diclofenac	1797513	ciprofloxacin
1195492	etodolac	1742253	levofloxacin
1178663	indomethacin	1716903	moxifloxacin
1185922	ketoprofen	1113648	Nabumetone
1136980	ketorolac	987406	Ethacrynic acid
1436678	tamoxifen		

Observation_Period

- person_id
- **observation_period_start_date**
- **observation_period_end_date**

Risk Model Validation

Predictive factor	Recurrence	Progression
#of tumors		
1	0	0
2-7	3	3
8 or more	6	3
Tumor diameter		
<3 cm	0	0
3 or more cm	3	3
Recurrence rate		
First tumor	0	0
1 recurrence per year	2	2
>1 recurrence per year	4	2
Stage		
Ta	0	0
T1	1	4
Presence of CIS		
No	0	0
Yes	1	6
Grade (1973 WHO)		
1	0	0
2	1	0
3	2	5
Total score	0-17	0-23
Low risk	0	0
Intermediate risk	1-9	2-6
High risk	10-17	7-23

Area under the receiver operator curve (AUC) was:

- 59.63% for recurrence
- 74.96% for progression

Challenges applying OMOP CDM

- Challenges identified revolved around the generalizability of OMOP CDM to represent oncology specific attributes, complex drug regimens, and time to outcome.

Tumor Characteristics Gap

- The current structure for storing tumor characteristics disperses them across multiple tables (e.g., condition occurrence, measurement, observation) without sufficient detail and linkages.
- Temporary Solution: We adapted existing concepts for tumor specification.

Tumor Characteristics Gap

- Recent Oncology Working Group Accepted Proposal: Store tumor characteristics in an extension to the Measurement table and connect everything together with a new Episode table for disease abstractions. Extend vocabulary support for diagnostic modifiers using NAACCR and Nebraska Lexicon.

Tumor Characteristics Gap

- Remaining Challenges of Accepted Proposal: Unfortunately, non-muscle invasive bladder cancer is not currently supported in the chosen vocabularies, thus a further extension of these vocabularies is needed.

Complex Drug Regimens Gap

- Gap: OMOP CDM lacked structure to express key drug regimen elements (e.g., induction vs maintenance, cycle number, protocol) such as the immunotherapy regimen for Bacillus Calmette-Guerin that are predictive of outcomes.

Complex Drug Regimens Gap

- Temporary Solution: We repurposed partially related unused fields (refills, quantity) for immunotherapy regimens.

Complex Drug Regimens Gap

- Proposed Solution: Add additional columns (cycle type, cycle number, protocol) to the drug_exposure table so that you can know if the patient is in compliance with the proposed protocol or has discontinued planned treatment

Time to Outcome Gap

- Gap: OMOP CDM does not allow for an explicit representation of time between recurrences for an individual patient.
- Temporary Solution: We created temporary categorical representations of otherwise continuous data.

Time to Outcome Gap

- Stepwise metrics:

  Recurrence < 1/year = 2 points in existing risk model

  Recurrence >= 1/year = 4 points in existing risk model

- OHDSI tools should be to be able to auto-calculate the different ranges and assign points differentially depending upon duration and utilize these as features in risk prediction algorithms

Time to Outcome Gap

- Proposed Solution: As temporally dynamic factors can be derived from the data, enable their use through appropriate expansion of the Feature Selection package. Two types of temporally derived variables would be most beneficial: 1) each step has a range and constant weight; and 2) derivatives that capture change over time

Infrastructure Opportunities

- Gap: Relational databases are not optimal for knowledge representation tasks such as named entity recognition, which is one of the main use cases of OMOP CDM
- Temporary Solution: We piloted OMOP CDM in PostgreSQL and Oracle relational databases

Infrastructure Opportunities

- Proposed Solution: Implement OMOP CDM in a graph database. This would be a major infrastructure change, but it is necessary change to create efficiencies in computation, especially as most of the tasks where the OMOP CDM is useful can be done using graph based algorithms.

Thank You

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

For questions, please contact Kourosh Ravvaz (Kourosh.Ravvaz@aurora.org) at the Advocate Aurora Health Research Institute