



DEPARTMENT OF
UROLOGY

VANDERBILT  UNIVERSITY
MEDICAL CENTER

Feasibility and validation of large-scale data acquisition from the electronic health record to a secure research database for nephrolithiasis

Presenter: Wilson Sui, MD

Disclosures

- none

Introduction: Big data









The six Vs of big data

Big data is a collection of data from various sources, often characterized by what's become known as the 3Vs: *volume*, *variety* and *velocity*. Over time, other Vs have been added to descriptions of big data:

Pros....

and Cons.

VOLUME	VARIETY	VELOCITY	VERACITY	VALUE	VARIABILITY
The amount of data from myriad sources.	The types of data: structured, semi-structured, unstructured.	The speed at which big data is generated.	The degree to which big data can be trusted.	The business value of the data collected.	The ways in which the big data can be used and formatted.
					

Steps in the typical chart review project

- 1) Identify patients
- 2) Variables
- 3) Extract each datapoint manually
- 4) Log it somewhere
- 5) Database maintenance

Current state of chart review

B	C	D	E	F	G	H	I
DOB	Sex	HTN	CHF	T2DM	Stone disease?	Location stone	Treatment
11/29/1978	M	0	1	1	1	RENAL PELVIS	PCNL
8/29/1999	F	1	0	0	0	0	0
5/4/1940	F	1	1	1	1	URETER	ESWL

Time per patient = number of variables * complexity

Current state of chart review

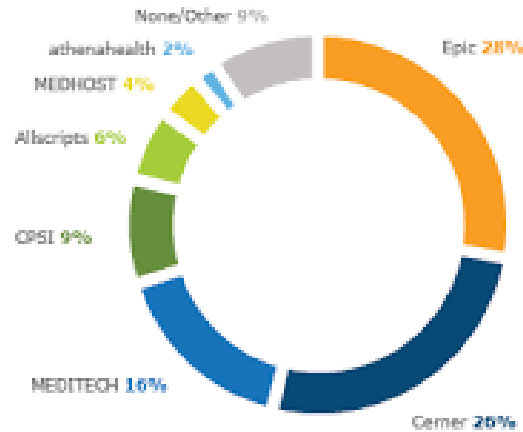
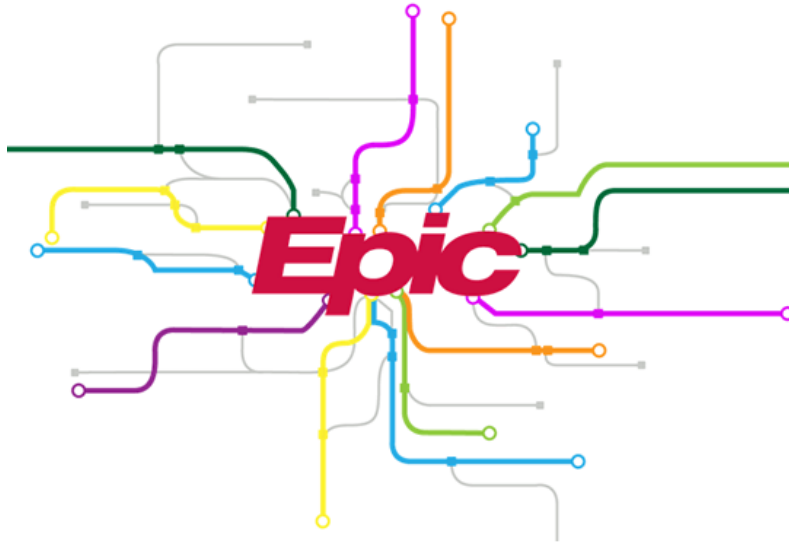
Chief Complaint/History of Present Illness: ^{Ca 9.4}
CT 11/08 ^{Ca 1.1}
RNP 7x6x4mm ^{3.4}
RNP 4mm ^{5.1}
KUB ^{Hydronephrosis}
Pineal calc
8-9mm calc
various lateral calc
CT 7/09 4mm UVS
stone
Impression: Nephrolithiasis
- Not interested in URS/stent
PMHx
? palpitations

Hx Nephrolithiasis - since 2003
- URS - stent
2nd episode was Aug 2009 - 1-2 days
pain. No stone collected
of pain now of UTI

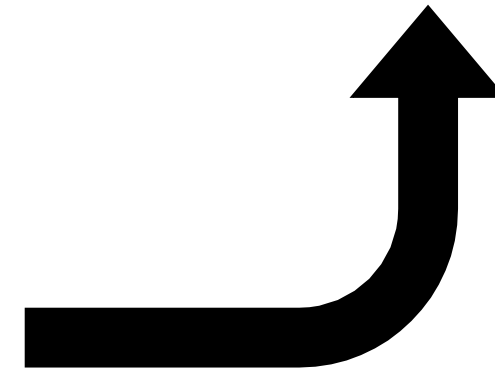
The EPIC link: CDP & CDM



2018 US Acute Care
Hospital Market Share
(n=5,447 acute care hospitals)



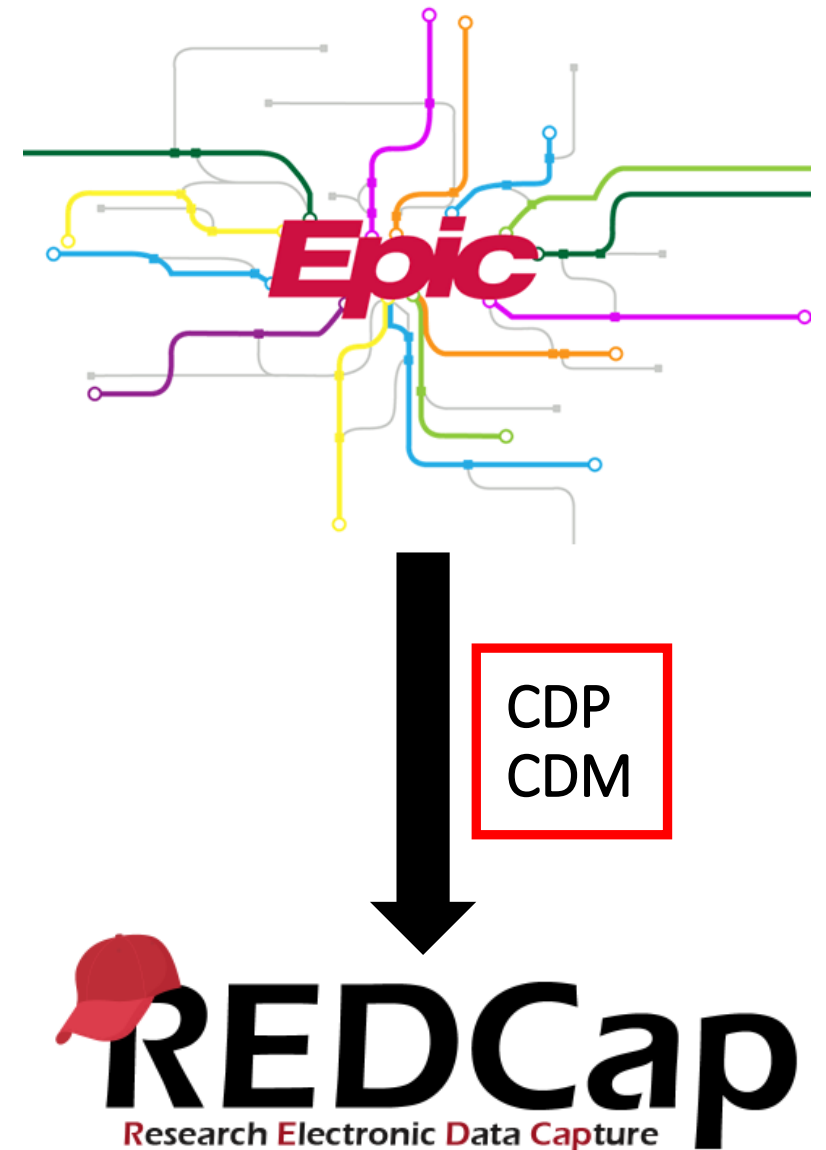
Clinical Data Pull
Clinical Data
mart



What is CDP & CDM?

Feature to fetch data from EHR

- Manual or automatic
- Able to pull temporal data
- Data fields:
 - Demographics (race, gender, ethnicity, DOB, address, phone)
 - Medications (active, completed, on-hold, stopped)
 - Allergies
 - Vitals
 - Labs
 - Problem list



What have we done so far?

LARGE-SCALE DATA ACQUISITION FROM THE ELECTRONIC HEALTH RECORD TO A SECURE RESEARCH DATABASE FOR NEPHROLITHIASIS: VALIDATION AND CLINICAL APPLICATION

ASSOCIATION OF CHRONIC KIDNEY DISEASE STAGE WITH 24-HOUR URINE VALUES AMONG PATIENTS WITH NEPHROLITHIASIS

URINARY CITRATE WASTING ASSOCIATES WITH OBESITY AND DIABETES MELLITUS AMONG NEPHROLITHIASIS PATIENTS

MACHINE LEARNING MODELS TO PREDICT KIDNEY STONE COMPOSITION AND 24-HOUR URINE ABNORMALITIES BASED ON ELECTRONIC HEALTH RECORD-DERIVED FEATURES

Validation

- PPV: 92-100%
- NPV: 89-100%
- Average time for manual collection of 5 meds & 5 comorbidities: 29.2 +/- 12.7 seconds

ASSOCIATION OF CHRONIC KIDNEY DISEASE STAGE WITH 24-HOUR URINE VALUES AMONG PATIENTS WITH NEPHROLITHIASIS

Table 1. Demographic and clinical characteristics of the study cohort

	Stage I (eGFR ≥90) n = 184 (%)	Stage II (eGFR ≥60-89.9) n = 1537	Stage IIIa (eGFR 45-59.9) n = 245	Stage IIIb (eGFR 30-44.9) n = 70	Stage IV (eGFR 15-29.9) n = 17	Stage V (eGFR < 15) n = 4	p-value
Age (mean ± SD)	44.0 ± 15.1	49.6 ± 14.5	59.8 ± 12.1	63.0 ± 11.2	59.9 ± 14.8	63.0 ± 17.0	<0.001
BMI (mean ± SD)	22.9 ± 28.6	26.2 ± 25.2	26.4 ± 23.9	26.7 ± 22.9	30.2 ± 7.4	31.1 ± 4.8	0.607
Gender							
Male	85 (46.2)	764 (49.7)	133 (54.3)	33 (47.1)	7 (41.2)	2 (50)	0.598
Female	99 (53.8)	773 (50.3)	112 (45.7)	37 (52.9)	10 (58.8)	2 (50)	
Race							
White	163 (91.1)	1381 (93.4)	223 (93.7)	67 (97.1)	17 (100)	3 (75)	0.26
Non-white	16 (8.9)	98 (6.6)	15 (6.3)	2 (2.2)	0 (0)	1 (25)	
Past medical history							
Inflammatory bowel disease or Diarrhea	18 (9.8)	127 (8.3)	15 (6.1)	6 (8.6)	2 (11.8)	0 (0)	0.748
Hypertension	72 (39.1)	721 (46.9)	154 (62.9)	49 (70.0)	12 (70.6)	3 (75)	<0.001
Gout	7 (3.8)	41 (2.7)	25 (10.2)	10 (14.3)	1 (5.9)	2 (50)	<0.001
Type 2 diabetes mellitus	23 (12.5)	291 (18.9)	75 (30.6)	26 (37.1)	5 (29.4)	1 (25)	<0.001
Osteoporosis / immobility / hyperparathyroidism	8 (4.3)	76 (5.0)	16 (6.5)	6 (8.6)	3 (17.6)	0 (0)	0.138
Coronary artery disease / myocardial infarction	10 (5.4)	134 (8.7)	31 (12.7)	12 (17.1)	4 (23.5)	1 (25)	0.004
Cerebrovascular accident	7 (3.8)	21 (1.4)	11 (4.5)	7 (10.0)	0 (0)	0 (0)	<0.001
Hyperlipidemia	32 (17.4)	423 (27.5)	88 (35.9)	30 (42.9)	7 (41.2)	2 (50)	<0.001
Gastroesophageal reflux disease	50 (27.2)	532 (34.6)	98 (40.0)	30 (42.9)	6 (35.3)	3 (75)	0.029
Epilepsy/migraine	9 (4.9)	57 (3.7)	5 (2.0)	7 (10.0)	1 (5.9)	0 (0)	0.067
Medications							
Potassium Citrate	13 (7.1)	100 (6.5)	22 (9.0)	4 (5.7)	0 (0)	0 (0)	0.587
Allopurinol	4 (2.2)	45 (2.9)	17 (6.9)	3 (4.3)	0 (0)	1 (25)	0.003
Thiazide (HCTZ, indapamide, chlorthalidone)	25 (13.6)	197 (12.8)	28 (11.4)	4 (5.7)	0 (0)	0 (0)	0.255
Stone comp							0.004
Calcium oxalate monohydrate	67 (54.0)	494 (56.5)	95 (60.5)	26 (52.0)	7 (63.6)	0 (0)	
Calcium oxalate dihydrate	18 (14.5)	117 (13.4)	11 (7.0)	2 (4.0)	1 (9.1)	0 (0)	
Hydroxyapatite	28 (22.6)	159 (18.2)	22 (14.0)	15 (30.0)	0 (0.0)	1 (33.3)	
Uric Acid	3 (2.4)	50 (5.7)	21 (13.4)	5 (10.0)	2 (18.2)	0 (0)	
Other	8 (6.5)	55 (6.3)	8 (5.1)	2 (4.0)	1 (9.1)	0 (0)	

ASSOCIATION OF CHRONIC KIDNEY DISEASE STAGE WITH 24-HOUR URINE VALUES AMONG PATIENTS WITH NEPHROLITHIASIS

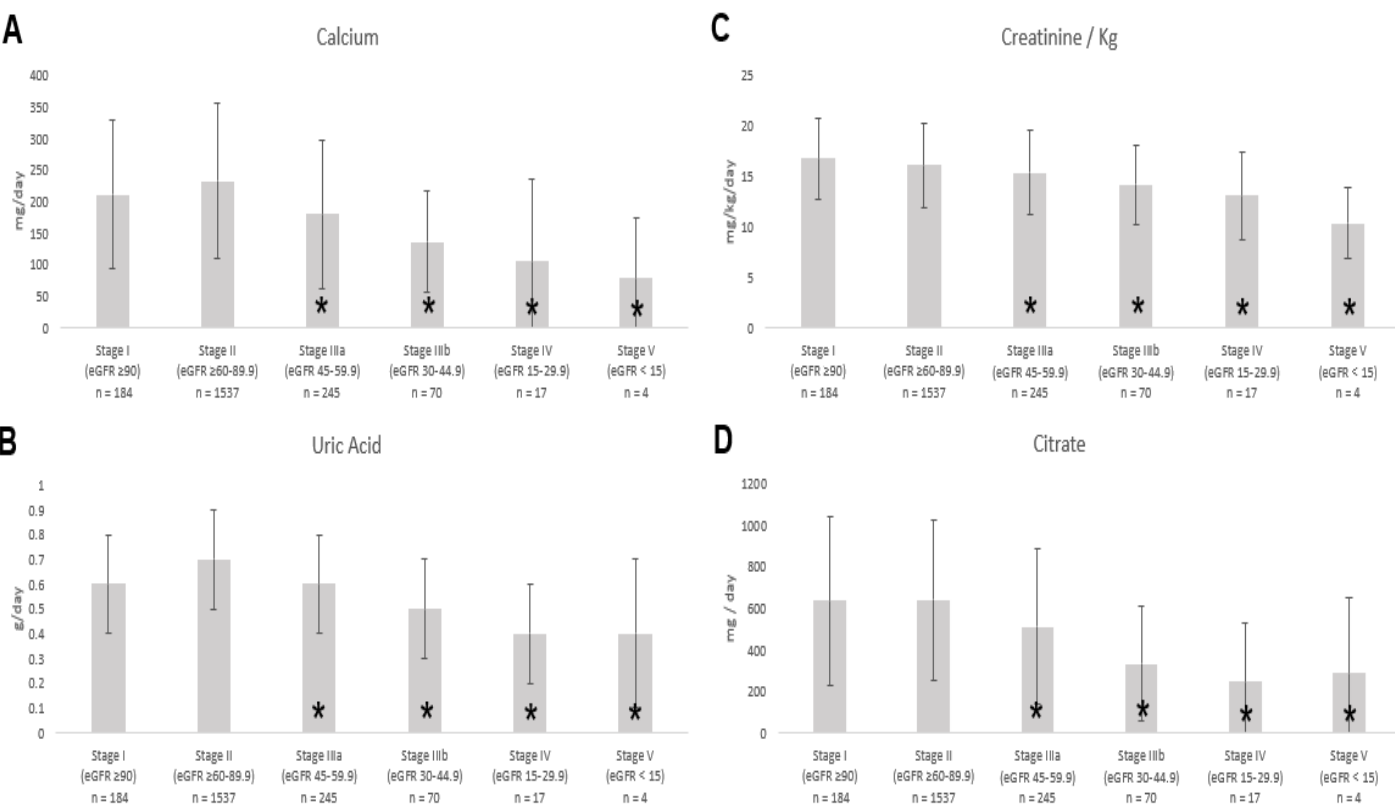


Table 3. Multivariable linear regression of 24hour urine analytes controlling for demographic and clinical factors

all units are per day	$\beta \pm \text{S.E.}$	95% C.I.	p-value
Volume (L)	0.160 \pm 0.732	-1.277 - 1.597	0.827
Calcium (mg)	0.013 \pm 0.005	0.003 - 0.022	0.008
Oxalate (mg)	0.127 \pm 0.038	0.053 - 0.202	0.001
Citrate (mg)	0.004 \pm 0.002	0.001 - 0.007	0.019
pH	0.968 \pm 1.152	-1.295 - 3.231	0.401
Uric acid (g)	6.631 \pm 2.463	1.795 - 11.467	0.007
Sodium (mmol)	0.012 \pm 0.007	-0.002 - 0.026	0.092
Potassium (mmol)	0.031 \pm 0.026	0.021 - 0.083	0.247
Magnesium (mg)	0.022 \pm 0.014	-0.005 0 0.049	0.11
Phosphorus (g)	3.248 \pm 1.576	0.154 - 6.343	0.04
Ammonium (mmol)	0.105 \pm 0.025	0.055 - 1.55	<0.001
Chloride (mmol)	0.014 \pm 0.008	-0.001 - 0.029	0.062
Sulfate (meq)	0.035 \pm 0.036	-0.036 - 0.106	0.333
Creatinine (mg)	-0.004 \pm 0.001	-0.006 - -0.001	0.003
Creatinine per Kg	0.204 \pm 0.144	-0.487 - 0.078	0.155
SSCaOx	0.646 \pm 0.161	0.330 - 0.964	<0.001
SSUA	-0.474 \pm 0.727	-1.901 - 0.952	0.514
SSCaP	1.415 \pm 0.622	0.195 - 2.635	0.023

Models adjusted for age, BMI, sex, hypertension, coronary artery disease, type 2 diabetes, hyperlipidemia, gastroesophageal reflux disease

Downsides

- TOO much data
- Time limit on lab value import
- Problem list is static
- Relies on coding data
- Availability is limited across institutions

Widely applicable

- Real-time updating of lab values (ex: PSA, tumor markers) and comorbidities
- Actively being updated by dedicated staff (CPT codes, radiology?)
- Multi-institutional datasets?

Thank you!

Dr. Hsi
Dr. Penson & VUMC Urology
The VUMC RedCap team