

Exhaustively Characterizing a Patient Cohort by Prevalence of EMR Facts: a Generalized, Vendor-Agnostic Method for Quality Control and Research

Alex F. Bokov, Ph.D. ¹, Gail P. Olin, M.S. R.N.²,
Angela Bos ¹, Alfredo Tirado-Ramos, Ph.D. ¹,
Pamela Kittle, R.N. ¹, Carlayne Jackson, M.D. ¹

¹University of Texas Health Science Center at San Antonio, San Antonio, TX

²University of Texas Health Science Center at Houston, Houston, TX

Key Steps in Analyzing Data from an EMR Research Data Warehouse

1. Convert verbal/conceptual cohort inclusion/exclusion criteria into an unambiguous computer executable query (e.g. i2b2) that relies only on data that is available.
2. Translate predictors, covariates, and response variables also into a computer executable query. Adjust plan for feasibility and interpretability.
3. When the two above queries are finalized, use them to pull visit-level data (for self-service systems such as at our institution) or to create a data request for the local informatics team.
4. Proceed with analysis plan-- development and validation subsets, model specification, variable selection, final analysis re-run on holdout data, reporting of results.

i2b2: Integrating Informatics from Bench to Bedside

i2b2 Query & Analysis Tool Project: CIRD User: AlexBokov Find Patients | Analysis Tools | Message Log | Help | Change Password | Logout

Navigate Terms

- [Cardiology Lab Results \[7,545 facts; 1,746 patients\]](#)
- [Cohort Manager](#)
- [Demographics \[4,242,691 facts; 383,752 patients\]](#)
- [Diagnoses \[17,772,539 facts; 318,849 patients\]](#)
 - [ICD10 \[16,748,690 facts; 317,133 patients\] - 317133](#)
 - [A00-B99 Certain infectious and parasitic diseases](#)
 - [C00-D49 Neoplasms \(C00-D49\) \[1,115,273 facts; 53,111 patients\]](#)
 - [D50-D89 Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism](#)
 - [E00-E89 Endocrine, nutritional and metabolic diseases](#)
 - [F01-F99 Mental, Behavioral and Neurodevelopmental disorders](#)
 - [G00-G99 Diseases of the nervous system \(G00-G99\)](#)
 - [G00-G09 Inflammatory diseases of the central nervous system](#)
 - [G10-G14 Systemic atrophies primarily affecting the nervous system](#)
 - [G10 Huntington's disease \[674 facts; 44 patients\]](#)
 - [G11 Hereditary ataxia \[2,516 facts; 219 patients\]](#)
 - [G12 Spinal muscular atrophy and related syndromes \[24,906 facts; 2,516 patients\]](#)
 - [G12.0 Infantile spinal muscular atrophy, type I](#)
 - [G12.1 Other inherited spinal muscular atrophy](#)
 - [G12.2 Motor neuron disease \[24,906 facts; 2,516 patients\]](#)
 - [G12.20 Motor neuron disease, unspecified](#)
 - [G12.21 Amyotrophic lateral sclerosis](#)

Query Tool

Query Name:

Temporal Constraint: Treat all groups independently

Group 1	Group 2	Group 3
Dates <input type="checkbox"/> Exclude <input type="checkbox"/>	Dates <input type="checkbox"/> Exclude <input type="checkbox"/>	Dates <input type="checkbox"/> Exclude <input type="checkbox"/>
Treat Independently	Treat Independently	Treat Independently

drop a term on here

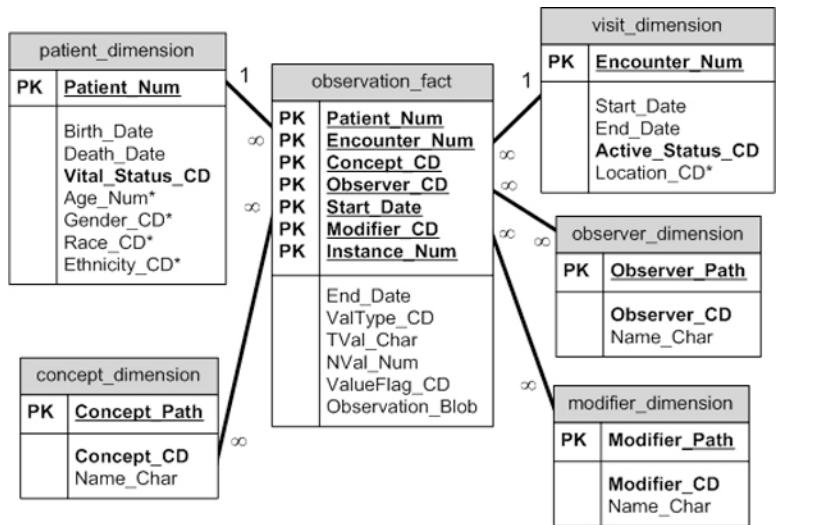
Run Query **Clear** **Print Query**

Query Status

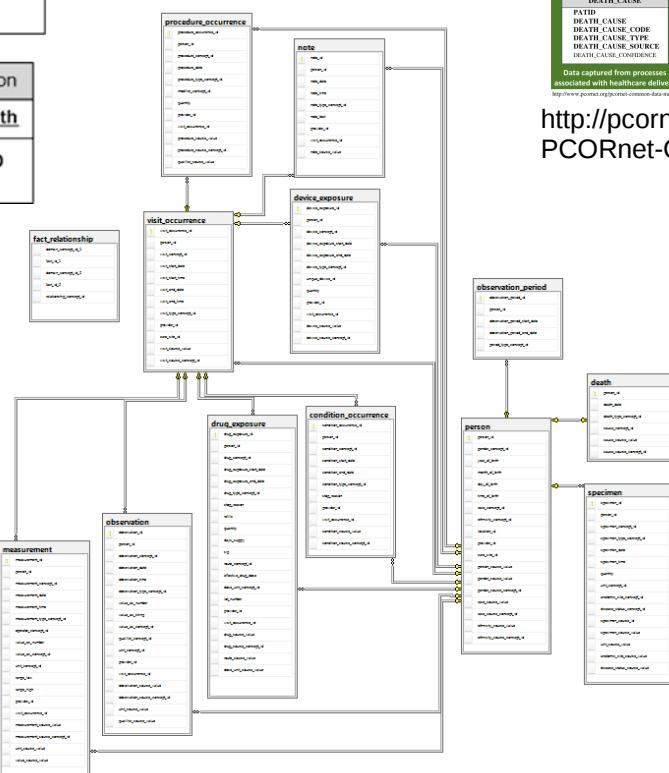


UT Health
San Antonio

The Star Schema and Alternatives



Bellazzi et al. BMC Bioinformatics 2012, 13(Suppl 14):S1
DOI: 10.1186/1471-2105-13-S14-S1



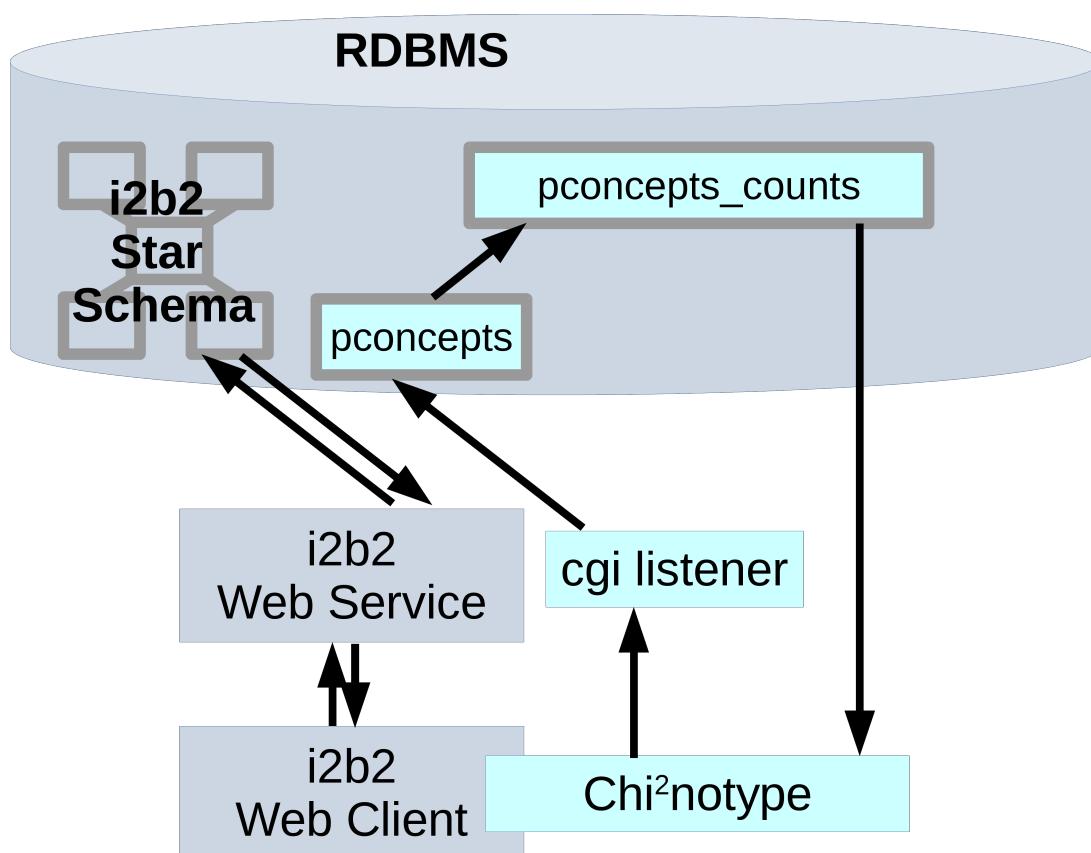
PCORnet Common Data Model v3



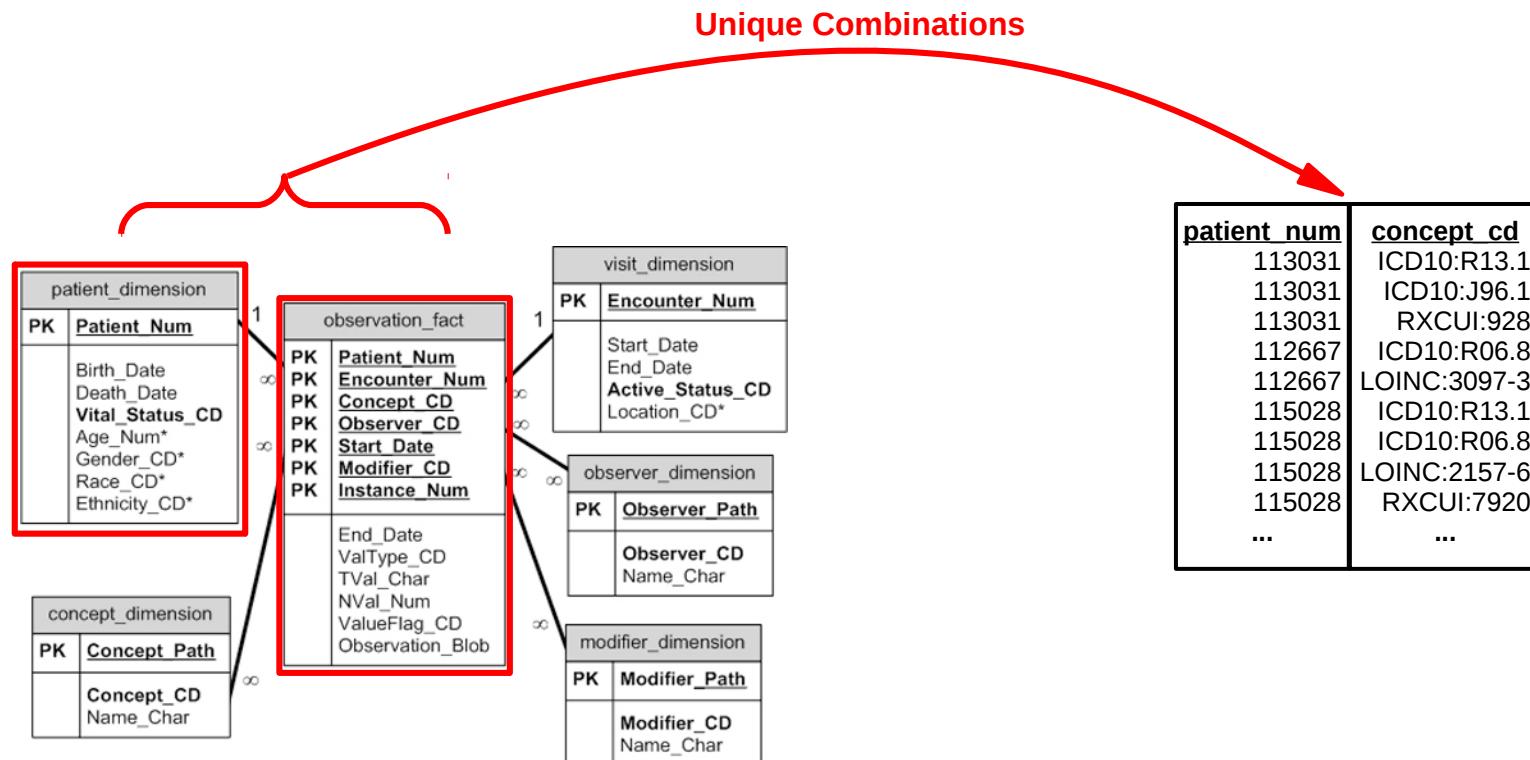
http://pcornet.org/wp-content/uploads/2016/11/2016-11-15-PCORnet-Common-Data-Model-v3.1_Specification.pdf

https://github.com/OHDSI/CommonDataModel/blob/master/OMOP_CDM_v5_2.pdf

How the Chi²notype Plugin Interoperates With i2b2

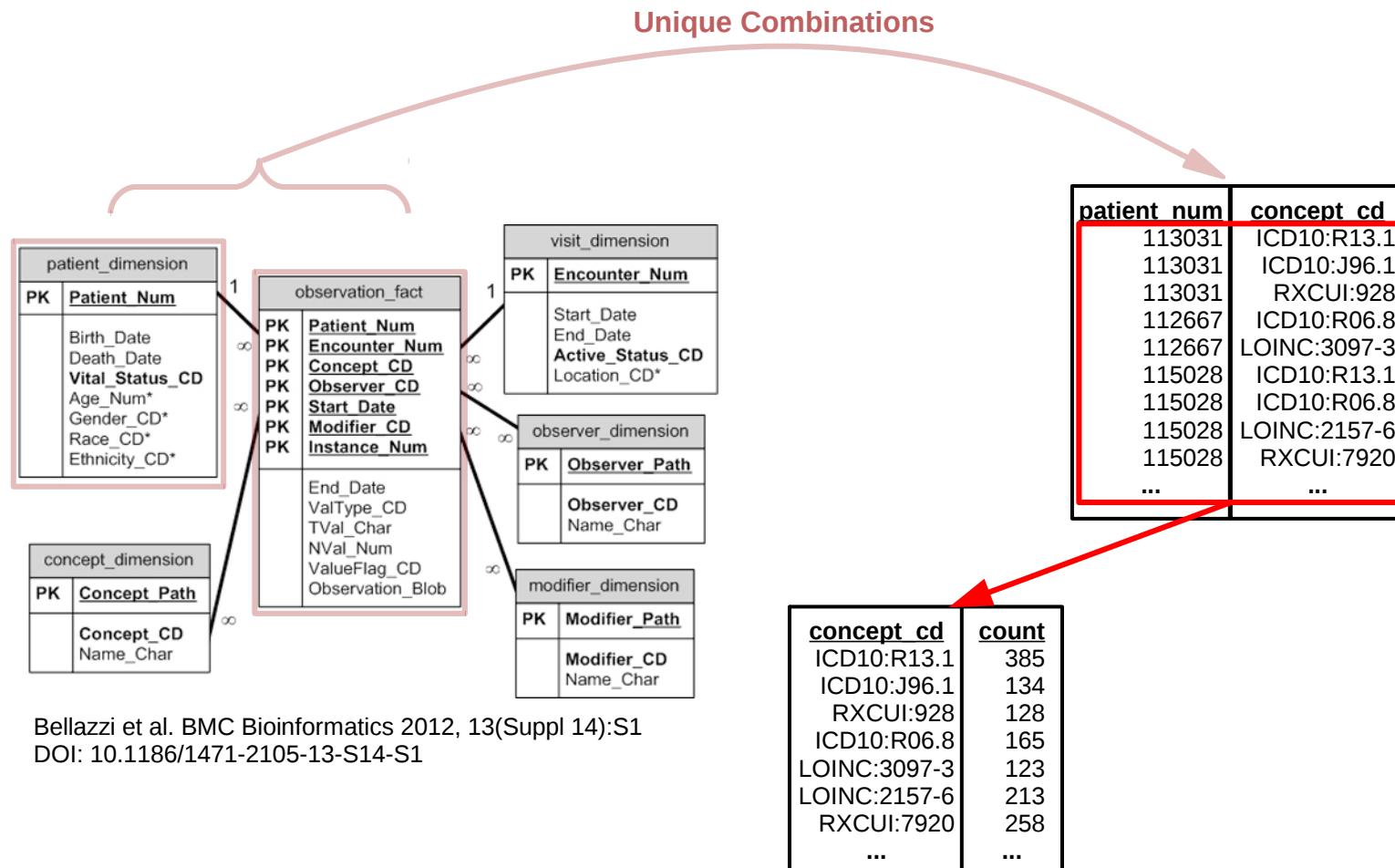


Star Schema to Frequency Tables



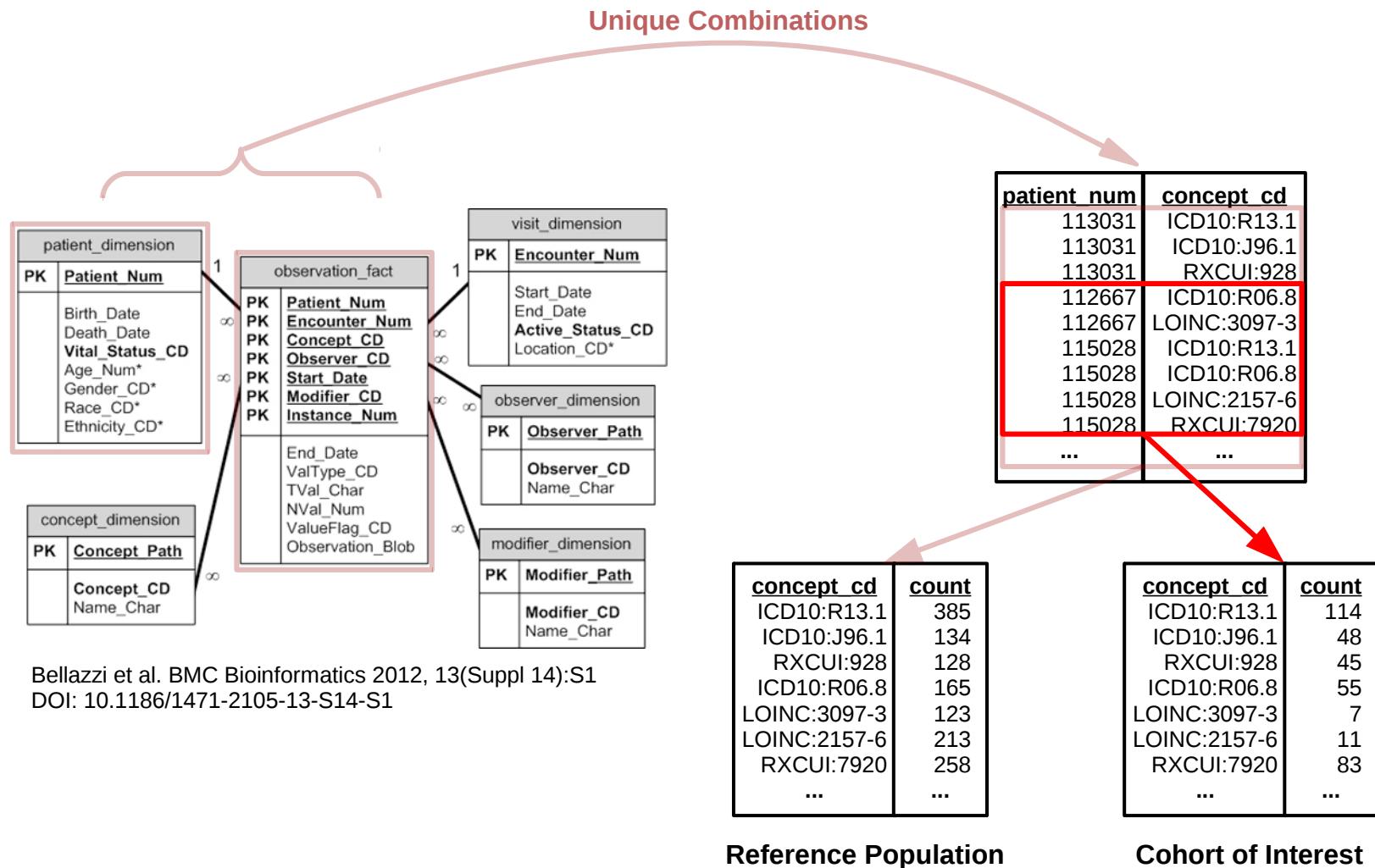
Bellazzi et al. BMC Bioinformatics 2012, 13(Suppl 14):S1
DOI: 10.1186/1471-2105-13-S14-S1

Star Schema to Frequency Tables



Reference Population

Star Schema to Frequency Tables



Bellazzi et al. BMC Bioinformatics 2012, 13(Suppl 14):S1
 DOI: 10.1186/1471-2105-13-S14-S1



Typical Chi²notype Output

i2b2 Query & Analysis Tool Project: CIRD User: Alex Bokov Find Patients | Analysis Tools | Message Log | Help | Change Password | Logout

Chi2

Specify Data View Results Plugin Help

size: cutoff:
20 10 concepts: RXCUI go export

CCD	NAME	REF_ALL	FRC_REF_ALL	ALS	FRC_ALS	CHISQ	DIR
TOTAL		383752		839			
RXCUI:691644	[CN900] CNS MEDICATIONS,OTHER	11871	0.03093	566	0.67461	11237.34	1
RXCUI:7920	[AU350] PARASYMPATHOLYTICS	5994	0.01562	258	0.30751	4576.49	1
RXCUI:692022	[OP600] MYDRIATICS/CYCLOPLEGICS,TOPICAL OPHTHALMIC	1015	0.00264	62	0.07390	1610.45	1
RXCUI:9813	[MS200] SKELETAL MUSCLE RELAXANTS	18990	0.04949	283	0.33731	1404.54	1
RXCUI:691599	[AU000] AUTONOMIC MEDICATIONS	31048	0.08091	347	0.41359	1147.72	1
RXCUI:927	[CN600] ANTIDEPRESSANTS	49236	0.12830	420	0.50060	906.36	1
RXCUI:691769	[CN309] SEDATIVES/HYPNOTICS,OTHER	10376	0.02704	160	0.19070	831.18	1
RXCUI:2225	[CN000] CENTRAL NERVOUS SYSTEM MEDICATIONS	126103	0.32861	733	0.87366	758.52	1
RXCUI:928	[CN601] TRICYCLIC ANTIDEPRESSANTS	7613	0.01984	128	0.15256	745.00	1
RXCUI:590969	[CN609] ANTIDEPRESSANTS,OTHER	45127	0.11759	357	0.42551	676.44	1
RXCUI:11256	[VT600] VITAMIN E	1375	0.00358	48	0.05721	673.43	1
RXCUI:691786	[DE890] DERMATOLOGICALS,SYSTEMIC,OTHER	66	0.00017	10	0.01192	673.16	1
RXCUI:691770	[CN300] SEDATIVES/HYPNOTICS	28952	0.07544	259	0.30870	605.06	1
RXCUI:975	[AP101] ANTIMALARIALS	2635	0.00687	64	0.07628	588.76	1
RXCUI:1540053	[VT000] VITAMINS	44080	0.11487	330	0.39333	566.36	1
RXCUI:998	[AP1001] ANTIPROTOZOALS	2740	0.00716	64	0.07628	559.52	1

The "Chi²" in Chi²notype

$$\sum \frac{(N_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_1 \text{ where } E_{ij} = N_i \frac{N_{\cdot j}}{N} \text{ and...}$$

X_{all} = any set of patients from the EHR system

N = total number of patients in X_{all} .

X_i = a cohort selected from X_{all} such that $X_i \subset X_{all}$.

N_i = total number of patients in X_i .

j = any fact from the EHR

N_{ij} = total number of patients in X_i who have fact j recorded at least once in their record.

$N_{\cdot j}$ = total number of patients in X_{all} who have fact j recorded at least once in their record.

Demographic Summary via Chi²notype

Description	# ALS		% ALS		# ALS	% ALS		
	All	All	+ PBA	+ PBA		All	All	
	# All	% All	# PBA	% PBA		# All	% All	
Ethnicity								
Non-Hispanic	488	58.16%	81	54.00%	Married	503	59.95%	
Hispanic	189	22.53%	43	28.67%	Single	127	15.14%	
Unknown/Other	143	17.04%	25	16.67%	Unknown	118	14.06%	
Not Recorded	12	1.43%	0	0.00%	Divorced	43	5.13%	
Answer Refused	7	0.83%	1	0.67%	Widowed	40	4.77%	
Race								
White or Caucasian	626	74.61%	118	78.67%	Legally Separated	6	0.72%	
African-American	18	2.15%	2	1.33%	Other	2	0.24%	
Asian	10	1.19%	1	0.67%	Distance of Residence from Clinic			
Unknown/Other	148	17.64%	25	16.67%	< 5 mi	94	11.20%	
Not Recorded	24	2.86%	0	0.00%	5-10 mi	110	13.11%	
Answer Refused	13	1.55%	4	2.67%	10-15 mi	132	15.73%	
Language								
English	762	90.82%	139	92.67%	15-20 mi	32	3.81%	
Spanish	27	3.22%	5	3.33%	20-50 mi	109	12.99%	
Unknown	48	5.72%	6	4.00%	50-100 mi	166	19.79%	
Other	2	0.24%	0	0.00%	100-200 mi	107	12.75%	
Sex								
Female	363	43.27%	78	52.00%	State of Residence			
Male	476	56.73%	72	48.00%	Texas (TX)	825	98.33%	
Vital Status								
Living	688	82.00%	114	76.00%	Other	13	1.55%	
Deceased	141	16.81%	36	24.00%				
Death Date Unknown	10	1.19%	0	0.00%				

What is Different Between ALS Patients with PBA and ALS Patients in General?

Description	#	ALS	%	ALS	%		
	All	All	+ PBA	+ PBA	χ^2	OR	
ICD-10 Diagnosis Codes							
R13.1 Dysphagia	385	45.89%	114	76.00%	283.00	3.73	
R53.8 Other malaise and fatigue	463	55.18%	106	70.67%	64.10	1.96	
R45.8 Other symptoms and signs involving emotional state	97	11.56%	97	64.67%	1584.66	14.00	
R47 Speech disturbances, not elsewhere	202	24.08%	73	48.67%	146.28	2.99	
R06.8 Other abnormalities of breathing	165	19.67%	55	36.67%	70.77	2.36	
J96.1 Chronic respiratory failure	134	15.97%	48	32.00%	70.94	2.48	
R49 Voice and resonance disorders	135	16.09%	43	28.67%	42.20	2.10	
F48 Other nonpsychotic mental disorders	37	4.41%	37	24.67%	339.82	7.10	
Medication Classes [VA Ingredient Codes] and Medication Names							
[CN900] CNS MEDICATIONS, OTHER	566	67.46%	144	96.00%	1447.03	11.58	
DEXTROMETHORPHAN-QUINIDINE 20-10 MG PO CAPS	159	18.95%	118	78.67%	1980.64	15.77	
RILUZOLE 50 MG PO TABS	488	58.16%	116	77.33%	122.58	2.45	
[RE301] OPIOID-CONTAINING ANTITUSSIVES/EXPECTORANTS	23	2.74%	16	10.67%	74.89	4.24	
[RE302] NON-OPIOID-CONTAINING ANTITUSSIVES/EXPECTORANTS	146	17.40%	48	32.00%	54.95	2.23	
[OP600] MYDRIATICS/CYCLOPLEGICS, TOPICAL OPHTHALMIC	62	7.39%	26	17.33%	47.89	2.63	
ATROPINE SULFATE 1 % OP SOLN	61	7.27%	26	17.33%	49.78	2.67	
[AU000] AUTONOMIC MEDICATIONS	347	41.36%	91	60.67%	81.67	2.19	
[AU350] PARASYMPATHOLYTICS	258	30.75%	77	51.33%	91.16	2.38	
GLYCOPYRROLATE 1 MG PO TABS	198	23.60%	65	43.33%	89.57	2.48	
<i>Skeletal Muscle Relaxants</i>							
TIZANIDINE HCL 4 MG PO TABS	83	9.89%	31	20.67%	44.15	2.37	
<i>Tube Feeding</i>							
JEVITY 1.5 CAL PO LIQD	223	26.58%	81	54.00%	183.44	3.24	
LOINC Laboratory Results							
[HIGH] BUN/Creat SerPl(3097-3)	43	34.68%	7	26.92%	122.20	0.69	
[HIGH] CK SerPl-cCnc (2157-6)	68	31.92%	11	22.00%	86.78	0.60	

Collaboration Ideas? Questions?

bokov@uthscsa.edu

alex.bokov@gmail.com

Thank You!

- This was developed as part of our work with the Greater Plains Collaborative CDRN, funded by PCORI grants CDRN-1306-04631 and CDRN-1501-26643NIH (Dr. Russ Waitman, PI)
- UT Health San Antonio (Dr. Ron Rodriguez, interim Dean of the Medical School)
- Department of Epidemiology and Biostatistics (Dr. Amelie Ramirez, interim Chair)
- The Institute for the Integration of Medicine and Science (Dr. Robert Clark and Dr. Paula Shireman, PIs)
- Advice from my colleagues Dr. Jon Gelfond and Dr. Chen-Pin Wang
- The i2b2 developer team at Partners Healthcare (Dr. Shawn Murphy, Director of Research Information Systems)
- The HERON developer team at KU Medical Center (Dan Connolly, Lead Developer)