Feasibility of Large-Scale Observational Cancer Research using the OHDSI Network—*Aim 2 Findings*

George Hripcsak, MD, MS

RuiJun Chen, MD

Thomas Falconer, MS

Biomedical Informatics, Columbia University

Medical Informatics Services, NewYork-Presbyterian

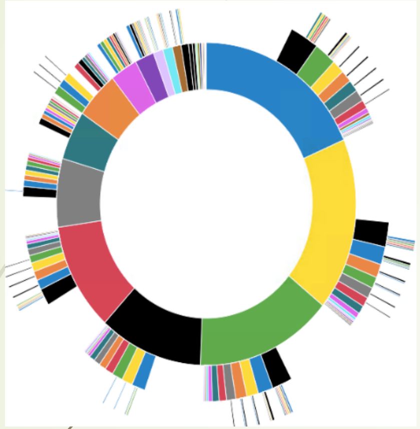
Columbia University Medical Center

NewYork-Presbyterian

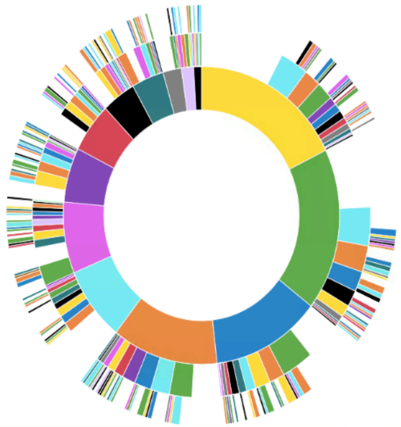
NCI Contract Aims

- Aim 1. Understand the sequence of treatments in cancer patients with diabetes, depression or high blood pressure
  - Presentation and webinar at NCI on 2/14/18
- Aim 2. Understand the feasibility of using existing data infrastructure to conduct cancer treatment and outcomes research.

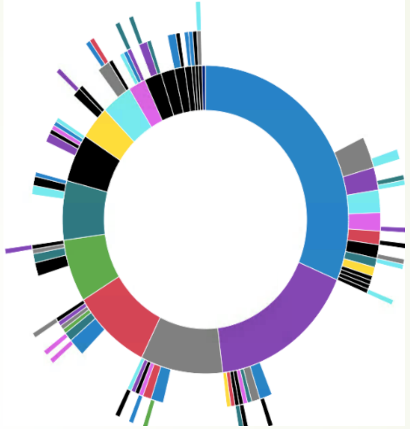
Aim 1 example: Depression treatment pathways in cancer care



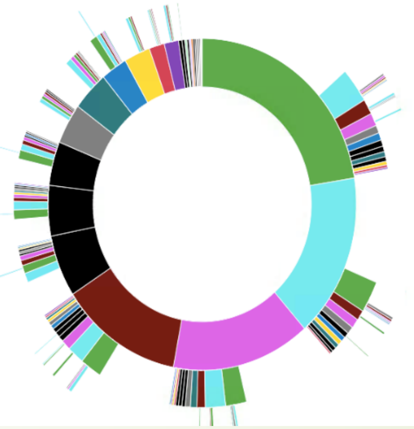
Truven CCAE



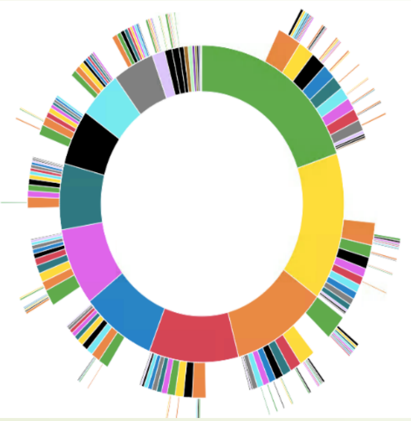
Columbia



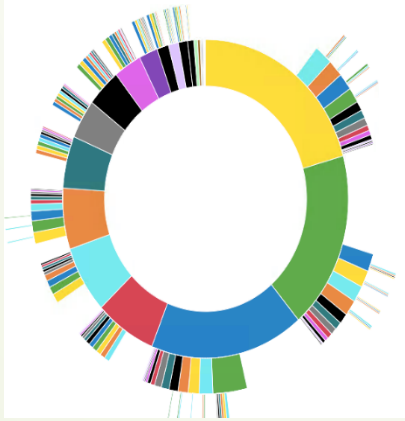
IMS France



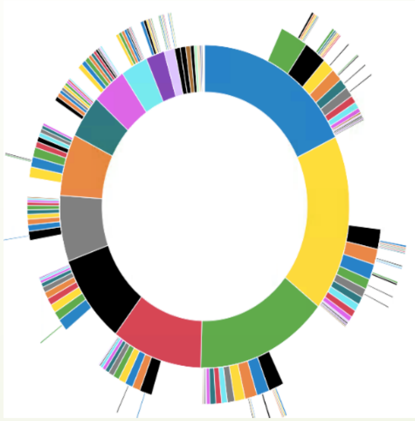
IMS Germany



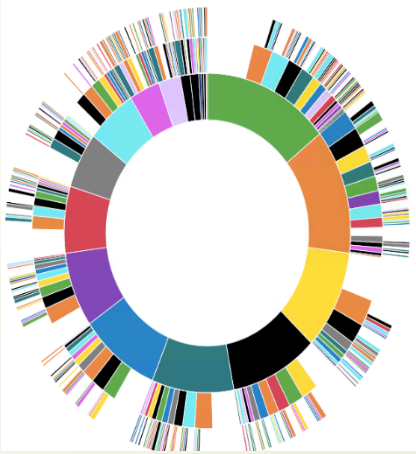
Truven Medicaid



Truven Medicare

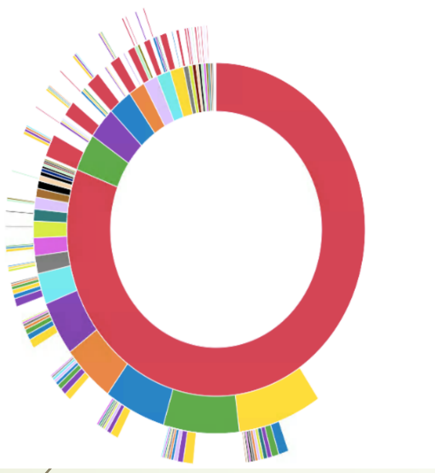


Optum Extended SES

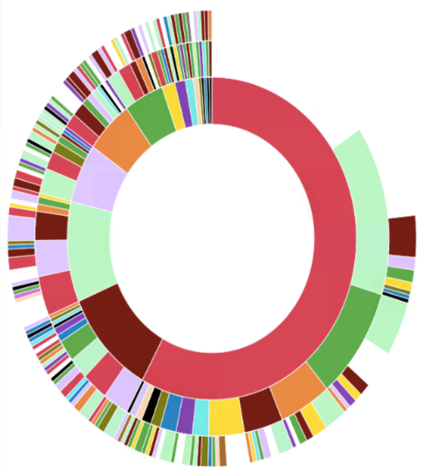


Stanford

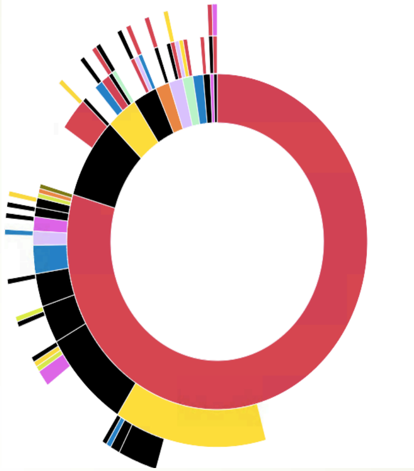
Aim 1 example: Type II DM treatment pathways in cancer care



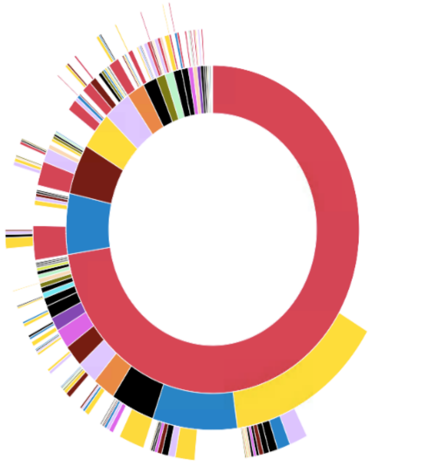
Truven CCAE



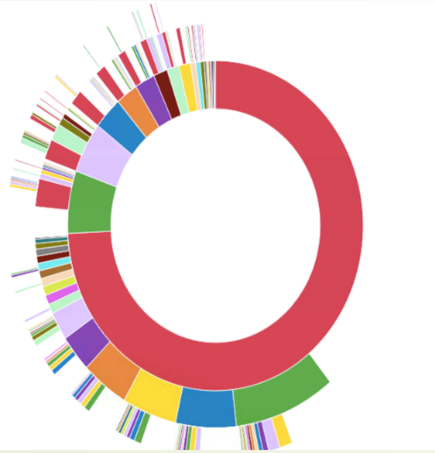
Columbia



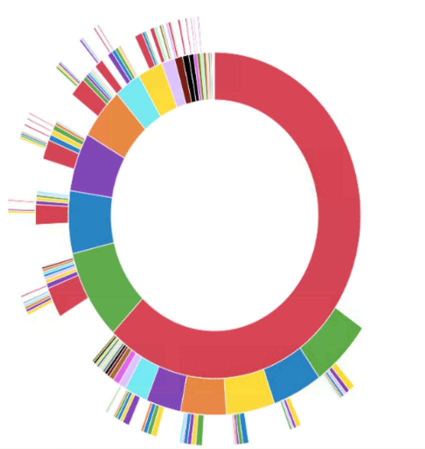
IMS France



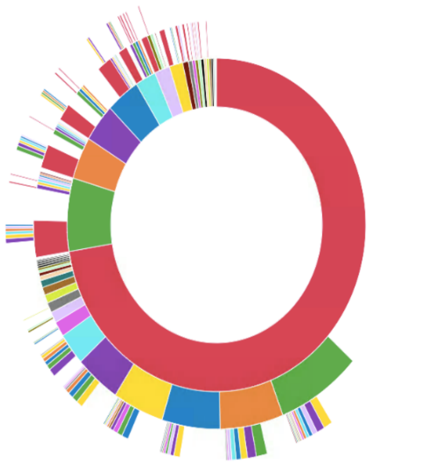
IMS Germany



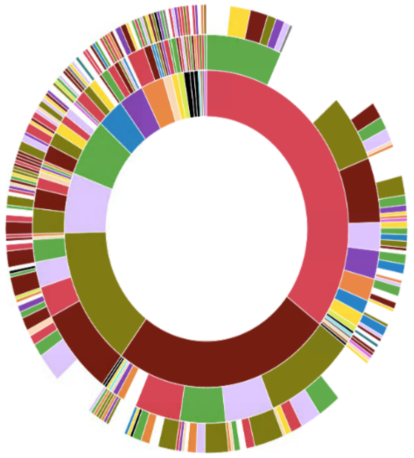
Truven Medicaid



Truven Medicare

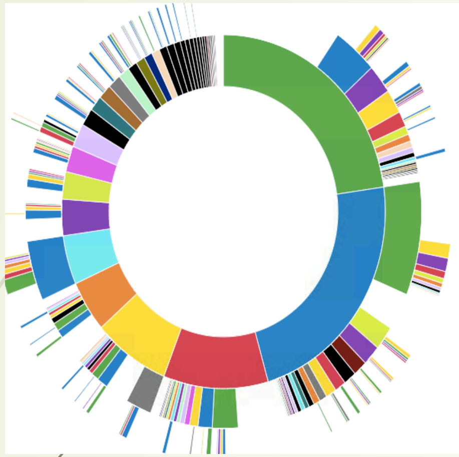


Optum Extended SES

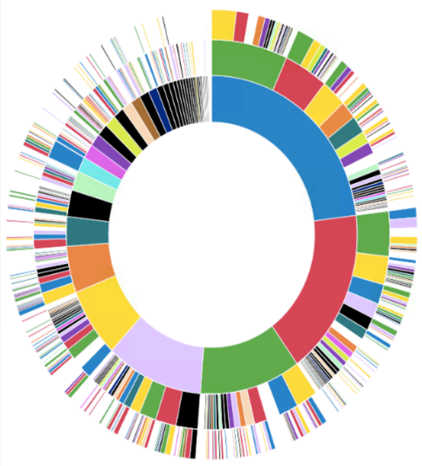


Stanford

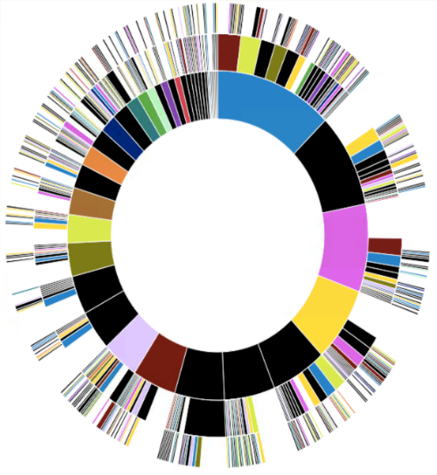
Aim 1 example: Hypertension treatment pathways in cancer care



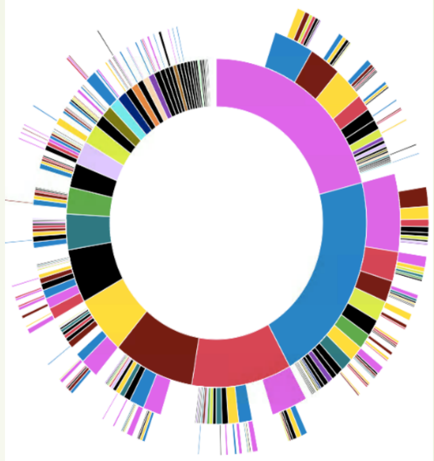
Truven CCAE



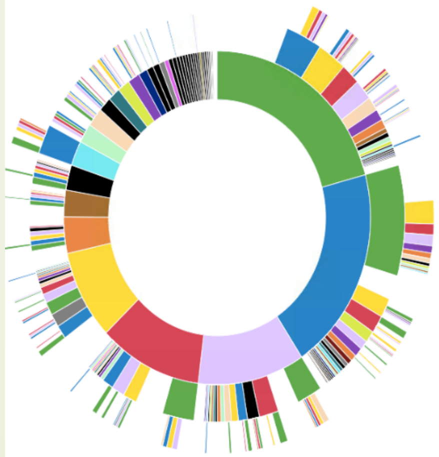
Columbia



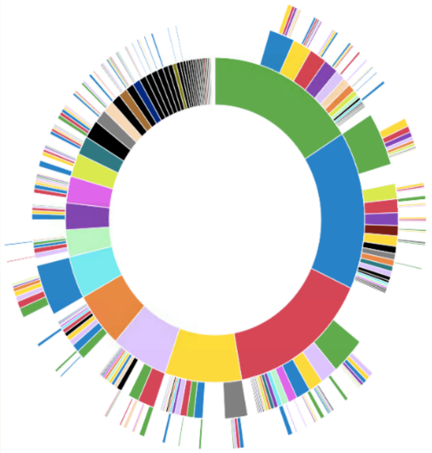
IMS France



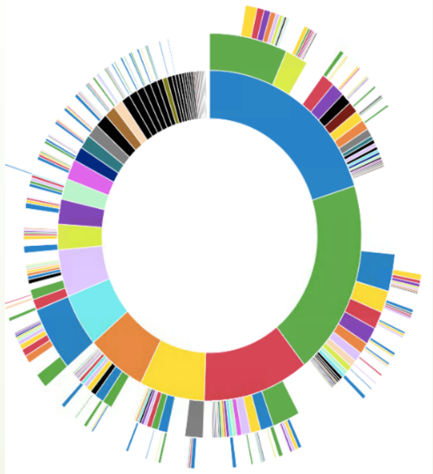
IMS Germany



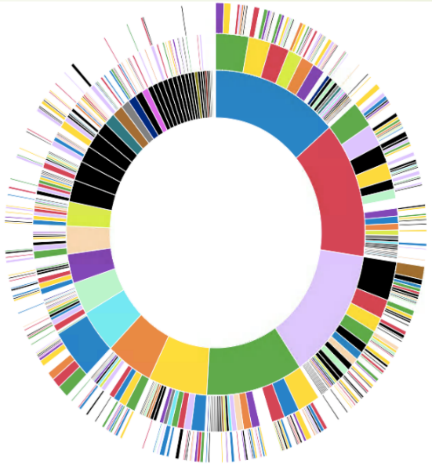
Truven Medicaid



Truven Medicare



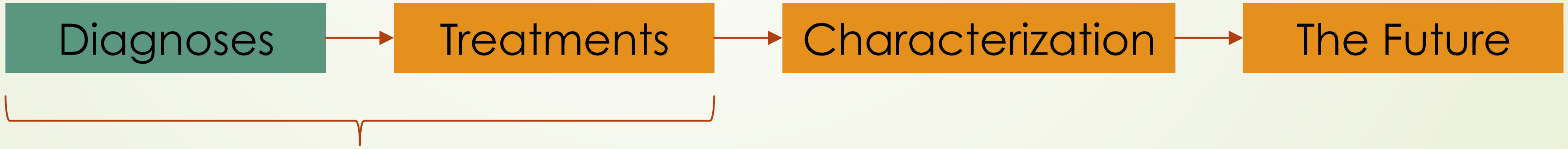
Optum Extended SES



Stanford

Aim 2: Phenotyping and Validation of Cancer Diagnoses

Any cancer, AML, CLL, pancreatic, and prostate cancer



Diagnoses → Treatments → Characterization → The Future

How good is the data?

Cancer Demographics at Columbia University Medical Center (CUMC)

- 6.38 million unique patient records
- 5.33 million unique patients with at least 1 diagnosis/condition
- 667,328 unique patients diagnosed with cancer
- 38,670 unique patients in cancer registry (NAACCR Tumor Registry)
  - Includes patients with reportable cancers actively being treated at NYP/CUMC
  - Collected as part of hospital's ACOS accreditation
  - Manually abstracted by contracted 3rd party of cancer registrars
  - Requires significant time/manpower, can lag behind real time especially with data dictionary updates

Cancer Demographics at Columbia University Medical Center (CUMC)

- Most prevalent diagnoses (SNOMED diagnosis):
  - Prostate Cancer (Primary malignant neoplasm of prostate)
    - 33,094 unique pts
  - Breast Cancer (Primary malignant neoplasm of female breast)
    - 31,281 unique pts
  - Unknown (Primary malignant neoplasm of unspecified site)
    - 28,428 unique pts
  - Lung cancer (Primary malignant neoplasm of respiratory tract)
    - 20,134 unique pts

Phenotyping and Validation of Cancer Diagnoses: Any Cancer

- All conditions ICD9CM, ICD10CM, SNOMED mapped to standardized SNOMED codes
  - SNOMED codes => condition_concept_id
    - **55342001-Neoplastic Disease => 438112**
  - Excluding:
    - **20376005-Benign neoplastic disease => 435506**
    - **254827004-Lipomatous tumor => 4112852**

Phenotyping and Validation of Cancer Diagnoses: Any Cancer

- Manual chart review of 100 patients (randomly chosen)
  - 2 pt charts unable to be found
  - 94/98 (95.9%) pts verified to have cancer
- **PPV: 95.9%**
- Using chemotherapy agent doxorubicin to identify false negatives and calculate sensitivity
  - 2294 patients with drug exposure to doxorubicin (ancestor is doxorubicin ingredient)
  - 2270 patients with diagnosis of cancer and drug exposure to doxorubicin
  - Missing: 24/2294 (1%) patients exposed to doxorubicin who were not captured as having cancer based on these SNOMED codes
- **Sensitivity: 99%**

Phenotyping and Validation of Cancer Diagnoses: Any Cancer

- Using other chemotherapy drugs to verify sensitivity:
  - 1589/1599 (**99.4%**) patients who got cisplatin correctly identified as having cancer
  - 1986/2133 (**93.1%**) patients who got fluorouracil (5FU) correctly identified as having cancer
  - 2351/2366 (**99.4%**) who got carboplatin correctly identified as having cancer
  - 107/112 (**95.5%**) who got abiraterone correctly identified as having cancer
  - 2217/2222 (**99.8%**) who got docetaxel correctly identified as having cancer

Phenotyping and Validation of Cancer Diagnoses: Any Cancer

- PPV 95.9% from manual review
- Sensitivity 99% using doxorubicin exposure
- Specificity (calculated): 99.87%

Phenotyping and Validation of Cancer Diagnoses: Next Step

- Chose 4 specific cancers for in-depth review
- Represent a variety of malignancies
  - Solid tumor vs hematologic
  - Aggressive vs indolent
  - Adult vs Pediatric

- AML
- CLL
- Pancreatic Cancer
- Prostate cancer

Phenotyping and Validation of Cancer Diagnoses: AML

- SNOMED codes => OMOP condition concept_id's
  - SNOMED: 91861009; Acute myeloid leukemia, disease (disorder)
  - Condition concept_id: 140352

- CUMC stats:
  - 2619 unique patients with AML
  - 95,875 condition occurrences of AML

Phenotyping and Validation of Cancer Diagnoses: AML

- Validation: random selection of 100 patients for chart review; manually reviewed first 51
- 50/51 had cancer
- 36/51 confirmed as AML (large portion was ALL incorrectly coded)
- **PPV: 70.6%**

Phenotyping and Validation of Cancer Diagnoses: AML

- Validation—Using Columbia's cancer registry as gold standard for sensitivity
  - Identify whether pts in registry with AML are identified using our phenotype => obtain false negative rate
- 190 patients in registry with morphotype of 9861/3 (ICD-O 9861/3--Acute myeloid leukemia, NOS)
- 184 found in CUMC_pending using SNOMED code/phenotype above
- **Sensitivity: 96.8%**
- Based on prevalence of .001,
- **Specificity: 99.9%**

Phenotyping and Validation of Cancer Diagnoses: CLL

- SNOMED codes => OMOP condition concept_id's
  - SNOMED: 92814006 ; Chronic lymphoid leukemia
  - Condition concept_id: 138379

- CUMC stats:
  - 3354 unique patients with CLL
  - 114,991 condition occurrences of CLL

Phenotyping and Validation of Cancer Diagnoses: CLL

- Validation: random selection of 90 patients for manual review
- 70/90 confirmed as CLL
- **PPV: 77.8%**

If addition of RxNorm codes in addition to SNOMED code to identify CLL patients:

- 397 unique patients identified using this method
  - Sampled 32 patients (~8% of identified patients in total),

    30 of whom were confirmed in source data to have CLL

    → PPV of 93.75%

Phenotyping and Validation of Cancer Diagnoses: CLL

- Validation—Using cancer registry as gold standard for sensitivity
- 421 patients in registry with morphotype of 9823/3 (9823/3 Chronic lymphocytic leukemia/small lymphocytic lymphoma)
- 403 found in CUMC_pending using SNOMED code/phenotype from earlier
- **Sensitivity: 95.7%**
- Based on prevalence of .00135,
- **Specificity: 99.9%**

Phenotyping and Validation of Cancer Diagnoses: Pancreatic Cancer

- SNOMED codes => OMOP condition concept_id's
  - SNOMED: 126859007; Neoplasm of pancreas
  - Condition concept_id: 4129886
- Exclude:
  - Benign neoplasm of pancreas SNOMED 92264007 (concept_id: 4243445)
  - Benign tumor of exocrine pancreas SNOMED 271956003 (concept_id: 4156048)
- CUMC stats:
  - 10,241 unique patients with pancreatic cancer
  - 199,988 condition occurrences of pancreatic cancer

Phenotyping and Validation of Cancer Diagnoses: Pancreatic Cancer

- Validation: random selection of 100 patients for chart review; manually reviewed first 50
- 50/50 had cancer
- 44/50 confirmed as pancreatic cancer (5 incorrectly diagnosed; 1 unclear because dx was in 1993)
- **PPV: 88%**

Phenotyping and Validation of Cancer Diagnoses: Pancreatic Cancer

- 1206 patients in registry with morphotype of 8140/3 (Adenocarcinoma) and primary site of C25.* (Pancreas)
- 1194 found in CUMC_pending using SNOMED code/phenotype above
- **Sensitivity: 99.0%**

- Based on prevalence of .004,
- **Specificity: 99.9%**

Phenotyping and Validation of Cancer Diagnoses: Prostate Cancer

- SNOMED codes => OMOP condition concept_id's
  - SNOMED: 399068003 ; Malignant tumor of prostate
  - Condition concept_id: 4163261
- Exclude:
  - Secondary malignant neoplasm of prostate, SNOMED 94503003 (concept_id=4314337)
  - Non-hogkin's lymphoma of prostate, SNOMED 449318001 (concept_id=40486666)
- CUMC stats:
  - 37,157 unique patients with prostate cancer
  - 455,562 condition occurrences

Phenotyping and Validation of Cancer Diagnoses: Prostate Cancer

- Validation: random selection of 100 patients for chart review; manually reviewed first 50
- 48/50 had cancer (1 unclear as it was never biopsied, just suspected)
- 47/50 confirmed as prostate cancer
- **PPV: 94%**

Phenotyping and Validation of Cancer Diagnoses: Prostate Cancer

- Validation—Using cancer registry as gold standard for sensitivity
- 3803 patients in registry with morphotype of 8140/3 adenocarcinoma and primary site =C61.9 (would have been 4807 with just C61.9 for prostate)
- 3786 found in CUMC_pending using SNOMED code/phenotype above
- **Sensitivity: 99.6%**
- Based on prevalence of .015,
- **Specificity: 99.9%**

Phenotyping and Validation of Cancer Diagnoses: Summary

	PPV	Sensitivity	Specificity
Any cancer	95.9%	98.9%	99.87%
AML	70.6%	96.8%	99.9%
CLL	77.8%	95.7%	99.9%
Pancreatic	88.0%	99.0%	99.9%
Prostate	94.0%	99.6%	99.9%

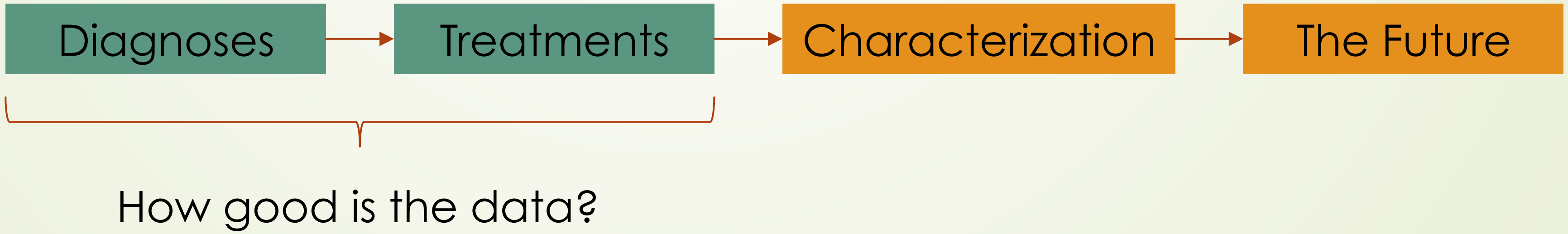
Phenotyping and Validation of Cancer Diagnoses

- **What we learned**

- Overall, feasible to accurately create phenotypes for subsets of cancer diagnoses and validate against chart and cancer registries
- Errors in coding can lead to lower PPV
  - AML miscoded as ALL or vice versa
  - Hematologic malignancies more likely to be miscoded
  - However, due to low prevalence, still high specificity and sensitivity
- Later dates/recent data are more reliable and accurate for coding

Aim 2: Phenotyping and Validation of Cancer Treatments

Chemotherapy, hormone therapy, immunotherapy, radiation therapy, and procedures



Diagnoses → Treatments → Characterization → The Future

How good is the data?

Phenotyping and Validation of Cancer Treatments: Chemotherapy

- Utilized WHO-ATC list of antineoplastic agents (L01)
- WHO: ATC list of Antineoplastic agents
  - 163 RxNorm codes
  - 162 concept_id's found from these RxNorm codes in CUMC (missing inotuzumab ozogamicin, 1942950)

- 536,082 drug exposures to 162 RxNorm codes found in CUMC
- Significant proportion included celecoxib and tretinoin
  - Excluded celecoxib as antineoplastic benefit not an indication and still being proven
  - Can tailor future studies to include/exclude tretinoin to improve accuracy, depending on whether it is used for the cancer of interest

Phenotyping and Validation of Cancer Treatments: Chemotherapy

- Validation: random selection of 100 patients for chart review; manually reviewed first 50 (if available in inpatient EMR)
- 50/50 received the drug at the time specified (correct drug exposure)
- 41/50 received drug for cancer
- **PPV: 100% for drug exposure; 82% as chemotherapy for cancer**

Phenotyping and Validation of Cancer Treatments: Hormone therapy

- Utilized WHO-ATC list for Endocrine therapy (L02) under ANTINEOPLASTIC AND IMMUNOMODULATING AGENTS (265)
- 27 RxNorm codes/drugs
- 27 concept_id's found in CUMC
- Limitations: includes estradiol in OCPs, medroxyprogesterone (depo-provera)
- Excluding estradiol and medoxyprogesterone:
  - 123012 drug exposures to hormone therapy
  - 20462 unique patients

Phenotyping and Validation of Cancer Treatments: Hormone therapy

- Validation: random selection of 100 patients for chart review; manually reviewed first 50 (in inpatient EMR)
- 49/50 received the drug specified
- 43/50 patients received the drug at the correct date/time listed in database
  - Most of discrepancies due to recording of med rec as exposure (will be fixed)
- 42/50 patients received the drug specified for cancer treatment
  - Megestrol was the primary medication identified not used for cancer
- **PPV: 98% for receiving the drug documented; 84% for receiving the drug as cancer therapy**

Phenotyping and Validation of Cancer Treatments: Immune therapy

- WHO-ATC list; L01 Antineoplastic Agents
  - L01X Other Antineoplastic Agents
    - L01XC Monoclonal Antibodies

- 23 concept_id's found from these RxNorm codes in CUMC
- 36,285 drug exposures
- 4,531 patients

Phenotyping and Validation of Cancer Treatments: Immune therapy

- Validation: random selection of 100 patients for chart review; manually reviewed first 50 (in inpatient EMR)
- 50/50 received the drug specified
- 49/50 received drug at the time specified
  - Because IV medications/infusions and considered chemo, generally great documentation; only wrong time was for old chart
- 47/50 patients received the drug specified for cancer treatment
  - All exceptions due to Rituximab; only drug used for non-cancer treatments
- **PPV: 100% for receiving the drug documented**; **94% for receiving the drug as cancer therapy**

Phenotyping and Validation of Cancer Treatments: Drug therapies

- **What we learned**

- Most of observational EHR data/OMOP accurately identifies drug exposures
- No currently curated list is perfect
  - Non-cancer related chemotherapy identified often for rheumatologic or other hematologic disorders: RA, sickle cell, polycythemia
  - Hormone therapy: Megace; Immunotherapy: rituximab
- Adding more criteria to phenotypes further improves accuracy (add cancer as condition to reduce false positives)
- Limitation: do not yet have access to outpatient EMR; may find more drugs used for non-chemotherapy purposes there

Phenotyping and Validation of Cancer Treatments: Radiation Therapy

- List of procedure codes (CPT4, ICD9Proc, HCPCS, Revenue codes) from NCI Cancer Research Network
- 266 codes total
- 190755 procedure exposures in CUMC_merged (all through CPT4 or ICD9Proc)
- 19,950 unique patients

Phenotyping and Validation of Cancer Treatments: Radiation Therapy

- Validation: random selection of 100 patients for chart review; manually reviewed first 50 (if in inpatient EMR)
- 49/50 patients had cancer (1 unknown; from 1989), despite no cancer condition codes
- 43/50 patients received RT ever; 2/50 clearly did not (wrongly coded); 5/50 patients unknown (lack of inpatient chart data)
- 11/18 with clear inpatient dates of administration received RT on the exact date specified by the procedure code
- **PPV: 86% for receiving RT**

Phenotyping and Validation of Cancer Treatments: Radiation Therapy

- **What we learned**

- Overall, can effectively identify patients receiving RT using procedure codes
- Dates may not be exact, sometimes date of note does not reflect date of procedure
- Dates better aligned, improved documentation in recent years
- RT frequently outpatient only
- Current code sets can be modified to improve/adjust PPV, sensitivity, specificity (i.e removing planning codes from NCI)

Phenotyping and Validation of Cancer Treatments: Registry

- As with diagnoses, used local NAACCR tumor registry as gold standard to determine sensitivity
- Registry treatments coded based on SEER*Rx categorization of medications
- Our phenotypes categorized treatments based on codes from WHO-ATC (for medications) and NCI Cancer Research Network (for RT)
- Often dramatic differences in code list
  - NAACCR registry and SEER*Rx include clinical trial/experimental drugs
    - For future studies, may be feasible to use NLP to extract from clinician notes if available
  - SEER*Rx always the larger code set
  - But only drug name, no mappings to any standardized vocabularies
  - For example, immunotherapy-27 codes in WHO-ATC; 2490 drugs in SEER*Rx

Phenotyping and Validation of Cancer Treatments: Registry

- **Chemotherapy**
- 162 RxNorm codes/drugs in our phenotype based on WHO-ATC
- 5066 drugs in SEER*Rx drug list
- Using RxDateChemo field in NAACCR Registry, determined if patient ever received chemotherapy
- 8476/12392 patients from registry also found based on WHO-ATC codes and current phenotype
    - Sensitivity: 68.4%

Phenotyping and Validation of Cancer Treatments: Registry

- **Hormone therapy**
- 27 RxNorm codes/drugs in our phenotype based on WHO-ATC
- 1460 drugs in SEER*Rx drug list
- Using RxDateHormone field in NAACCR Registry, determined if patient ever received hormone therapy
- 2863/5846 patients from registry also found based on WHO-ATC codes and current phenotype
  - Sensitivity: 49.0%

Phenotyping and Validation of Cancer Treatments: Registry

- **Immunotherapy**
- 23 RxNorm codes/drugs in our phenotype based on WHO-ATC
- 2429 drugs in SEER*Rx drug list for 'Biologic therapy (BRM, immunotherapy)
- Using RxDateBRM field in NAACCR Registry, determined if patient ever received biologic therapy (not specific for immunotherapy)
- 232/1466 patients from registry also found based on WHO-ATC codes and current phenotype
  - Sensitivity: 15.8%
- Then redefined phenotype to use broader set of WHO-ATC does
- 735/1466 patients from registry found using new phenotype
  - Sensitivity: 50.1%

Phenotyping and Validation of Cancer Treatments: Registry

- **Radiation therapy**
- Using RxDateRadiation field in NAACCR Registry, determined if patient ever received radiation therapy
- 5081/7539 patients from registry also found based on current phenotype of procedure/revenue codes
  - Sensitivity: 67.4%

Phenotyping and Validation of Cancer Treatments: Registry

	PPV (from chart review)	Sensitivity (from registry)	Prevalence	Specificity
Chemotherapy	100%	68.4%	0.23%	99.9%
Hormone therapy	98%	49.0%	0.11%	99.9%
Immuno-therapy	100%	15.8%/50.1%*	0.03%	99.9%
Radiation therapy	86%	67.4%	0.14%	99.9%

*Based on different phenotypes from narrow and broader code sets, respectively, from WHO-ATC

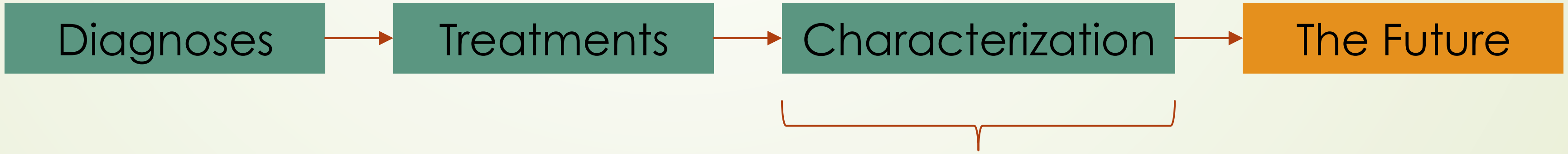
Phenotyping and Validation of Cancer Treatments: Registry

- **What we learned**

- Feasible to create phenotypes for various types of treatments for cancer but may require more modification and testing than diagnoses
- Sensitivity may vary widely depending on phenotype and code set used as 'gold standard'
  - Low when only capturing a small subset of the coded treatments in gold standard (large discrepancy in number of codes between phenotype and registry)
  - Some drug and procedure codes may miss clinical trial/experimental drugs and treatments
- Sensitivity can be improved by modifying the created phenotype
  - i.e. broadening immunotherapy codes to better match registry improved sensitivity more than 4-fold
- Specificity remains high due to low prevalence

Aim 2: Characterizing Treatments over Time

One example of clinical characterization study



Diagnoses → Treatments → Characterization → The Future

What can we do with the data?

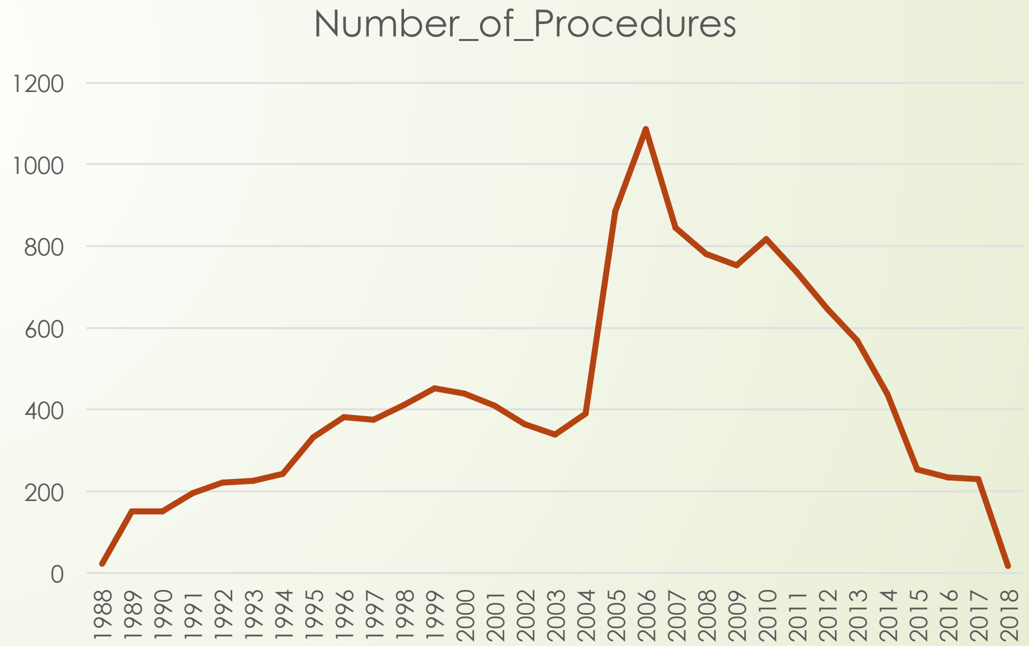
Treatments over Time-Prostate Cancer

- Prostatectomy codes:
  - SNOMED 90470006 (concept_id=4235738)
  - MedDRA 10061916 (concept_id=37521400)
  - CPT: 2109825 (transurethral electrosurgical resection), 2110031 (perineal, partial resection), 2110032 (perineal, radical),
  - 2110033 (perineal, radical, with lymph node biopsy), 2110034 (perineal, radical, with bilateral pelvic lymphadenectomy)
  - 2110036 (retropubic, partial resection), 2110037 (retropubic, radical), 2110038 (retropubic, radical, with lymp node biopsy)
  - 2110039 (retropubic, radical, with bilateral pelvic lymphadenectomy)
  - ICD-10-CM PCS: 2805820 (excision), 2899589 (resection)

Treatments over Time-Prostate Cancer

- Prostatectomy

Year_of_Procedure_Start	Number_of_Procedures
1988	23
1989	151
1990	151
1991	196
1992	221
1993	225
1994	242
1995	332
1996	381
1997	374
1998	412
1999	452
2000	440
2001	410
2002	365
2003	339
2004	389
2005	885
2006	1087
2007	845
2008	781
2009	754
2010	816
2011	736
2012	647
2013	569
2014	436
2015	253
2016	234
2017	229
2018	17



Treatments over Time-Prostate Cancer

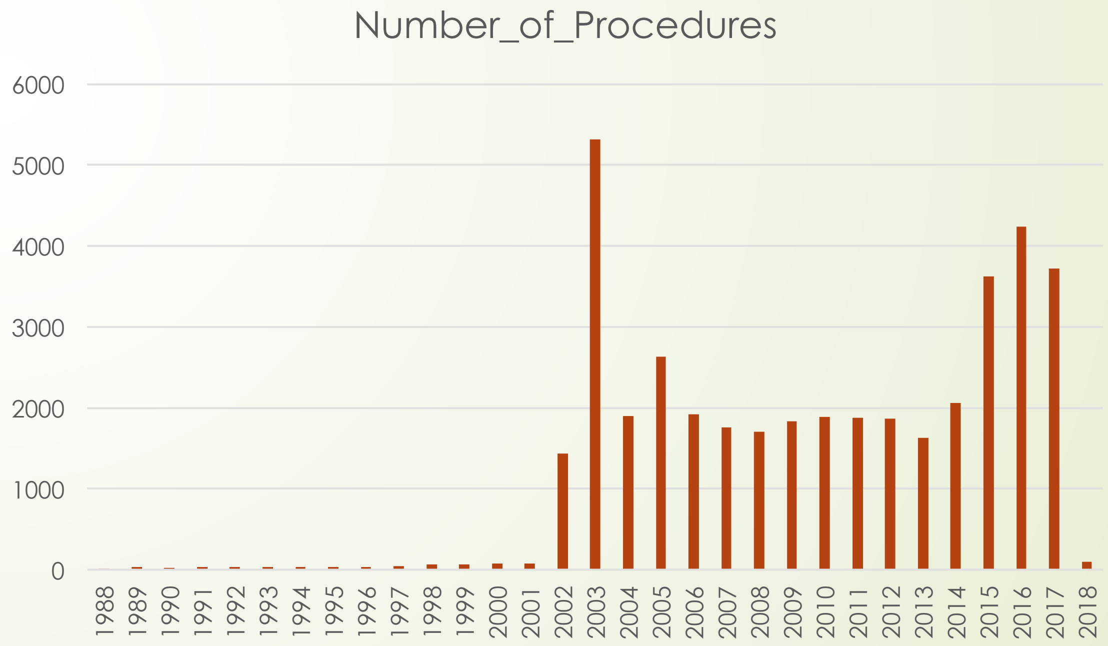
- Radiation Therapy codes

  -
    ```
    where (vocabulary_id like 'CPT4' and procedure_source_value in
    ('0073T','0082T','0083T','0182T','0190T','0197T','19296','19297','19298','20555','20660','31463','32553','41019','49411','4
    9412','52250','55859','55860','55875','55876','55920','57155','57156','58346','61720','61735','61770','61781','61782','61
    783','61793','61795','61796','61797','61798','61799','61800','63620','63621','73670','76950','76965','77014','77261','772
    62','77263','77280','77285','77290','77295','77299','77300','77301','77305','77306','77307','77310','77315','77321','7732
    6','77326','77327','77327','77328','77328','77331','77332','77333','77334','77336','77338','77370','77370','77371','77372'
    ,'77373','77380','77381','77385','77386','77387','77399','77400','77401','77402','77403','77404','77405','77406','77407','
    77408','77409','77410','77411','77412','77413','77414','77415','77416','77417','77418','77419','77420','77421','77422','7
    7423','77425','77427','77430','77431','77432','77435','77469','77470','77499','77520','77522','77523','77525','77750','77
    761','77762','77763','77776','77777','77778','77781','77782','77783','77784','77785','77786','77787','77789','77790','777
    99','79005','79030','79035','79100','79101','79200','79300','79400','79403','79420','79440','79445','79900','79999')) /
    **** CPT4 codes ****/
    or (vocabulary_id like 'HCPCS' and procedure_source_value in
    ('A4650','A9606','A9699','C1715','C1716','C1717','C1718','C1719','C1720','C1728','C2616','C2633','C2634','C2635','
    C2636','C2637','C2638','C2639','C2640','C2641','C2642','C2643','C2698','C2699','C9726','C9728','G0173','G0174','
    G0242','G0243','G0251','G0338','G0339','G0340','G6003','G6004','G6005','G6006','G6007','G6008','G6009','G6010'
    ,'G6011','G6012','G6013','G6014','G6015','G6016','Q3001','S2270','S8049','C1325','C1348','C1350','C1700','','C1701'
    ,'C1702','C1703','C1704','C1705','C1706','C1707','C1708','C1709','C1710','C1711','C1712','C1790','','C1791','C1792
    ','C1793','C1794','C1795','C1796','C1797','C1798','C1799','C1800','C1801','C1802','C1803','C1804','C1805','C1806',
    'C2632','C9714','C9715','G0178','G0256','G0273','G0274','G0338','G0339','G0340','G0458','C2644','C2645')) /****
    HCPCS codes (CUMC doesn't use) ****/
    or  (vocabulary_id like 'ICD9Proc' and procedure_source_value in
    ('I9:92.2','I9:92.21','I9:92.22','I9:92.23','I9:92.24','I9:92.25','I9:92.26','I9:92.27','I9:92.28','I9:92.29','I9:92.3','I9:92.31','I9:92
    .32','I9:92.33','I9:92.39','I9:92.41')) /**** ICD9 codes ****/
    or  (vocabulary_id like 'Revenue' and procedure_source_value in ('0333','0344')) /**** Revenue codes (CUMC
    doesn't use) ****/)
    ```

Treatments over Time-Prostate Cancer

- Radiation Therapy

Year_of_Procedure_Start	Number_of_Procedures
1988	3
1989	32
1990	26
1991	38
1992	30
1993	32
1994	35
1995	36
1996	31
1997	40
1998	67
1999	67
2000	71
2001	72
2002	1431
2003	5314
2004	1901
2005	2632
2006	1920
2007	1755
2008	1708
2009	1829
2010	1892
2011	1880
2012	1868
2013	1625
2014	2065
2015	3624
2016	4235
2017	3725
2018	101



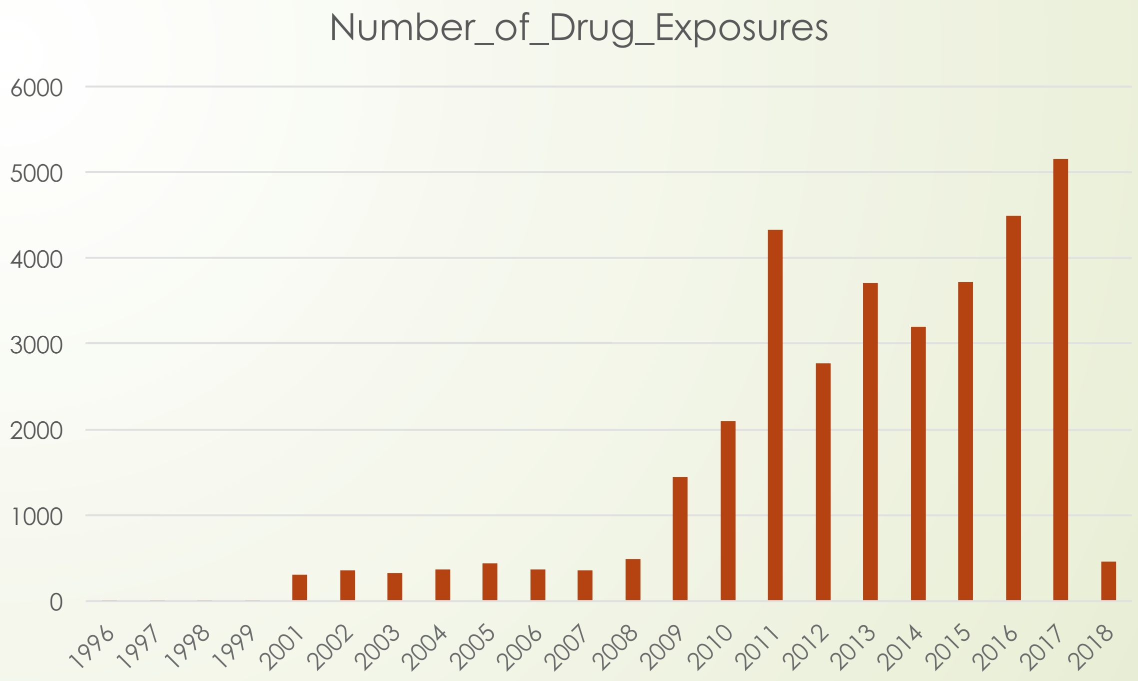
Treatments over Time-Prostate Cancer

- Hormone Therapy codes
  - Adrogen Deprivation Therapies
  - LHRH agonists
    - Goserelin, 1366310
    - Histrelin, 1366773
    - Leuprolide, 1351541
    - Triptorelin, 1343039
  - LHRH agonists (as above) plus first generation antiandrogen
    - LHRH agonist plus nilutamide, 1315286
    - LHRH agonist plus Flutamide, 1356461
    - LHRH agonist plus bicalutamide, 1344381
  - LHRH agonist (as above) plus second generation antiandrogen
    - LHRH agonist plus enzalutamide, 42900250
  - LHRH antagonist
  - ----Degarelix, 19058410--PROS11-PROS14
  - first and second generation antiandrogens (see above)
  - ketoconazole, 985708
  - ketoconazole plus hydrocortisone, 975125
  - PROS12-PROS14
    - abiraterone (40239056)

Treatments over Time-Prostate Cancer

- Hormone Therapy

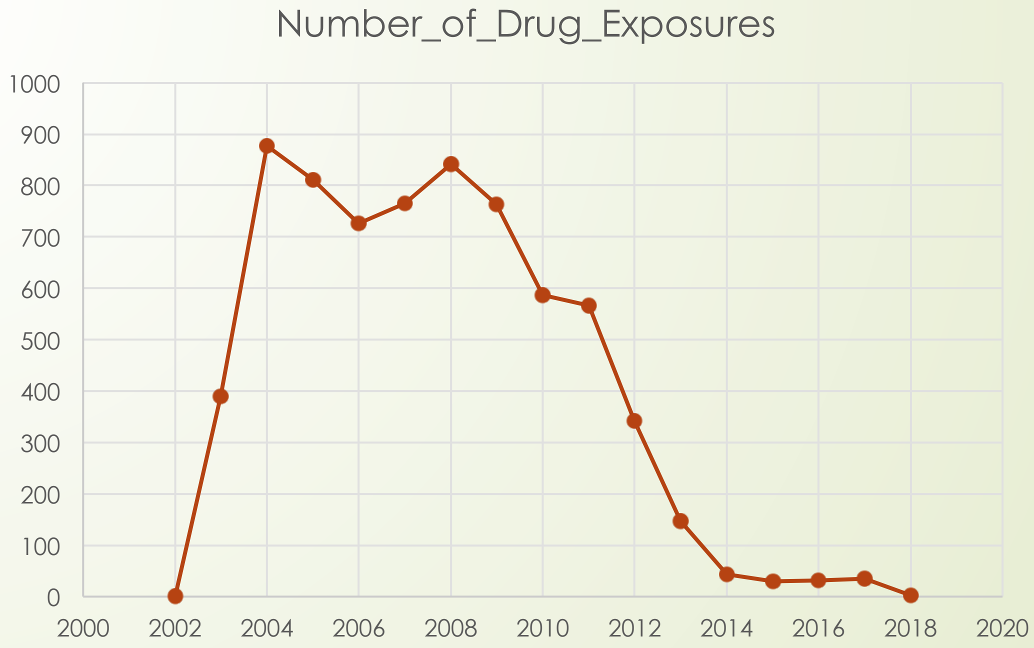
Year_of_Drug_Start	Number_of_Drug_Exposures
1996	2
1997	6
1998	12
1999	5
2001	307
2002	357
2003	325
2004	364
2005	437
2006	371
2007	354
2008	494
2009	1446
2010	2096
2011	4332
2012	2774
2013	3703
2014	3201
2015	3718
2016	4494
2017	5149
2018	462



Treatments over Time-Prostate Cancer

- Chemotherapy codes
  - Cisplatin, 1397599
  - Carboplatin, 1344905
  - Docetaxel, 1315942
  - Etoposide, 1350504

Year_of_Drug_Start	Number_of_Drug_Exposures
2002	1
2003	390
2004	878
2005	811
2006	726
2007	765
2008	842
2009	763
2010	586
2011	566
2012	343
2013	147
2014	44
2015	30
2016	31
2017	34
2018	2



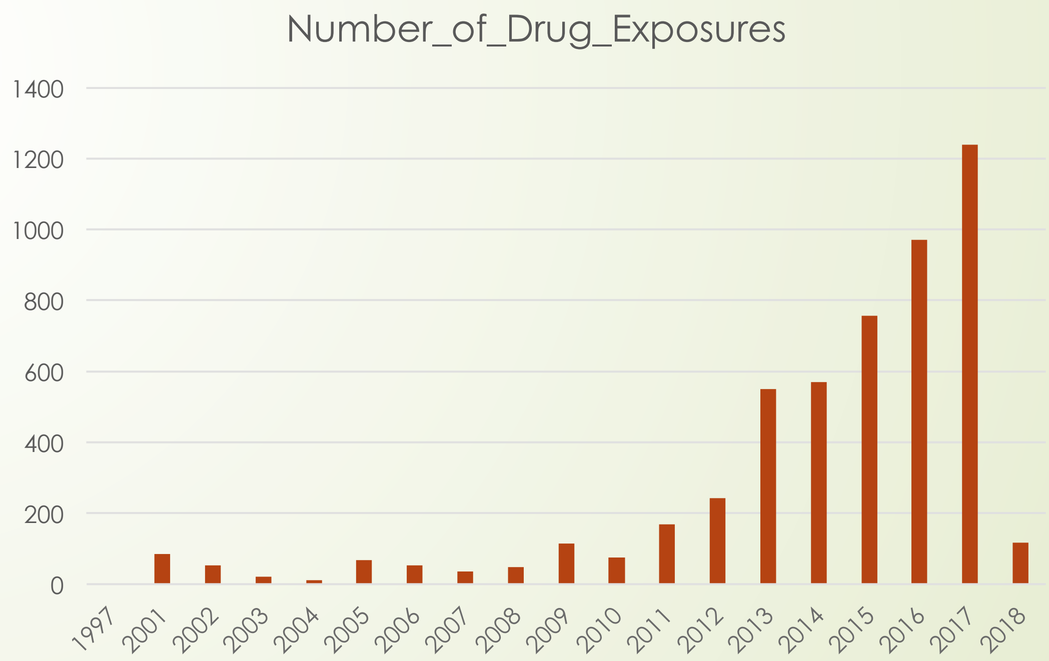
Treatments over Time-CLL

- Chemotherapy codes
  - Chlorambucil, 1390051 chemotherapy
  - Ibrutinib, 44507848 tyrosine kinase inhibitor
  - Bendamustine, 19015523 chemotherapy
  - Fludarabine, 1395557 chemotherapy
  - Cyclophosphamide, 1310317 chemotherapy
  - Pentostatin, 19031224 chemotherapy
  - Idelalisib, 45776944 kinase inhibitor
  - Venetoclax, 35604205 chemotherapy
  - (SEER categorized kinase inhibitors as chemo)

Treatments over Time-CLL

- Chemotherapy

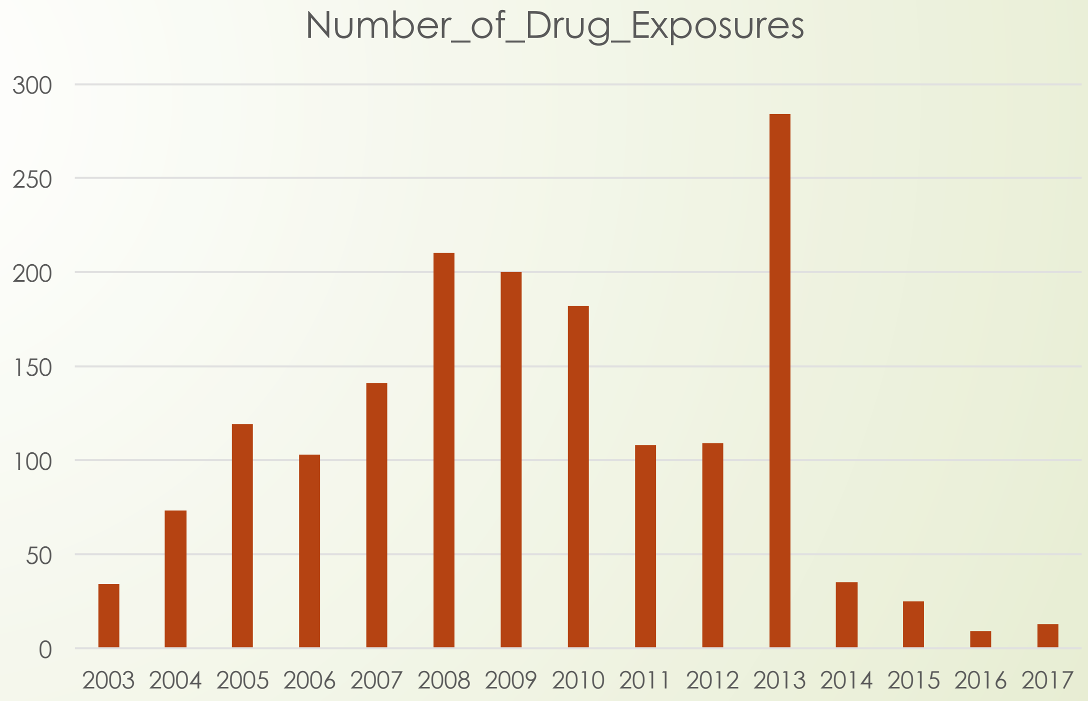
Year_of_Drug_Start	Number_of_Drug_Exposures
1997	1
2001	85
2002	54
2003	22
2004	12
2005	69
2006	52
2007	37
2008	49
2009	115
2010	76
2011	169
2012	243
2013	550
2014	571
2015	758
2016	972
2017	1240
2018	116



Treatment over Time-CLL

- Immune therapy
  - Obinutuzumab, 44507676 immune therapy/antibody
  - Ofatumumab, 40167582 immune therapy/antibody
  - Rituximab, 1314273 immune therapy/antibody

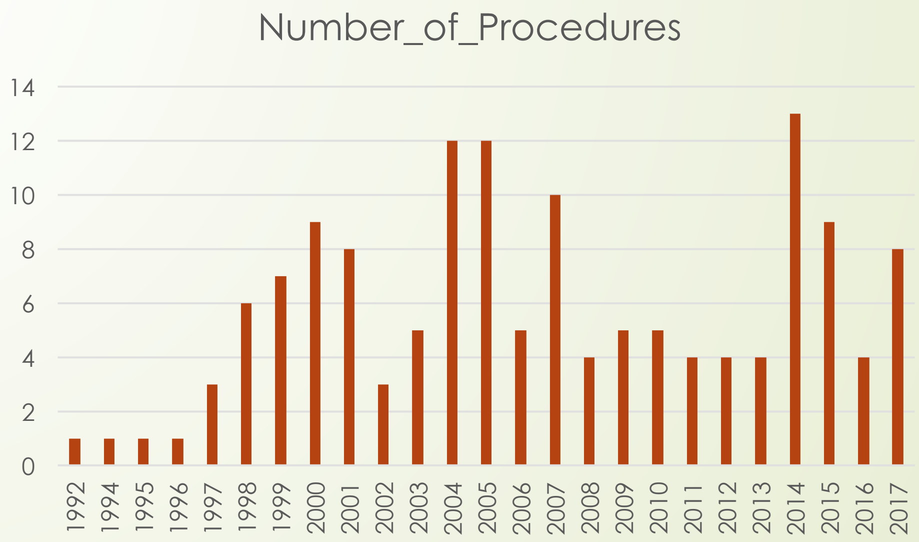
Year_of_Drug_Start	Number_of_Drug_Exposures
2003	34
2004	73
2005	119
2006	103
2007	141
2008	210
2009	200
2010	182
2011	108
2012	109
2013	284
2014	35
2015	25
2016	9
2017	13


Number_of_Drug_Exposures

Treatment over Time-CLL

- Procedures
  - Stem Cell Transplantation, concept_id: 4120445

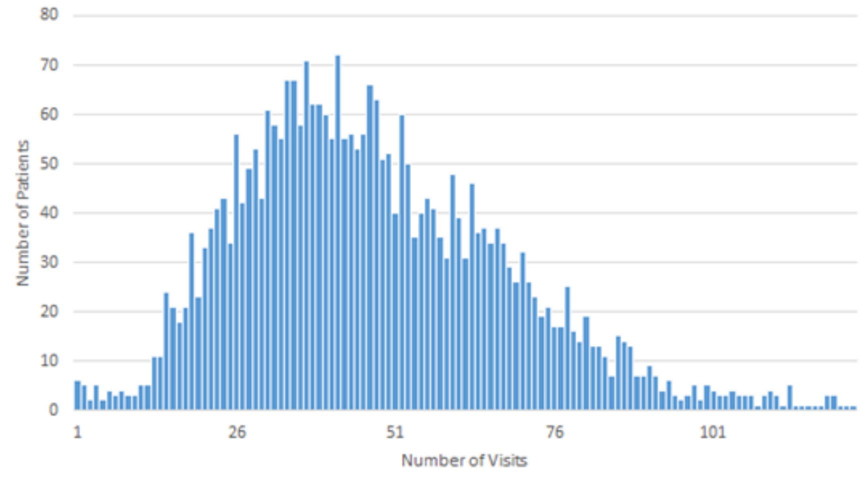
Year_of_Procedure_Start	Number_of_Procedures
1992	1
1994	1
1995	1
1996	1
1997	3
1998	6
1999	7
2000	9
2001	8
2002	3
2003	5
2004	12
2005	12
2006	5
2007	10
2008	4
2009	5
2010	5
2011	4
2012	4
2013	4
2014	13
2015	9
2016	4
2017	8



Feasibility of Additional Clinical Questions/Topics

- Treatment Burden: How many different clinical care providers/specialties do patients see in the first year after cancer diagnosis? How many visits?
  - Visits, procedures are all feasible to characterize now
  - Provider data not currently part of our merged (inpatient + outpatient database) but will be available on next reload; currently outpatient database only
  - In querying other sites, provider data is available in most databases; feasible to run network study

Outpatient visits/inpatient days over 1st year-All cancers



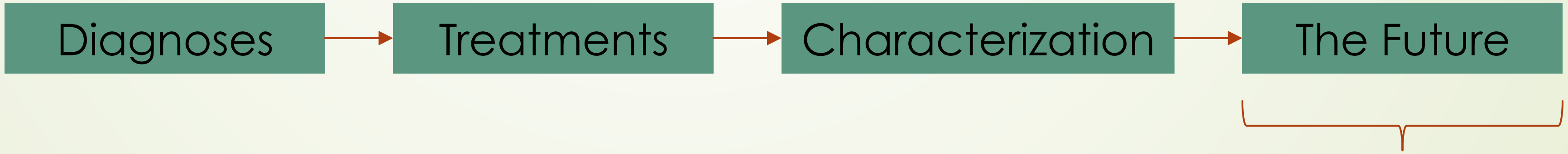
Feasibility of Additional Clinical Questions/Topics

- Treatment Burden: How many different clinical care providers/specialties do patients see in the first year after cancer diagnosis? How many visits?
    - Specialty data not consistently available but National Provider Identification (NPI) number often is
    - Created Python tool to identify and extract provider information from National Plan and Provider Enumeration System (NPPES) registry
    - Validated tool by comparing extracted taxonomy codes and specialty with specialty data from local credentialing system
    - 3752/3838 (97.7%) physicians had existing and valid NPI number identifiable in NPPES from which corresponding taxonomy data was successfully extracted
    - 3659/3752 (97.5%) concordance between taxonomy data extracted by tool and specialty 'gold standard' from credentialing system

Feasibility of Additional Clinical Questions/Topics

- Location data
  - For example: Where do Medicaid patients living in highly rural areas (e.g. counties with RUCC category codes of 7,8,9) receive their diagnostic services (e.g. imaging and laboratory facilities) and treatment (e.g. community oncology or academic oncology practices) and how far are these facilities from the patients
  - Currently do not have location information but can easily input; patient location is part of local source data
  - Inconsistent and vary wifely across other OHDSI sites and databases
    - Some have zip code, some have region, some only classify urban vs rural

OHDSI Oncology Working Group



Diagnoses → Treatments → Characterization → The Future

What are we working toward for the future?

Thank You!
Questions?

OHDSI Oncology Working Group-Challenges

- **Source data challenges**
  - In cancer registries, data are cleaned and abstracted, but limited in time and feature coverage. In Electronic Medical Records, oncology data are arguably the least structured type of data.
- **Modeling and terminology challenges**
  - In order to represent and reconcile these data in OMOP CDM, significant model, vocabulary, and convention extensions are required
- **Analytical derivation of the key disease features challenge**
  - To identify treatment episodes and response to treatment, cancer recurrences and progression of disease, we need to build derivation methods and tools

OHDSI Oncology Working Group-Goals

- Identification and follow-up of patients with a certain disease phenotype
- Identification of treatment regimen and response to treatment
- Identification of recurrences and progression of disease
- Prediction of recurrences, length of remissions, end of life events

OHDSI Oncology Working Group-Progress

- Mapping of ICD-O to SNOMED to represent cancer diagnosis
  - Extended OMOP vocabulary to support cancer diagnosis representation at the most granular level.
    - Extended SNOMED anatomy and morphology terminology to ensure direct mapping from ICD-O topography and histology respectively.
    - Extended pre-coordinated concepts representing intersection of anatomy and morphology to cover all SEER reported combinations of ICD-O histology and topography
  - Testing of new mappings is in progress at Columbia
- Representation of cancer diagnostic features in OMOP
  - Extended OMOP CDM to represent cancer occurrences and diagnostic features, like stage, grade, and others

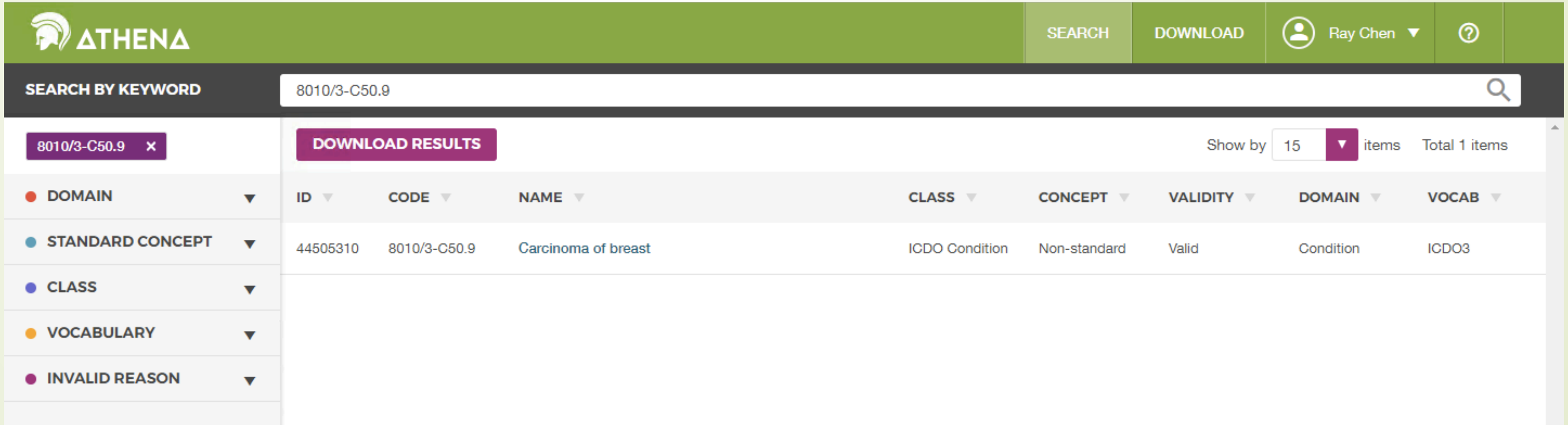
OHDSI Oncology Working Group

- OMOP vocabulary expanded to include mapping of ICD-O diagnoses to SNOMED
  - Extended SNOMED anatomy and morphology terminology
  - Direct mapping from ICD-O topography and histology to SNOMED
  - Covers all SEER reported combinations of ICD-O histology and topography

ICD-O Histology	ICDO Histology Desc	OMOP Morphology Concept ID	OMOP Morphology Concept Name	ICDO Topology	ICDO Topology Desc	OMOP Anatomy Concept ID	OMOP Anatomy Concept Name	OMOP precoordinated Concept ID	OMOP precoordinated Concept Name
8010/3	Carcinoma, NOS	4287106	Carcinoma	C50.9	Breast, NOS	4298444	Breast structure	4116071	Carcinoma of breast

OHDSI Oncology Working Group

- OMOP vocabulary expanded to include mapping of ICD-O diagnoses to SNOMED
  - Extended SNOMED anatomy and morphology terminology
  - Direct mapping from ICD-O topography and histology to SNOMED
  - Covers all SEER reported combinations of ICD-O histology and topography

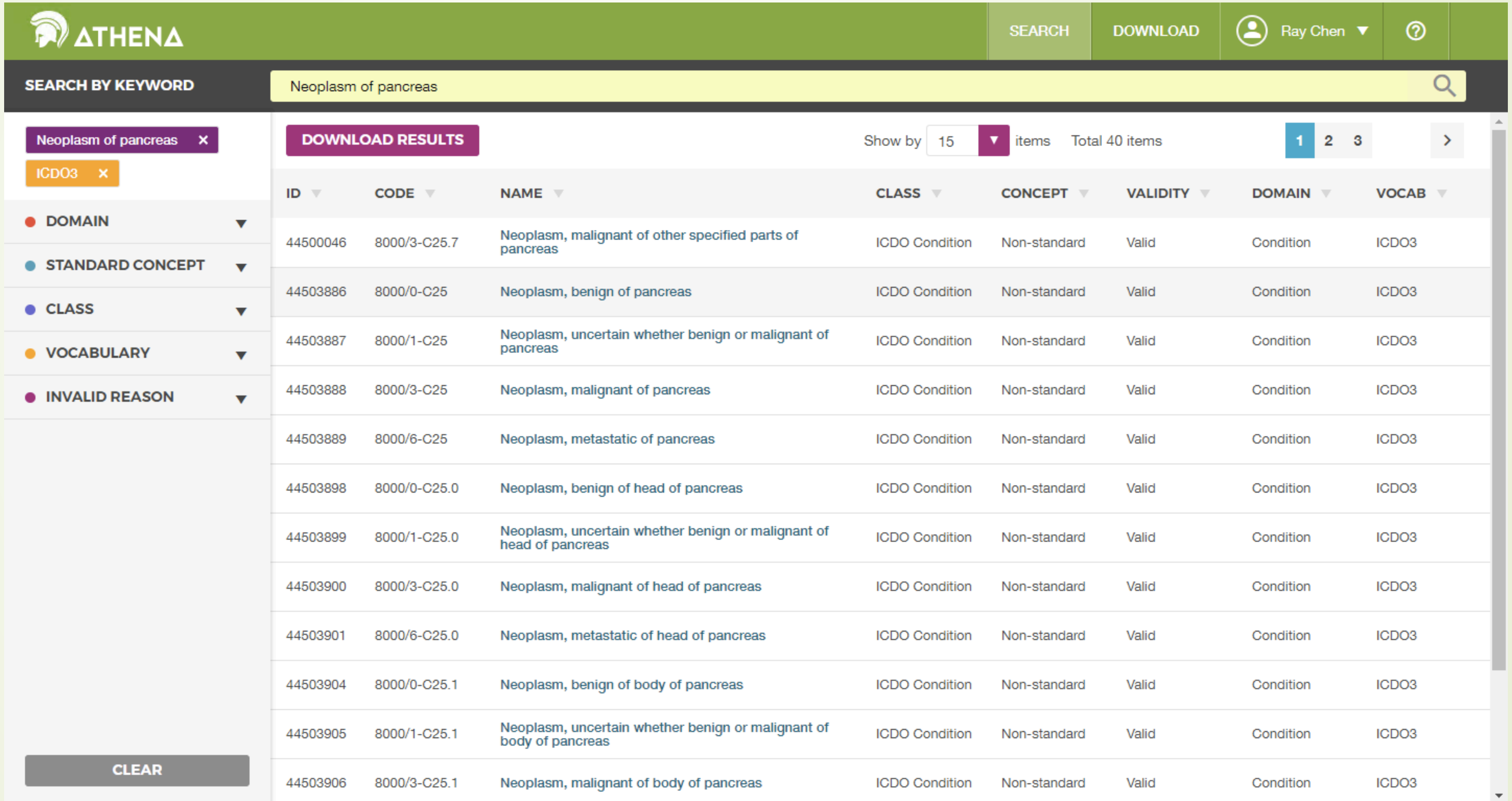


SEARCH BY KEYWORD: 8010/3-C50.9

Show by 15 items Total 1 items

ID	CODE	NAME	CLASS	CONCEPT	VALIDITY	DOMAIN	VOCAB
44505310	8010/3-C50.9	Carcinoma of breast	ICDO Condition	Non-standard	Valid	Condition	ICDO3

OHDSI Oncology Working Group



Search: Neoplasm of pancreas

Filters: Neoplasm of pancreas ✕ ICDO3 ✕

Show by 15 items Total 40 items 1 2 3 >

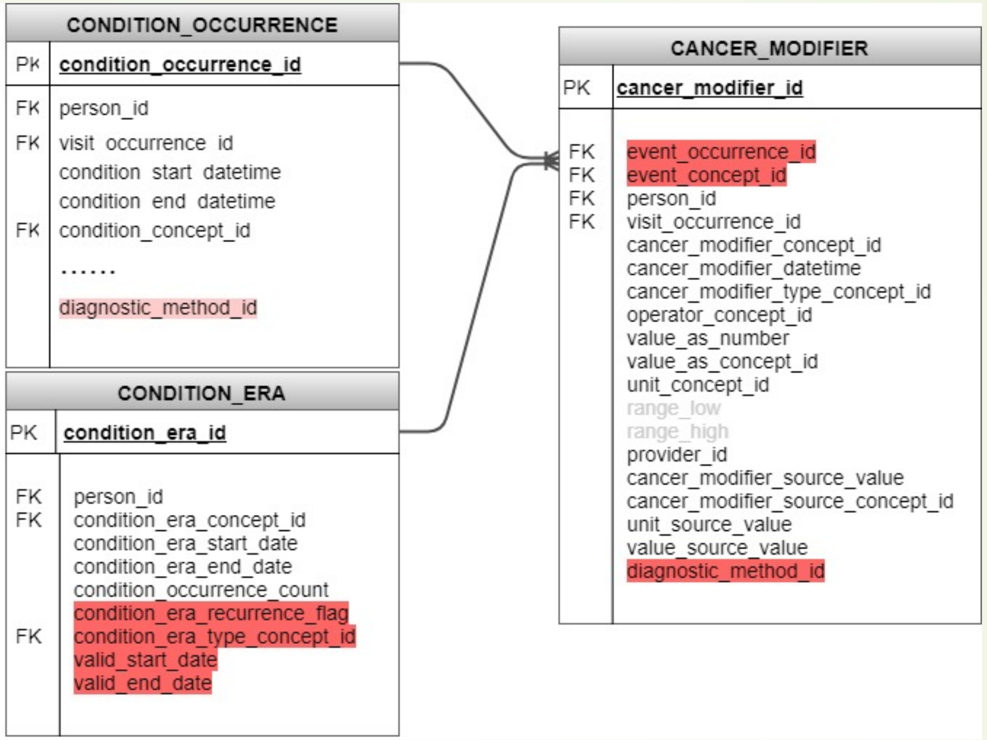
ID	CODE	NAME	CLASS	CONCEPT	VALIDITY	DOMAIN	VOCAB
44500046	8000/3-C25.7	Neoplasm, malignant of other specified parts of pancreas	ICDO Condition	Non-standard	Valid	Condition	ICDO3
44503886	8000/0-C25	Neoplasm, benign of pancreas	ICDO Condition	Non-standard	Valid	Condition	ICDO3
44503887	8000/1-C25	Neoplasm, uncertain whether benign or malignant of pancreas	ICDO Condition	Non-standard	Valid	Condition	ICDO3
44503888	8000/3-C25	Neoplasm, malignant of pancreas	ICDO Condition	Non-standard	Valid	Condition	ICDO3
44503889	8000/6-C25	Neoplasm, metastatic of pancreas	ICDO Condition	Non-standard	Valid	Condition	ICDO3
44503898	8000/0-C25.0	Neoplasm, benign of head of pancreas	ICDO Condition	Non-standard	Valid	Condition	ICDO3
44503899	8000/1-C25.0	Neoplasm, uncertain whether benign or malignant of head of pancreas	ICDO Condition	Non-standard	Valid	Condition	ICDO3
44503900	8000/3-C25.0	Neoplasm, malignant of head of pancreas	ICDO Condition	Non-standard	Valid	Condition	ICDO3
44503901	8000/6-C25.0	Neoplasm, metastatic of head of pancreas	ICDO Condition	Non-standard	Valid	Condition	ICDO3
44503904	8000/0-C25.1	Neoplasm, benign of body of pancreas	ICDO Condition	Non-standard	Valid	Condition	ICDO3
44503905	8000/1-C25.1	Neoplasm, uncertain whether benign or malignant of body of pancreas	ICDO Condition	Non-standard	Valid	Condition	ICDO3
44503906	8000/3-C25.1	Neoplasm, malignant of body of pancreas	ICDO Condition	Non-standard	Valid	Condition	ICDO3

OHDSI Oncology Working Group

- Representation of cancer diagnostic features in OMOP
  - Extended OMOP CDM to represent cancer occurrences and diagnostic features, like stage, grade, and others

- Goal is to represent cancer registries in OMOP CDM
  - Allow for as much coverage as is currently in registries at local, regional or national level
  - Important benefits of also having EHR data to provide greater breadth and longitudinal information about patients
  - Represent whole patient rather than cancer-centric view only

OMOP CDM Extension for Cancer Diagnosis



CONDITION_OCCURRENCE

PK condition_occurrence_id

FK person_id
FK visit_occurrence_id
condition_start_datetime
condition_end_datetime
FK condition_concept_id
……
diagnostic_method_id

CONDITION_ERA

PK condition_era_id

FK person_id
FK condition_era_concept_id
condition_era_start_date
condition_era_end_date
condition_occurrence_count
condition_era_recurrence_flag
FK condition_era_type_concept_id
valid_start_date
valid_end_date

CANCER_MODIFIER

PK cancer_modifier_id

FK event_occurrence_id
FK event_concept_id
FK person_id
FK visit_occurrence_id
cancer_modifier_concept_id
cancer_modifier_datetime
cancer_modifier_type_concept_id
operator_concept_id
value_as_number
value_as_concept_id
unit_concept_id
range_low
range_high
provider_id
cancer_modifier_source_value
cancer_modifier_source_concept_id
unit_source_value
value_source_value
diagnostic_method_id

- **event_occurrence_id** is a reference to an event the modifier modifies, in this case condition_occurrence_id or condition_era_id
- **event_concept_id** is a table an event is stored in, in this case ‘CONDITION_OCCURRENCE’ or ‘CONDITION_ERA’
- **condition_era_recurrence_flag** (Y/N) indicates if an era record represents cancer first occurrence or recurrence
- **condition_era_type_concept_id** identifies method of era derivation.
- **diagnostic_method_id** indicates how cancer diagnosis/diagnostic feature was diagnosed (e.g. pathology, symptomatically, record abstraction, etc.)

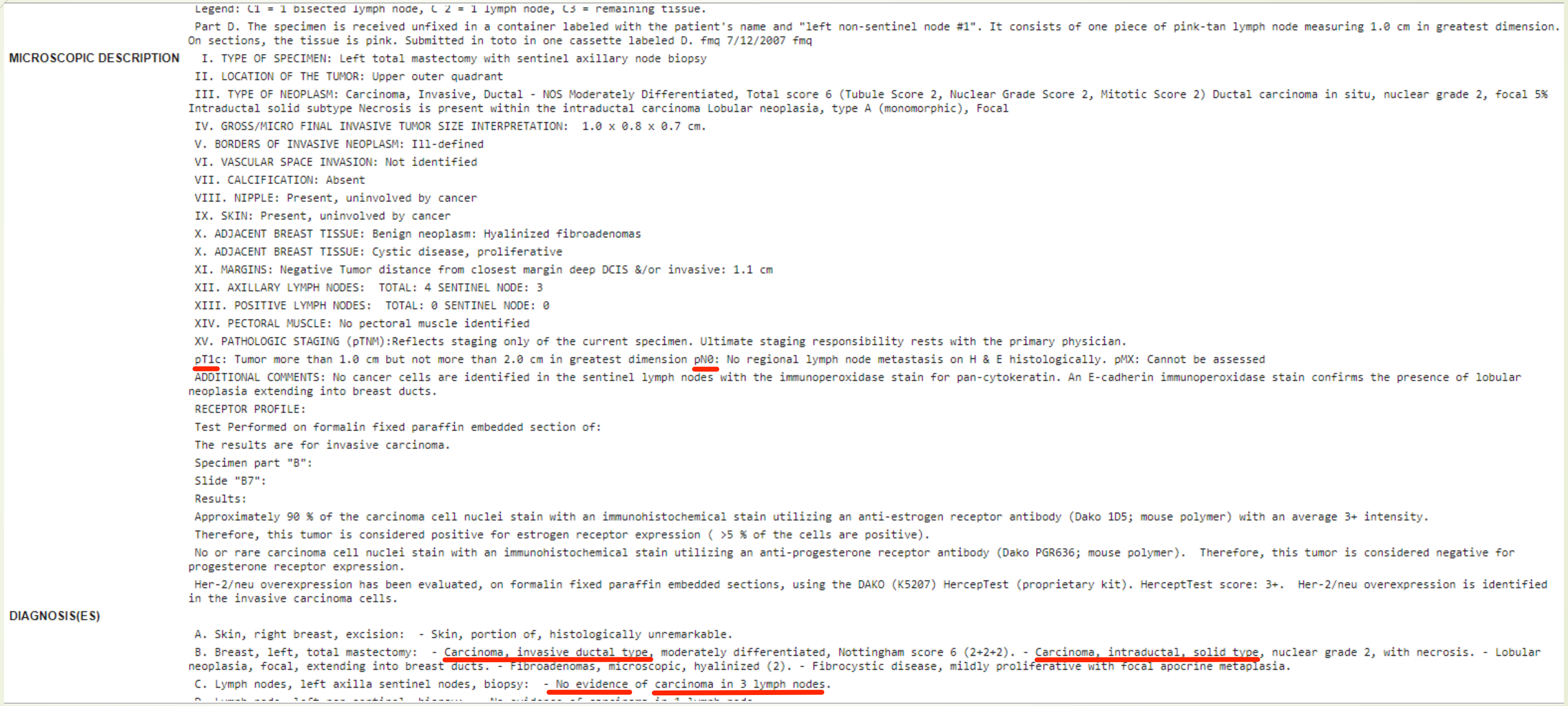
OHDSI Oncology Working Group-Next Steps/Future Work

- Treatment regimens: ongoing work additions to vocabulary and development of algorithms to derive treatment regimens
  - Feasible to execute studies on treatment regimen now if phenotype is developed and validated for each research study
  - i.e. for R-CHOP (drug exposure for Doxorubicin, Rituximab, Prednisone Cyclophosphamide, Vincristine within 2 days)

- Pathology reports included in notes as text
  - Can utilize NLP to extract desired features
  - In addition to representing diagnostic data from registry

- Continue ongoing work re: validation of ICD-O vocabulary, building of NAACR/registry elements into OMOP, reconciliation of cancer diagnoses derived from Cancer Registry and EHR

Pathology Report-NLP example



Legend: C1 = 1 bisected lymph node, C 2 = 1 lymph node, C3 = remaining tissue.

Part D. The specimen is received unfixed in a container labeled with the patient's name and "left non-sentinel node #1". It consists of one piece of pink-tan lymph node measuring 1.0 cm in greatest dimension. On sections, the tissue is pink. Submitted in toto in one cassette labeled D. fmq 7/12/2007 fmq

MICROSCOPIC DESCRIPTION

I. TYPE OF SPECIMEN: Left total mastectomy with sentinel axillary node biopsy

II. LOCATION OF THE TUMOR: Upper outer quadrant

III. TYPE OF NEOPLASM: Carcinoma, Invasive, Ductal - NOS Moderately Differentiated, Total score 6 (Tubule Score 2, Nuclear Grade Score 2, Mitotic Score 2) Ductal carcinoma in situ, nuclear grade 2, focal 5% Intraductal solid subtype Necrosis is present within the intraductal carcinoma Lobular neoplasia, type A (monomorphic), Focal

IV. GROSS/MICRO FINAL INVASIVE TUMOR SIZE INTERPRETATION: 1.0 x 0.8 x 0.7 cm.

V. BORDERS OF INVASIVE NEOPLASM: Ill-defined

VI. VASCULAR SPACE INVASION: Not identified

VII. CALCIFICATION: Absent

VIII. NIPPLE: Present, uninvolved by cancer

IX. SKIN: Present, uninvolved by cancer

X. ADJACENT BREAST TISSUE: Benign neoplasm: Hyalinized fibroadenomas

X. ADJACENT BREAST TISSUE: Cystic disease, proliferative

XI. MARGINS: Negative Tumor distance from closest margin deep DCIS &/or invasive: 1.1 cm

XII. AXILLARY LYMPH NODES: TOTAL: 4 SENTINEL NODE: 3

XIII. POSITIVE LYMPH NODES: TOTAL: 0 SENTINEL NODE: 0

XIV. PECTORAL MUSCLE: No pectoral muscle identified

XV. PATHOLOGIC STAGING (pTNM):Reflects staging only of the current specimen. Ultimate staging responsibility rests with the primary physician.

pT1c: Tumor more than 1.0 cm but not more than 2.0 cm in greatest dimension pN0: No regional lymph node metastasis on H & E histologically. pMX: Cannot be assessed

ADDITIONAL COMMENTS: No cancer cells are identified in the sentinel lymph nodes with the immunoperoxidase stain for pan-cytokeratin. An E-cadherin immunoperoxidase stain confirms the presence of lobular neoplasia extending into breast ducts.

RECEPTOR PROFILE:

Test Performed on formalin fixed paraffin embedded section of:

The results are for invasive carcinoma.

Specimen part "B":

Slide "B7":

Results:

Approximately 90 % of the carcinoma cell nuclei stain with an immunohistochemical stain utilizing an anti-estrogen receptor antibody (Dako 1D5; mouse polymer) with an average 3+ intensity.

Therefore, this tumor is considered positive for estrogen receptor expression (>5 % of the cells are positive).

No or rare carcinoma cell nuclei stain with an immunohistochemical stain utilizing an anti-progesterone receptor antibody (Dako PGR636; mouse polymer). Therefore, this tumor is considered negative for progesterone receptor expression.

Her-2/neu overexpression has been evaluated, on formalin fixed paraffin embedded sections, using the DAKO (K5207) HercepTest (proprietary kit). HerceptTest score: 3+. Her-2/neu overexpression is identified in the invasive carcinoma cells.

DIAGNOSIS(ES)

A. Skin, right breast, excision: - Skin, portion of, histologically unremarkable.

B. Breast, left, total mastectomy: - Carcinoma, invasive ductal type, moderately differentiated, Nottingham score 6 (2+2+2). - Carcinoma, intraductal, solid type, nuclear grade 2, with necrosis. - Lobular neoplasia, focal, extending into breast ducts. - Fibroadenomas, microscopic, hyalinized (2). - Fibrocystic disease, mildly proliferative with focal apocrine metaplasia.

C. Lymph nodes, left axilla sentinel nodes, biopsy: - No evidence of carcinoma in 3 lymph nodes.

D. Lymph node, left non-sentinel, biopsy: - No evidence of carcinoma in 1 lymph node.

Other Ongoing OHDSI/Federal Projects

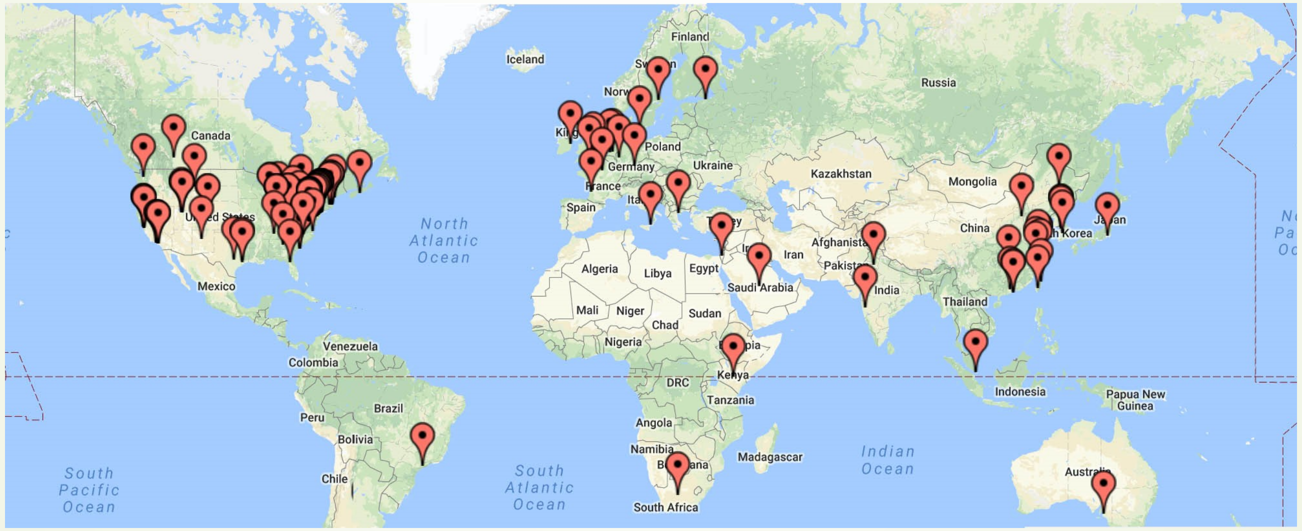
- FDA (Food and Drug Administration) BEST (Biologics Effectiveness and Safety) Award for biologics with CBER --Sentinel Initiative Center for Biologics Evaluation and Research
  - BEST Initiative: Blood and Blood Product Safety Surveillance
  - BEST Initiative: Develop New, Innovative Methods for Automation of Blood Product Adverse Event Reporting
- Dr. Hripcsak's R01 from NLM: Discovering and applying knowledge in clinical databases, LM006910--includes OHDSI supplement for vocabulary evaluation
- All of Us Research Program uses OMOP data model. Implemented by its Data and Research Center, Columbia is subcontractor (grant U2COD023196); From NHGRI.
- The eMERGE Consortium, funded by NHGRI, uses OMOP and provided a supplement to each of its 10 sites to implement OHDSI. (Grant U01HG008680)

Summary

- OHDSI/OMOP CDM can effectively identify patients with cancer as well as patients with specific diagnoses of cancer using SNOMED
- Identify receipt of drug therapies with very high accuracy using basic RxNorm phenotypes
- Procedures and timing, while already accurate, can be further enhanced by additions to the phenotype or curated code lists
- Observational data represents opportunity to obtain more complete and longitudinal view of patients with cancer (vs cancer-centric view of registries)

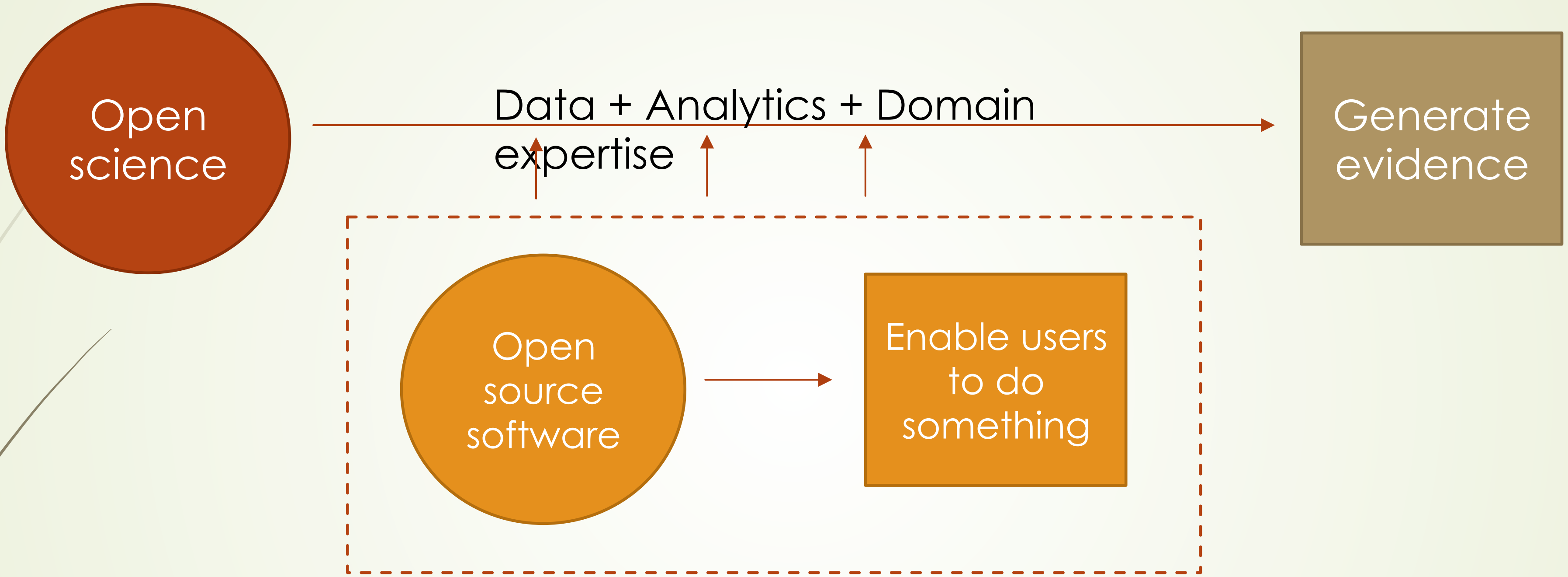
Observational Health Data Sciences and Informatics (OHDSI.org)

Mission: To improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care



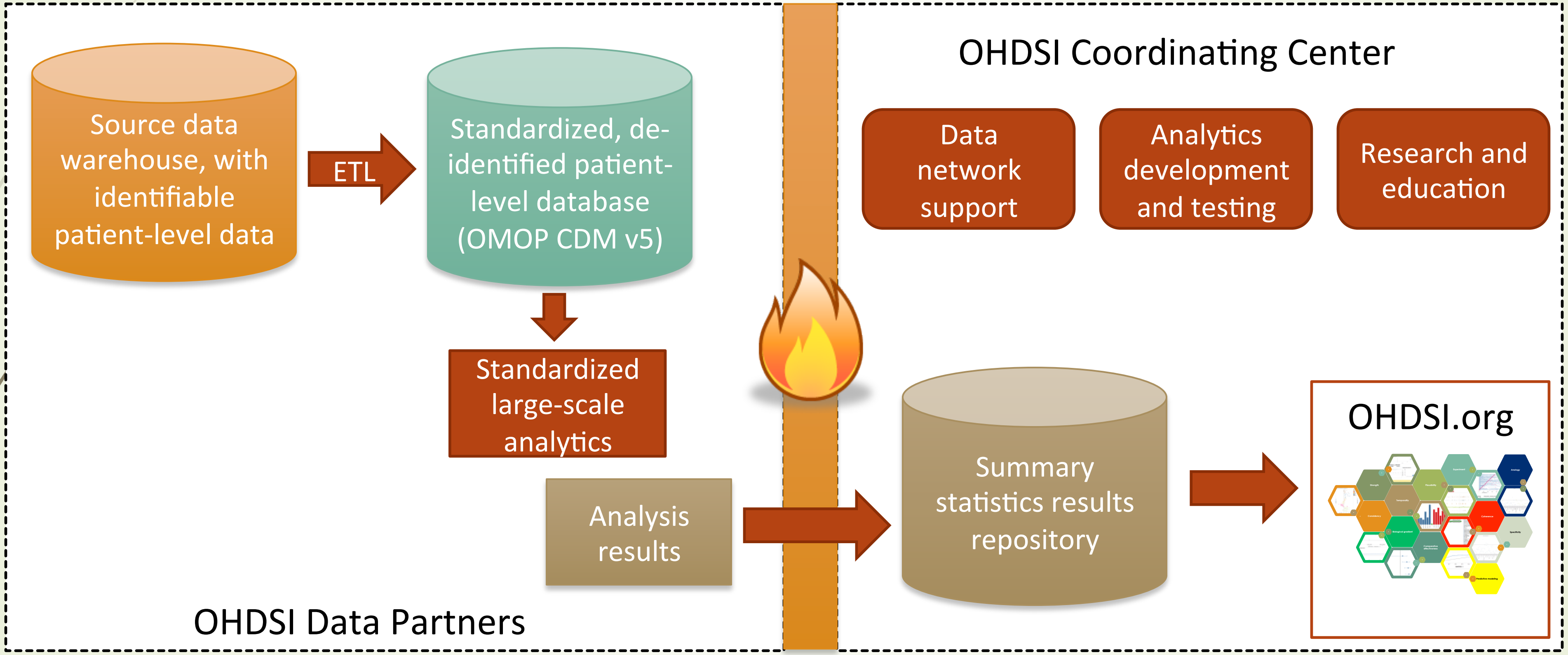
- >200 collaborators from 25 different countries
- Experts in informatics, statistics, epidemiology, clinical sciences
- Active participation from academia, government, industry, providers
- Over a billion records on >400 million patients in 80 databases

OHDSI’s approach to open science


Open science
Data + Analytics + Domain expertise
Generate evidence
Open source software
Enable users to do something

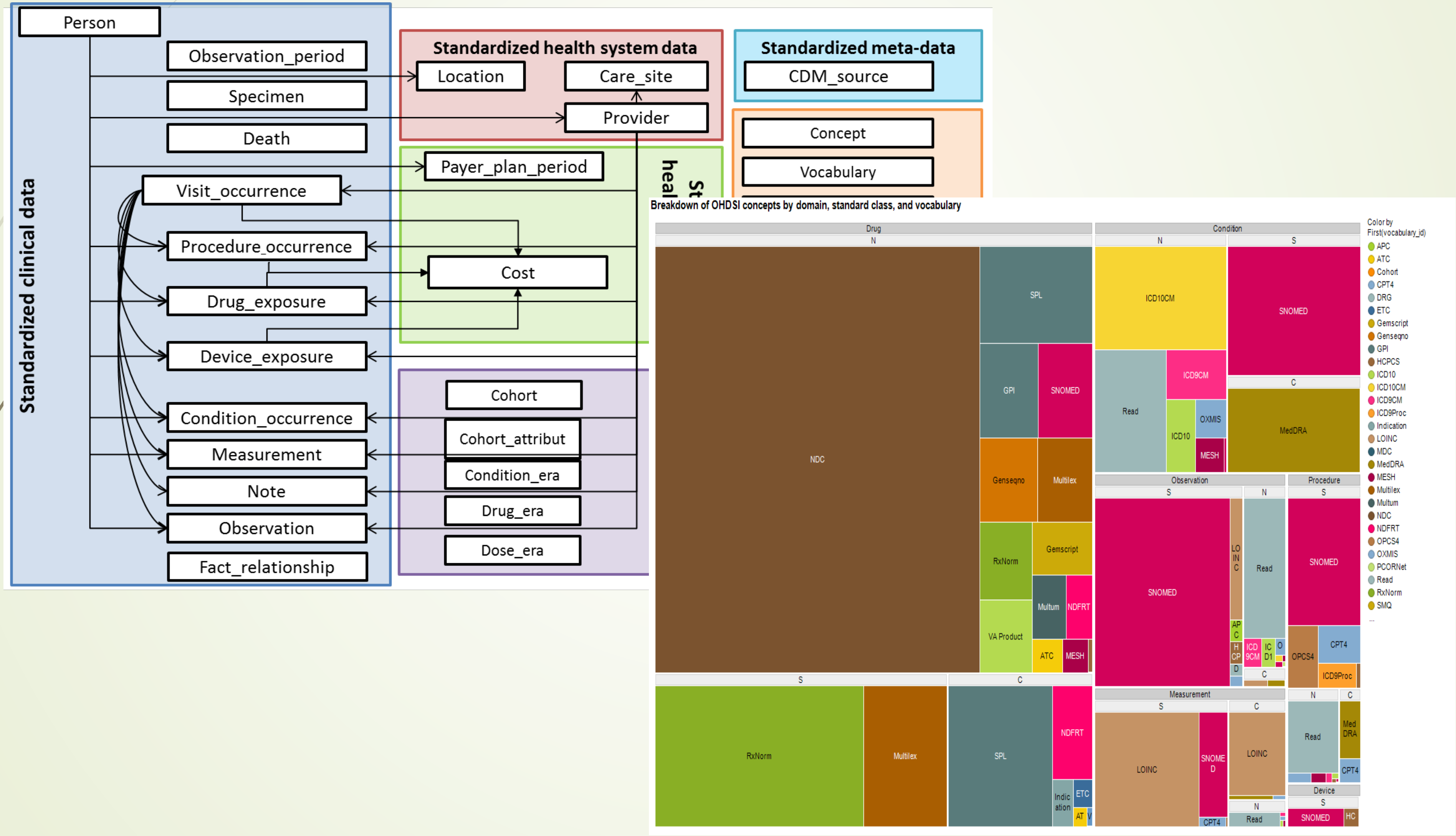
- Open science is about sharing the journey to evidence generation
- Open-source software can be part of the journey, but it’s not a final destination
- Open processes can enhance the journey through improved reproducibility of research and expanded adoption of scientific best practices

How OHDSI works:
Data stay local, total open science

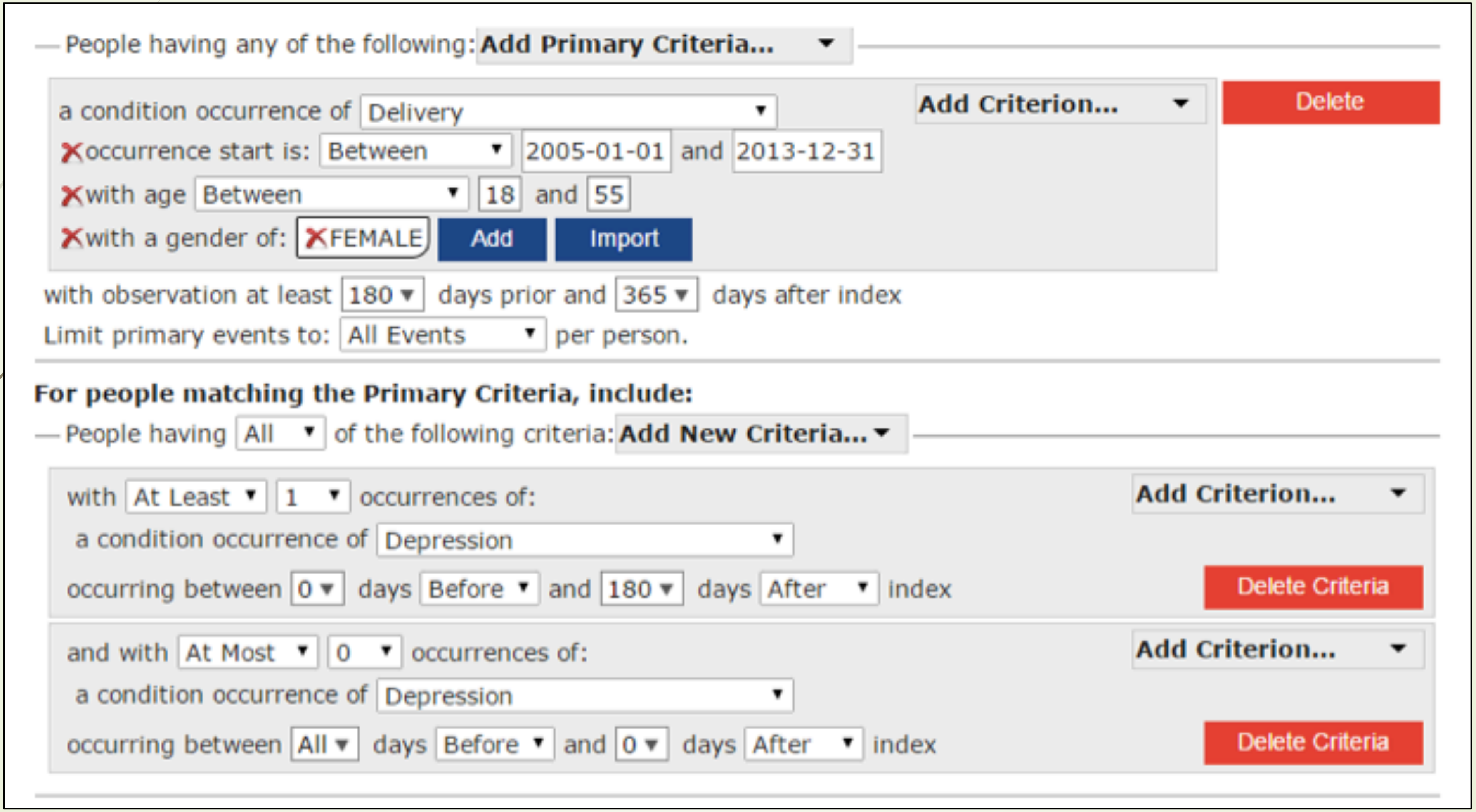


OHDSI Coordinating Center
Source data warehouse, with identifiable patient-level data
ETL
Standardized, de-identified patient-level database (OMOP CDM v5)
Standardized large-scale analytics
Analysis results
Data network support
Analytics development and testing
Research and education
Summary statistics results repository
OHDSI.org
OHDSI Data Partners

OHDSI OMOP CDM: Deep information model with extensive vocabularies (80)



ATLAS to build, visualize, and analyze cohorts



People having any of the following: Add Primary Criteria...

a condition occurrence of Delivery

occurrence start is: Between 2005-01-01 and 2013-12-31

with age Between 18 and 55

with a gender of: FEMALE

with observation at least 180 days prior and 365 days after index

Limit primary events to: All Events per person.

For people matching the Primary Criteria, include:

People having All of the following criteria: Add New Criteria...

with At Least 1 occurrences of:

a condition occurrence of Depression

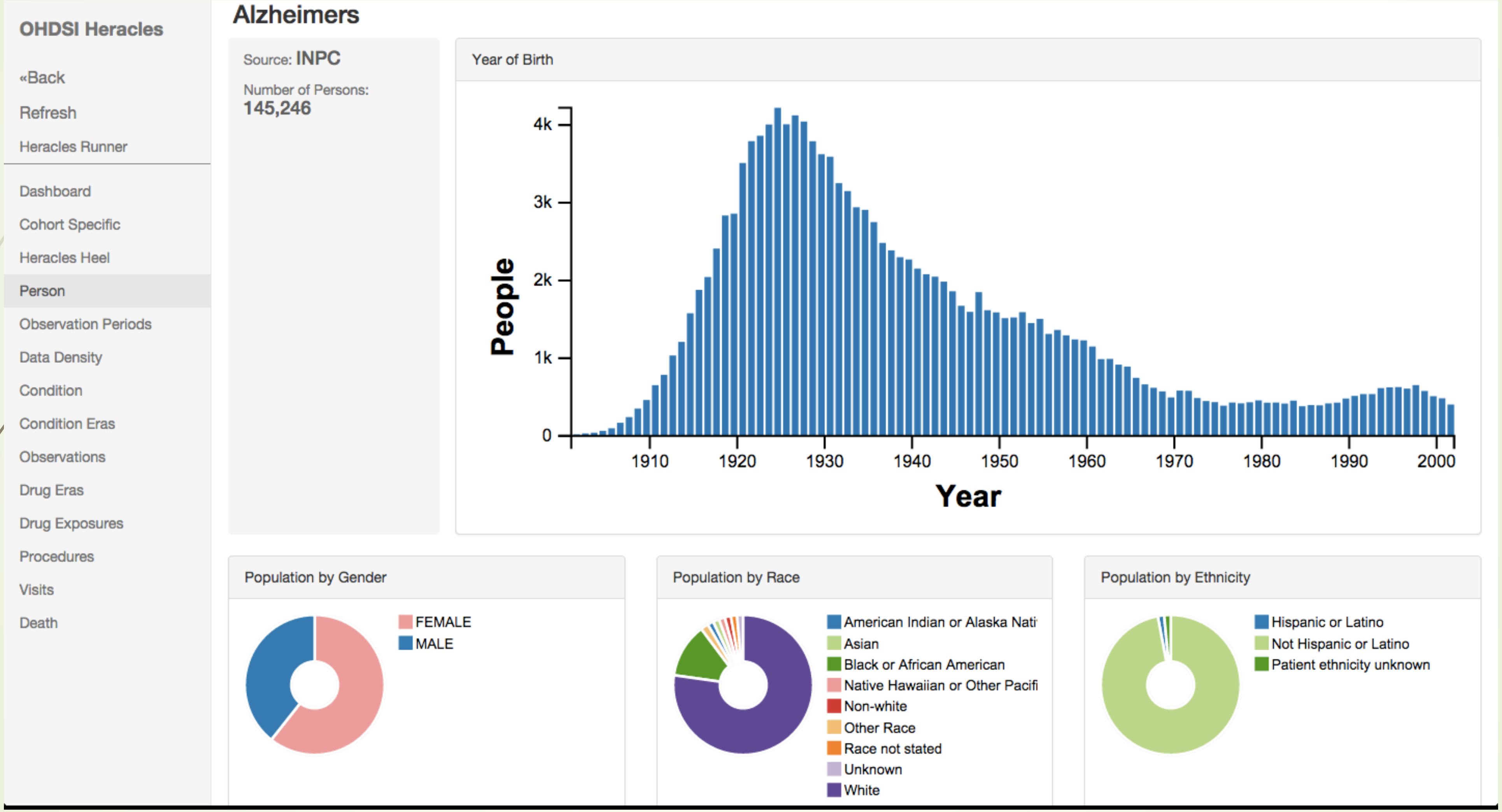
occurring between 0 days Before and 180 days After index

and with At Most 0 occurrences of:

a condition occurrence of Depression

occurring between All days Before and 0 days After index

ATLAS to build, visualize, and analyze cohorts

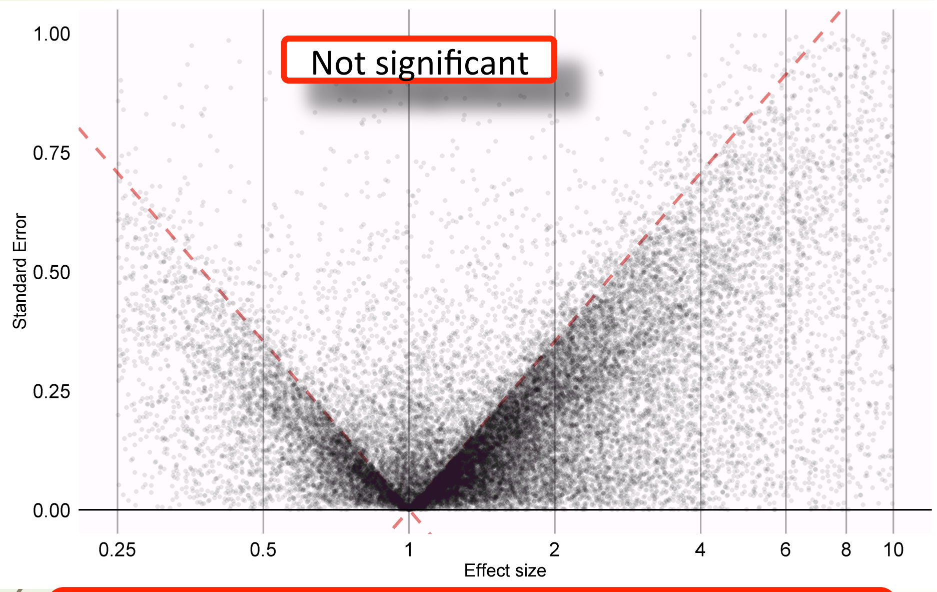


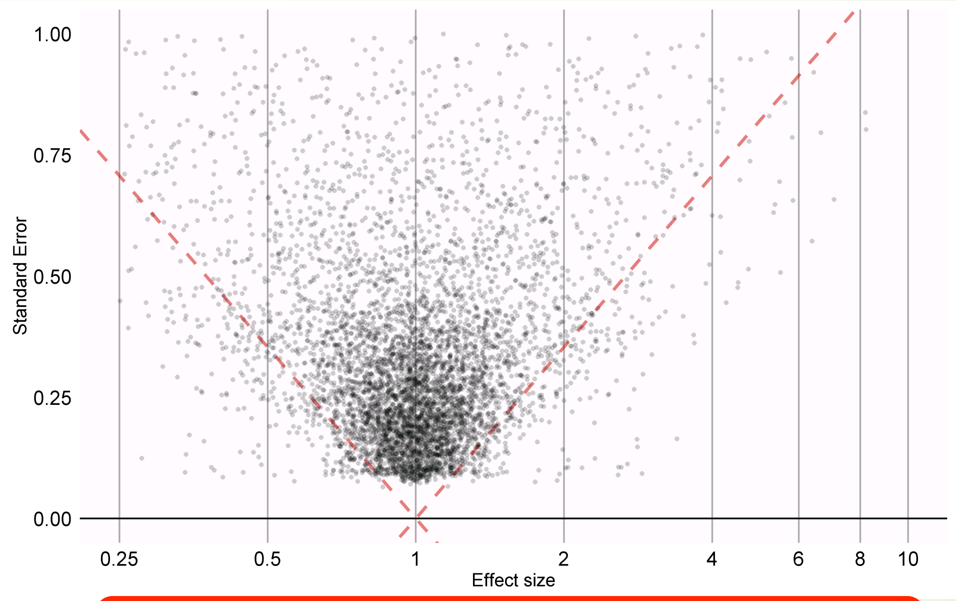
Alzheimers

Source: INPC

Number of Persons: 145,246

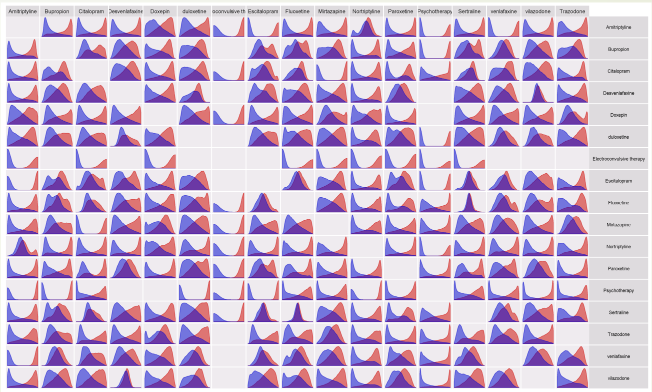
Improving reproducibility through large scale research





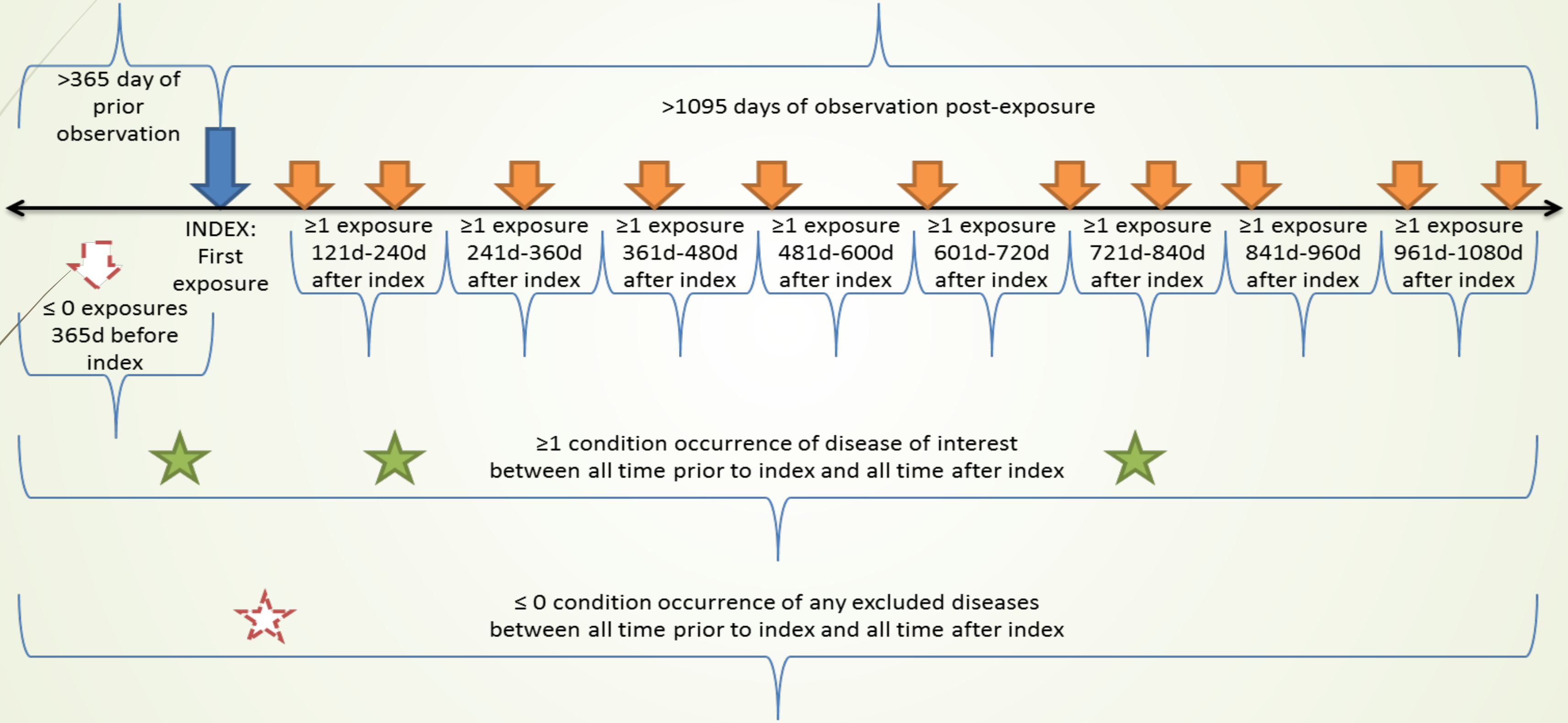
Literature is severely biased and 85% positive with p-value hacking

OHDSI 11% of exposure-outcome pairs have calibrated $p < 0.05$



Schuemie, Phil Trans A 2018

Treatment pathway event flow



>365 day of prior observation

>1095 days of observation post-exposure

INDEX: First exposure

≤ 0 exposures 365d before index

≥1 exposure 121d-240d after index

≥1 exposure 241d-360d after index

≥1 exposure 361d-480d after index

≥1 exposure 481d-600d after index

≥1 exposure 601d-720d after index

≥1 exposure 721d-840d after index

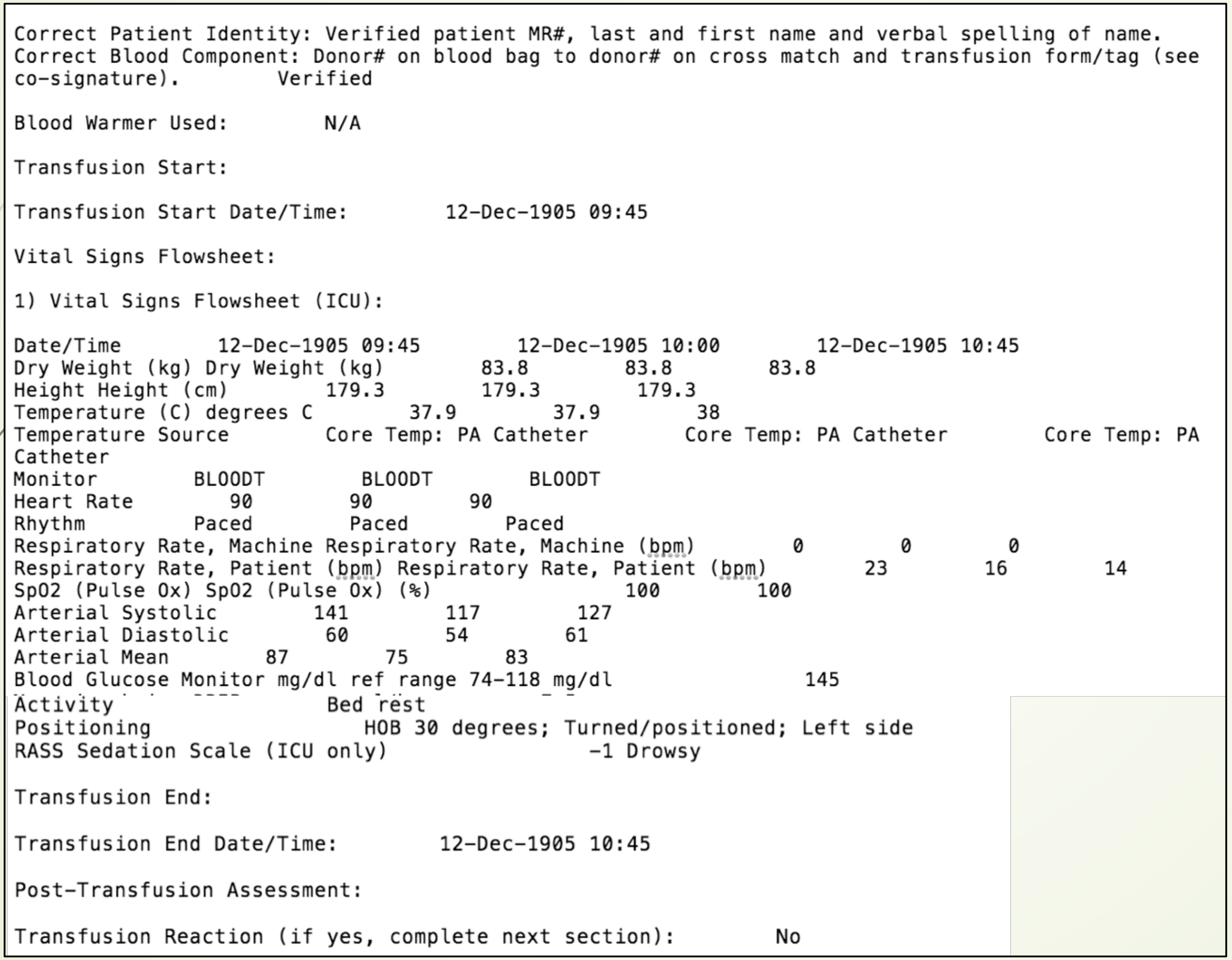
≥1 exposure 841d-960d after index

≥1 exposure 961d-1080d after index

≥1 condition occurrence of disease of interest between all time prior to index and all time after index

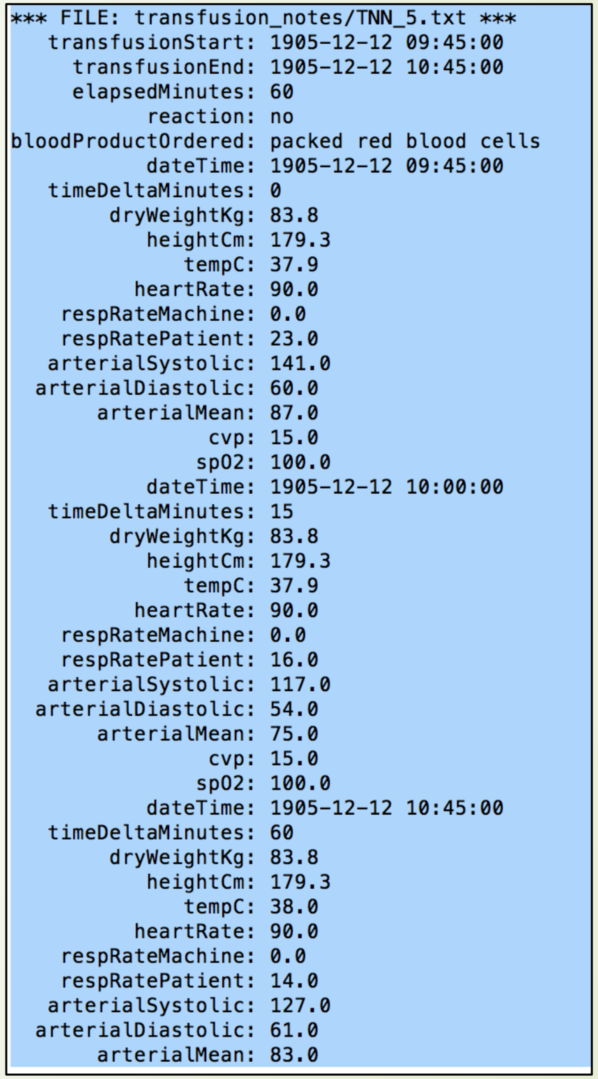
≤ 0 condition occurrence of any excluded diseases between all time prior to index and all time after index

FDA BEST-NLP Example



```
Correct Patient Identity: Verified patient MR#, last and first name and verbal spelling of name.
Correct Blood Component: Donor# on blood bag to donor# on cross match and transfusion form/tag (see
co-signature).        Verified

Blood Warmer Used:        N/A

Transfusion Start:

Transfusion Start Date/Time:        12-Dec-1905 09:45

Vital Signs Flowsheet:

1) Vital Signs Flowsheet (ICU):

Date/Time        12-Dec-1905 09:45        12-Dec-1905 10:00        12-Dec-1905 10:45
Dry Weight (kg) Dry Weight (kg)        83.8        83.8        83.8
Height Height (cm)        179.3        179.3        179.3
Temperature (C) degrees C        37.9        37.9        38
Temperature Source        Core Temp: PA Catheter        Core Temp: PA Catheter        Core Temp: PA
Catheter
Monitor        BLOODT        BLOODT        BLOODT
Heart Rate        90        90        90
Rhythm        Paced        Paced        Paced
Respiratory Rate, Machine Respiratory Rate, Machine (bpm)        0        0        0
Respiratory Rate, Patient (bpm) Respiratory Rate, Patient (bpm)        23        16        14
SpO2 (Pulse Ox) SpO2 (Pulse Ox) (%)        100        100
Arterial Systolic        141        117        127
Arterial Diastolic        60        54        61
Arterial Mean        87        75        83
Blood Glucose Monitor mg/dl ref range 74-118 mg/dl        145
Activity        Bed rest
Positioning        HOB 30 degrees; Turned/positioned; Left side
RASS Sedation Scale (ICU only)        -1 Drowsy

Transfusion End:

Transfusion End Date/Time:        12-Dec-1905 10:45

Post-Transfusion Assessment:

Transfusion Reaction (if yes, complete next section):        No
```

Source Transfusion Nursing Note



```
*** FILE: transfusion_notes/TNN_5.txt ***
   transfusionStart: 1905-12-12 09:45:00
     transfusionEnd: 1905-12-12 10:45:00
     elapsedMinutes: 60
           reaction: no
bloodProductOrdered: packed red blood cells
           dateTime: 1905-12-12 09:45:00
   timeDeltaMinutes: 0
        dryWeightKg: 83.8
           heightCm: 179.3
              tempC: 37.9
          heartRate: 90.0
     respRateMachine: 0.0
     respRatePatient: 23.0
   arterialSystolic: 141.0
  arterialDiastolic: 60.0
       arterialMean: 87.0
                cvp: 15.0
               sp02: 100.0
           dateTime: 1905-12-12 10:00:00
   timeDeltaMinutes: 15
        dryWeightKg: 83.8
           heightCm: 179.3
              tempC: 37.9
          heartRate: 90.0
     respRateMachine: 0.0
     respRatePatient: 16.0
   arterialSystolic: 117.0
  arterialDiastolic: 54.0
       arterialMean: 75.0
                cvp: 15.0
               sp02: 100.0
           dateTime: 1905-12-12 10:45:00
   timeDeltaMinutes: 60
        dryWeightKg: 83.8
           heightCm: 179.3
              tempC: 38.0
          heartRate: 90.0
     respRateMachine: 0.0
     respRatePatient: 14.0
   arterialSystolic: 127.0
  arterialDiastolic: 61.0
       arterialMean: 83.0
```

Extracted Structured Data