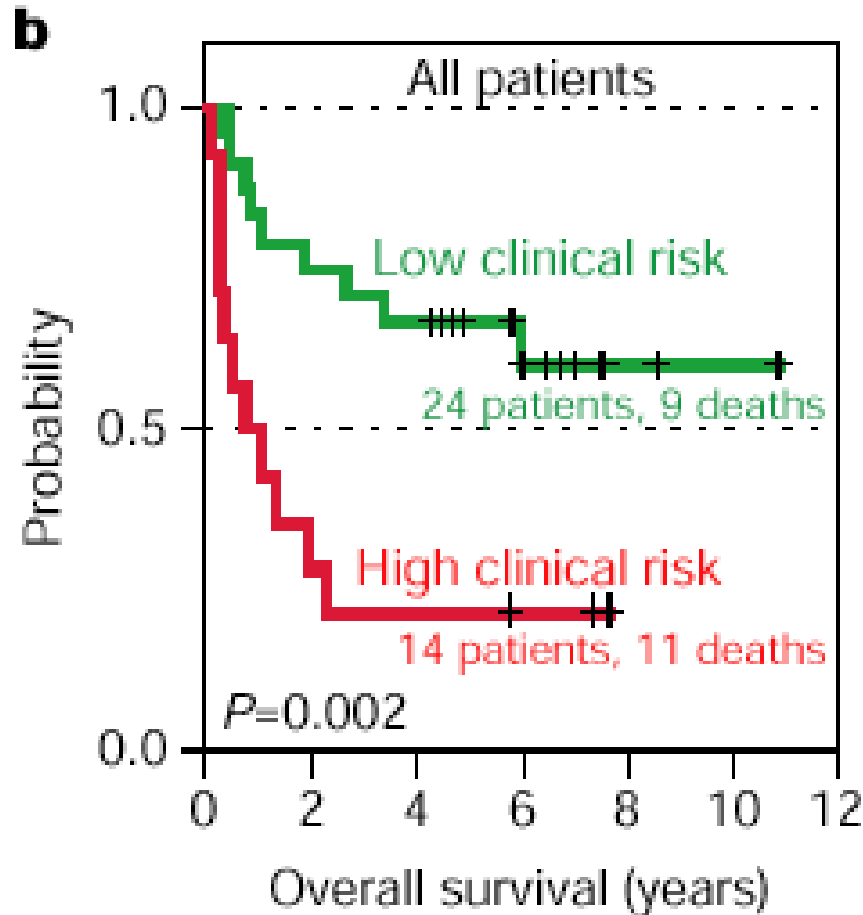


I2b2 - 10 Strategic Lessons in Instrumenting the Health Care Enterprise for Discovery Research in the Genomic Era

Shawn Murphy MD, Ph.D.

The paper that Launched at 100,000 chips: Alizadeh 2000



Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling

Ash A. Alizadeh^{1,2}, Michael B. Eisen^{2,3,4}, R. Eric Davis⁵, Chi Ma⁵, Izidore S. Lossos⁶, Andreas Rosenwald⁵, Jennifer C. Boldrick¹, Hajeer Sabet⁵, Truc Tran⁵, Xin Yu⁵, John I. Powell⁷, Liming Yang⁷, Gerald E. Marti⁸, Troy Moore⁹, James Hudson Jr.⁹, Lisheng Lu¹⁰, David B. Lewis¹⁰, Robert Tibshirani¹¹, Gavin Sherlock⁴, Wing C. Chan¹², Timothy C. Greiner¹², Dennis D. Weisenburger¹², James O. Armitage¹³, Roger Wamke¹⁴, Ronald Levy⁶, Wyndham Wilson¹⁵, Michael R. Grever¹⁶, John C. Byrd¹⁷, David Botstein⁴, Patrick O. Brown^{1,18} & Louis M. Staudt⁵

Departments of ¹Biochemistry, ³Genetics, ¹⁴Pathology, ⁶Medicine, ¹⁰Pediatrics and ¹¹Health Research & Policy and Statistics, and ¹⁸Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, California 94305, USA

⁵Metabolism Branch, Division of Clinical Sciences, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

⁷Bioinformatics and Molecular Analysis Section, CBEL, CIT, NIH, Bethesda, Maryland 20892, USA

⁸CBER, FDA, Bethesda, Maryland 20892, USA

⁹Research Genetics, Huntsville, Alabama 35801, USA

Departments of ¹²Pathology and Microbiology, and ¹³Internal Medicine, University of Nebraska Medical Center, Omaha, Nebraska 68198, USA

¹⁵Medicine Branch, Division of Clinical Sciences, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

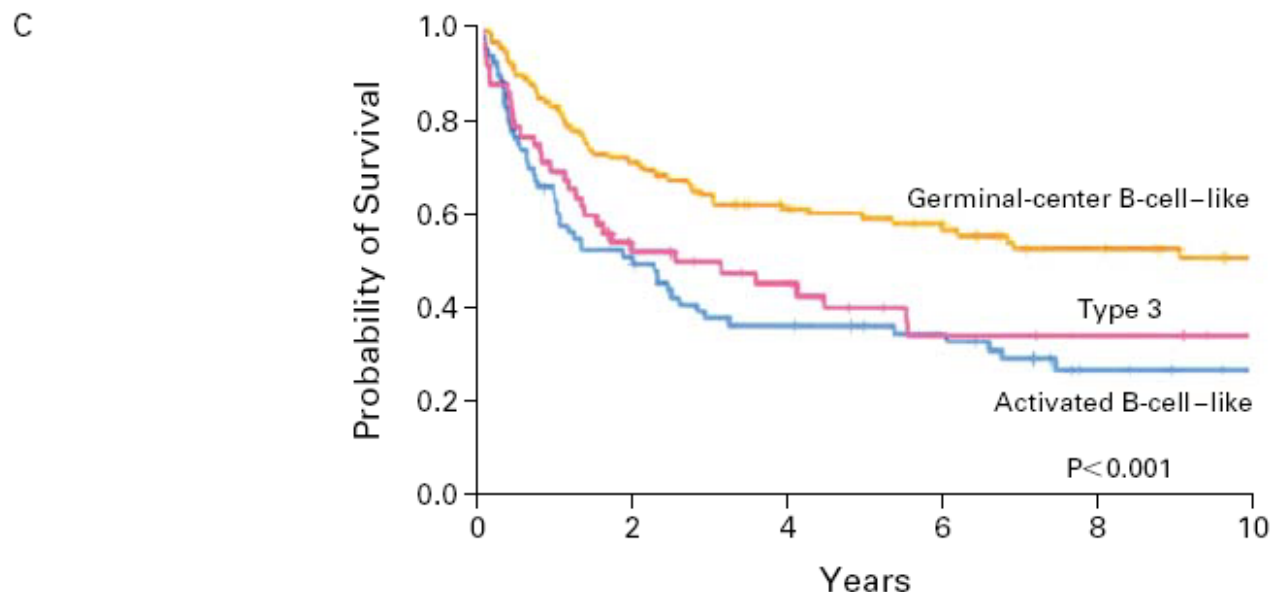
¹⁶Johns Hopkins Oncology Center, Johns Hopkins School of Medicine, Baltimore, Maryland 21287, USA

¹⁷Walter Reed Army Medical Center, Washington, DC 20307, USA

²These authors contributed equally to this work

Alizadeh et al., Nature 2000

Example: Rosenwald



No. AT RISK

Germinal-center B-cell-like	115	81	60	46	32	19
Type 3	52	24	18	10	8	5
Activated B-cell-like	73	35	23			

The New England
Journal of Medicine

Copyright © 2002 by the Massachusetts Medical Society

VOLUME 346

JUNE 20, 2002

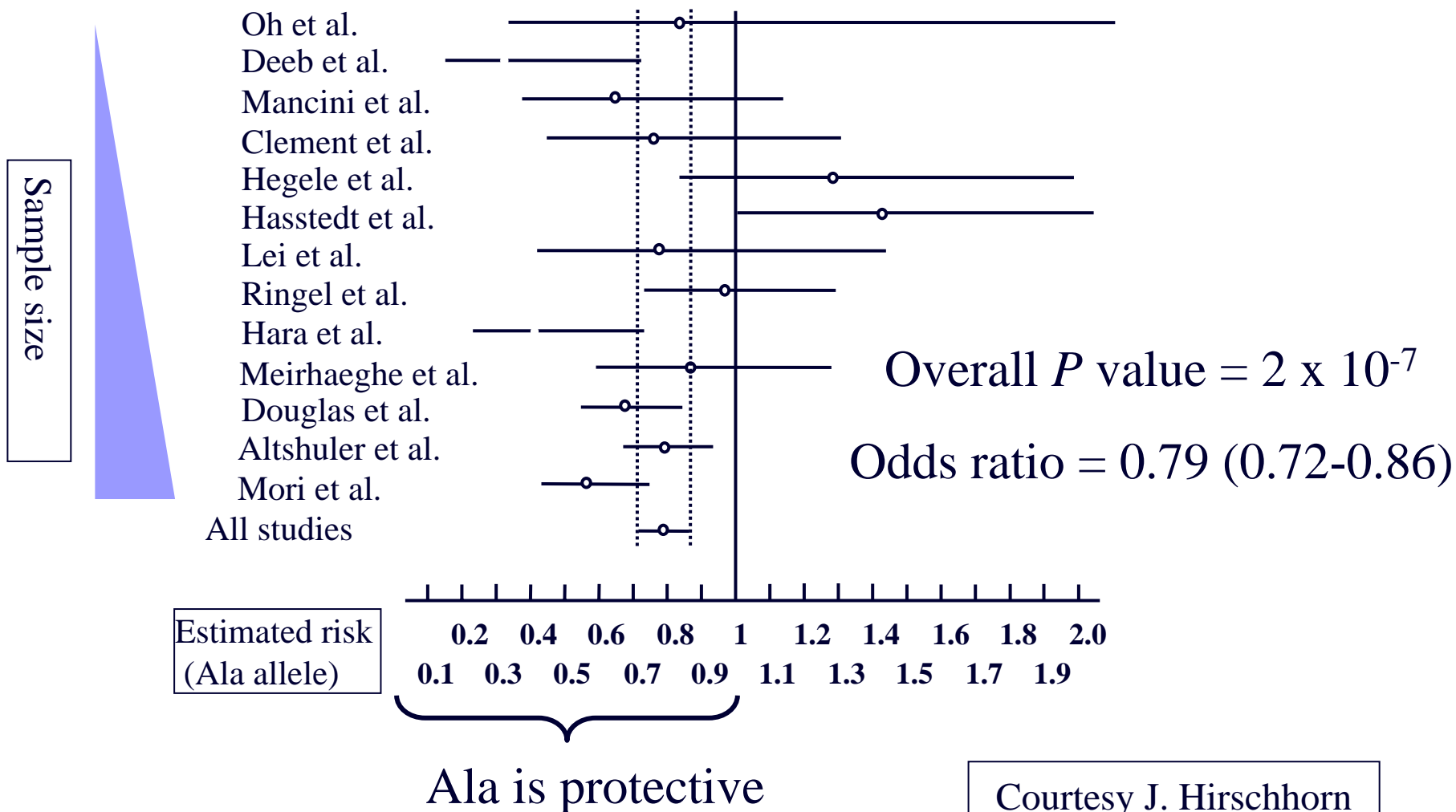
NUMBER 25



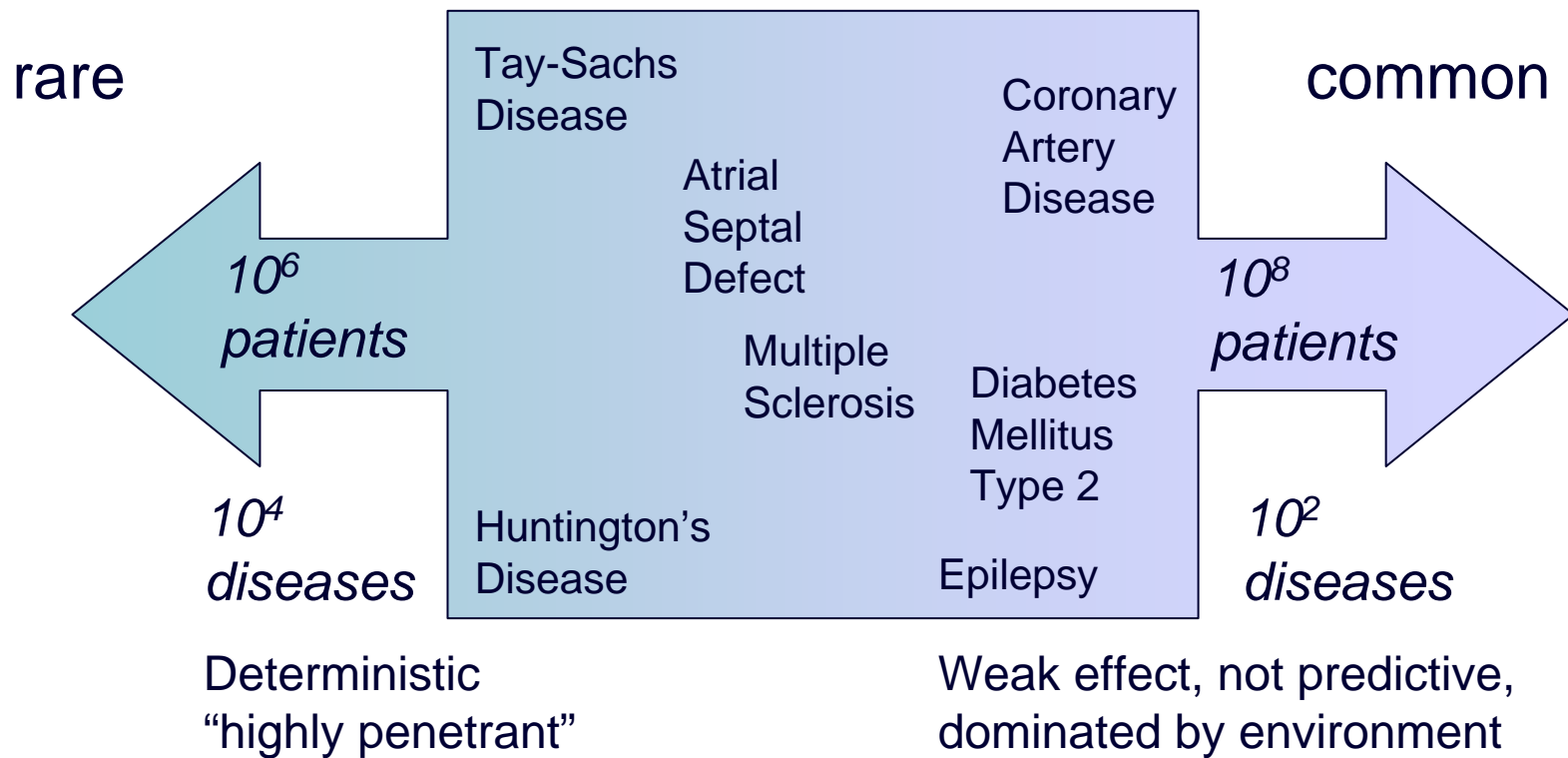
THE USE OF MOLECULAR PROFILING TO PREDICT SURVIVAL
AFTER CHEMOTHERAPY FOR DIFFUSE LARGE-B-CELL LYMPHOMA

ANDREAS ROSENWALD, M.D., GEORGE WRIGHT, Ph.D., WING C. CHAN, M.D., JOSEPH M. CONNORS, M.D.,
ELIAS CAMPO, M.D., RICHARD I. FISHER, M.D., RANDY D. GASCOYNE, M.D., H. KONRAD MÜLLER-HERMELINK, M.D.,
ERLEND B. SMELAND, M.D., Ph.D., AND LOUIS M. STAUDT, M.D., Ph.D.,
FOR THE LYMPHOMA/LEUKEMIA MOLECULAR PROFILING PROJECT

Example: PPAR γ Pro12Ala and Diabetes



Common-Rare: Weak-Strong Spectrum



The Power of Numbers: Efficiently Reaching a Large N

- High throughput genotyping
- High throughput phenotyping
- High throughput sample acquisition

DHHS Secretary's Advisory Committee on Genetics, Health, and Society (SACGHS) argues for the health value of a 500,000 to 1M subject study. Estimated cost: \$3,000,000,000

Cost of the pediatric 100,000 study recently launched >> \$1B + decades.

High Throughput Methods for supporting Research at Partners Healthcare

- Set of patients is selected from medical record data in a high throughput fashion
- Investigators work with the data of these patients using new i2b2 tools and a specialized team, both developed to work specifically with medical record data
- Using the Crimson system, tissues of these patients can be made available for genomic and biochemical analysis
- Automated discovery can be created from these projects to support further hypothesis-driven research

High Throughput Methods for supporting Research at Partners Healthcare

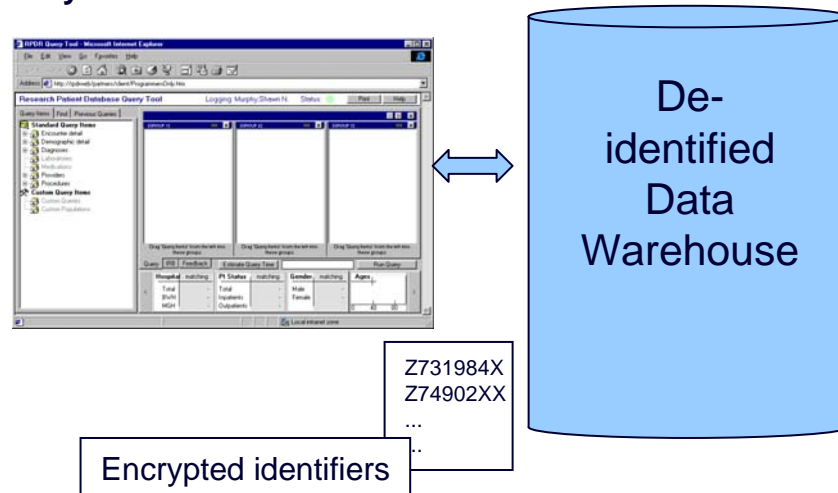
- Set of patients is selected from medical record data in a high throughput fashion
 - Partners Research Computing – Eugene Brunwald, John Glaser, Diane Keogh
- Investigators work with the data of these patients using new i2b2 tools and a specialized team, both developed to work specifically with medical record data
- Using the Crimson system, tissues of these patients can be made available for genomic and biochemical analysis
- Automated discovery can be created from these projects to support further hypothesis-driven research

Current Operations of the Research Patient Data Registry

Queries for aggregate patient numbers

- Warehouse of in & outpatient clinical data
- 4.6 million Partners Healthcare patients
- 1.2 billion diagnoses, medications, procedures, laboratories, physical findings, & genomics coupled to demographics & visits
- Authorized use by faculty status
- Clinicians can construct complex queries
- Queries cannot identify individuals, internally can produce identifiers for detailed data sets

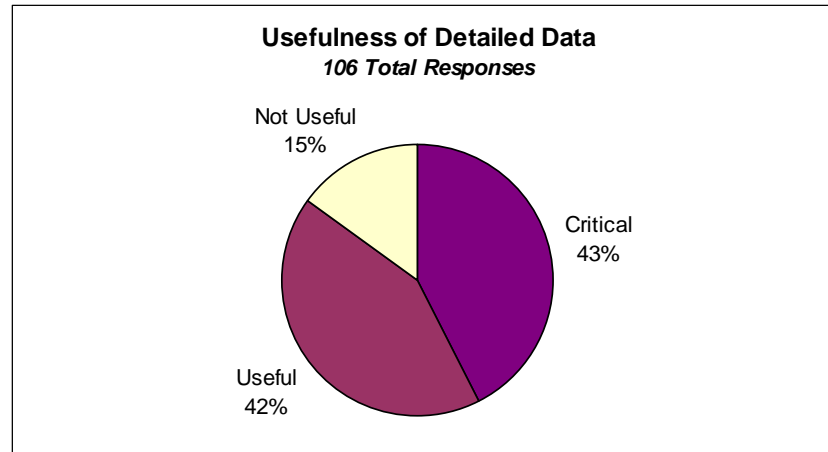
Query construction in web tool



1,970 registered users, 332 new in 2008

915 detailed data sets = research study returned to these teams, containing data for of 8.8 million patient records.

\$94-136 million total research support critically dependent on RPDR from patient data received throughout life of funding.



Enterprise web client

<http://services.i2b2.org/webclient/>

i2b2 Web Client - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites RSS Feeds Check for updates

Address <http://services.i2b2.org/webclient/#> Go

Google RS 113 blocked ABC Check Settings **lenovo**

i2b2 Query & Analysis Tool Find Patients | Analysis Tools | Message Log | Help | Logout

Navigate Terms Find Terms

Search by Names Search by Codes

contains multiple sclerosis

Find Any Category

Multiple sclerosis

Workplace

SHARED

demo

Previous Queries

- Moderat-Multipl@02:14:47 [2-24-2009] [demo]
- Moderat-Multipl@02:14:24 [2-24-2009] [demo]
- BetaBlockerStudy [2-24-2009] [demo]
- Coronal@21:39:04 [2-24-2009] [demo]
- Age-Circu-WBC (@21:36:01 [2-24-2009] [demo]
- Age-Circu-WBC (@21:32:35 [2-24-2009] [demo]

Query Tool

Query Name: Moderat-Multipl@02:14:47

Group 1			Group 2			Group 3		
Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude
Moderate Alzheimers dise			Multiple sclerosis					

none of these AND one or more of these AND drop a term on here

Run Query New Query 2 Groups New Group

Query Status

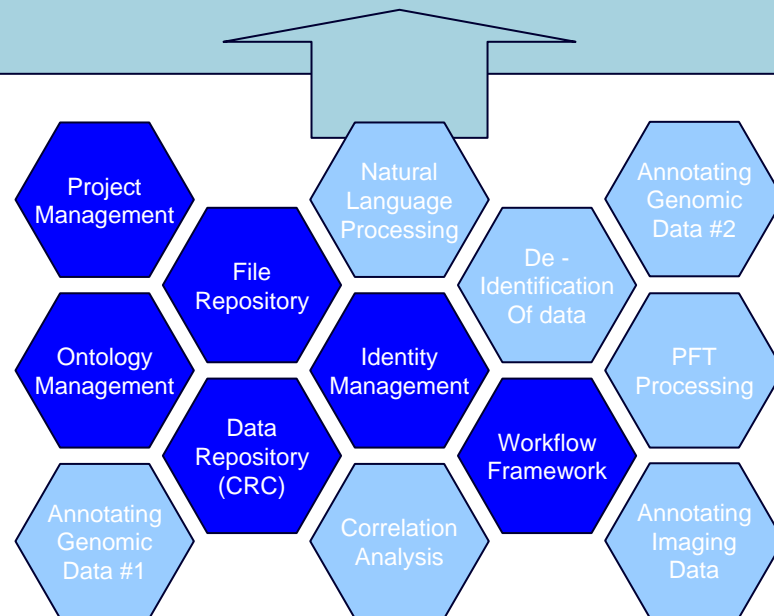
Executing query...

Elapsed time (seconds): 2.8

Query Finished...

Matching patients: 462

Enterprise-wide repurposing and distribution of medical record data for research

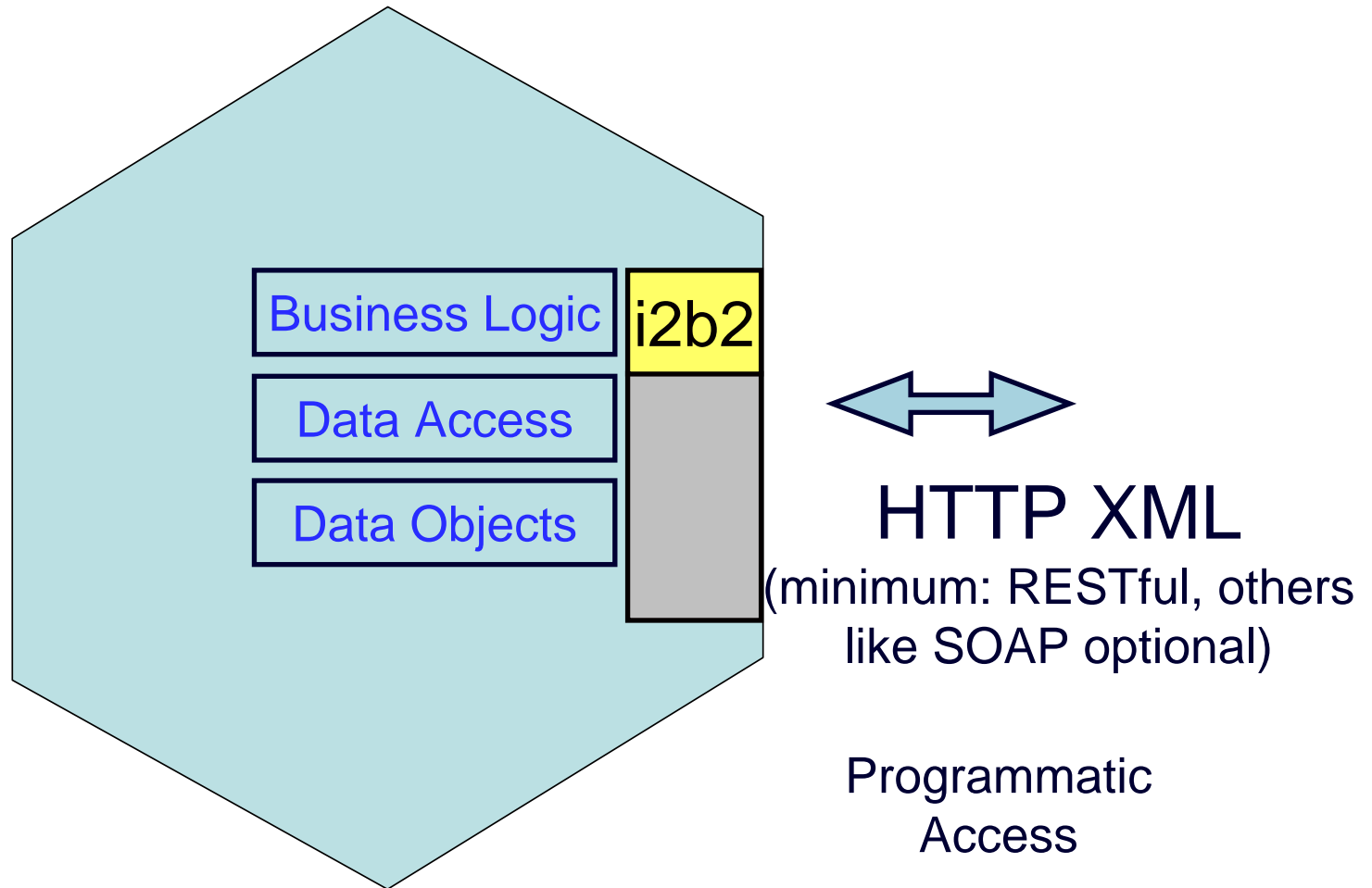


- Enable high performance collection of medical record data for querying and distribution
 - Enterprise web client
 - Create patient cohorts for further investigation
- Enable discovery within data on enterprise wide scale
 - Relevance networks
 - Pharmacovigilance

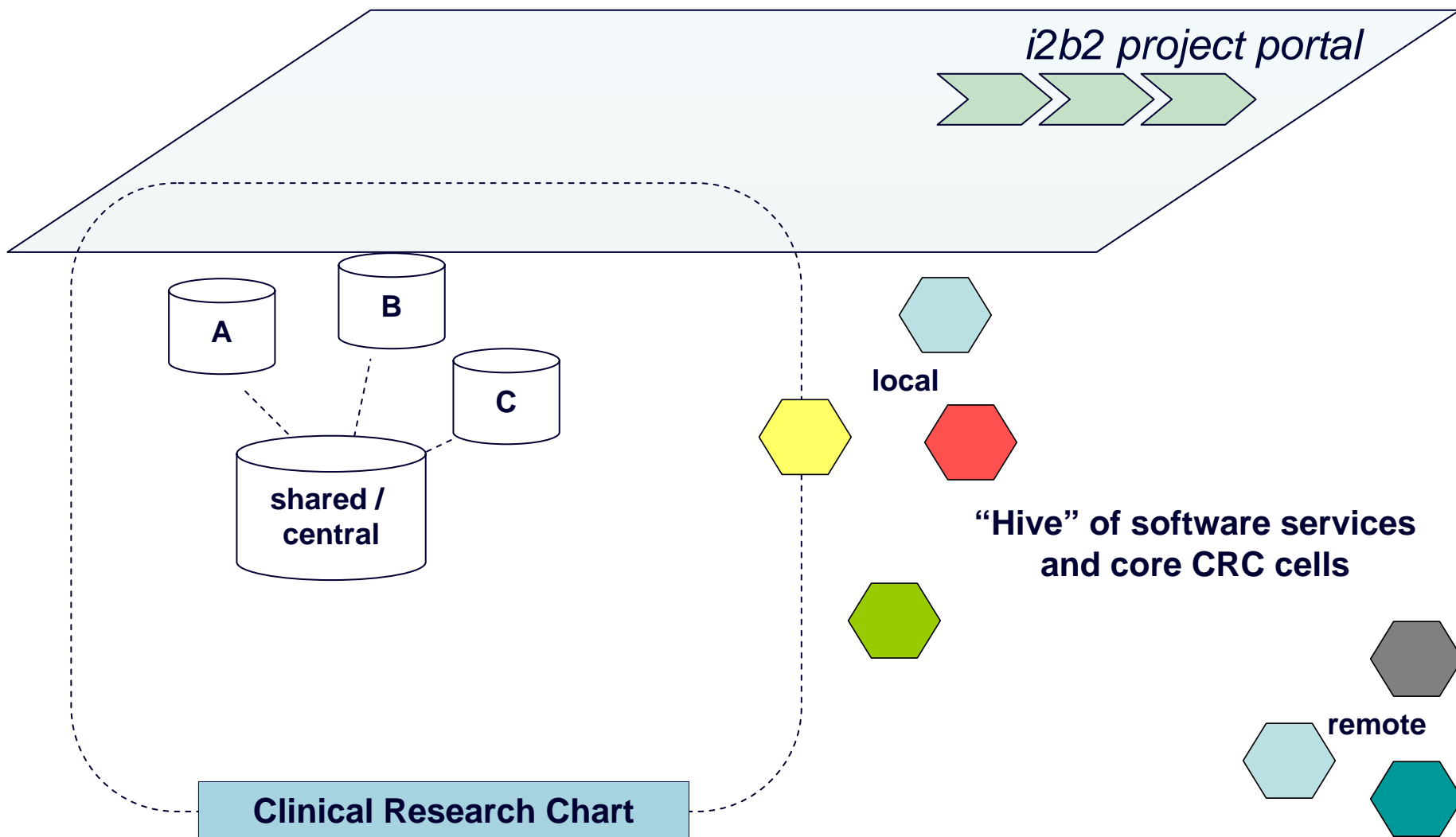
Technical Overview

- Formed as a collection of interoperable services provided by i2b2 Cells
- Loosely coupled
- Makes no assumptions about proximity
- Connected by Web services
- Activity can be directed manually or automatically

i2b2 Cell: Canonical Hive Unit



i2b2 Environment





-  CTSAs* Adopting i2b2
-  CTSAs* Evaluating i2b2 platform
-  Academic Medical Centers Adopting i2b2 Platform
-  Foreign Medical Centers Adopting i2b2 Platform

Community

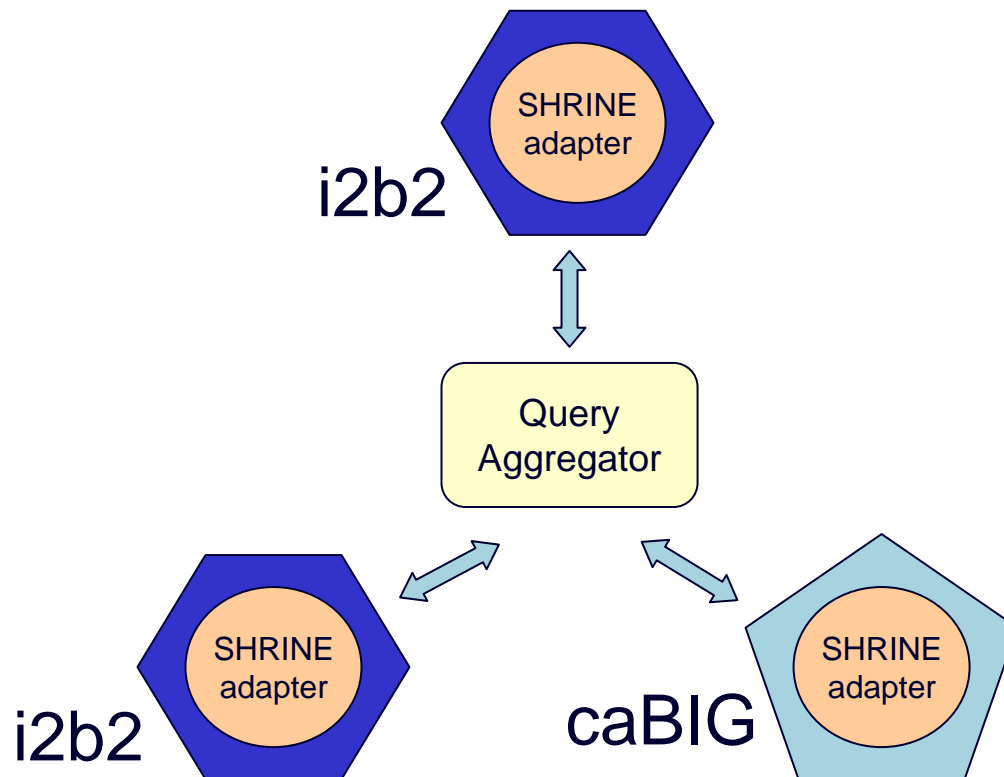
United States

- Beth Israel Deaconess Hospital, Boston, MA
- Boston University School of Medicine, Boston, MA
- Brigham and Women's Hospital, Boston, MA
- Children's Hospital, Boston, MA
- Denver Children's Hospital, Denver, CO
- Cincinnati Children's Hospital, Cincinnati, OH
- Cleveland Clinic, Cleveland, OH
- Weil Medical College of Cornell, NYC, NY
- Group Health Cooperative
- Harvard Medical School, Boston, MA
- Massachusetts General Hospital, Boston, MA
- Maine Medical Center, Portland, ME
- Marshfield Clinic, Wisconsin
- Morehouse School of Medicine, Atlanta, GA
- Oregon Health & Science University, Portland, OR
- Ohio State University Medical Center, Columbus, OH
- Philadelphia Children's Hospital, Philadelphia, PA
- Renaissance Computing Institute, Chapel Hill, NC
- Tufts New England Medical Center, Boston, MA
- University of California Davis, Davis, CA
- University of California San Francisco, SF, CA
- University of Massachusetts Medical School, Worcester, MA
- University of Michigan Medical Center, Ann Arbor, MI
- University of Pennsylvania School of Medicine, Philadelphia, PA
- University of Rochester Medical Center, Rochester, NY
- University of Texas Health Sciences Center Houston, Houston, TX
- University of Texas Health Sciences Center San Antonio, SA, TX
- University of Texas Health Sciences Center Southwestern,
- Utah Health Science Center, Salt Lake City, UT
- University of Washington, Seattle, WA

International

- Georges Pompidou Hospital, Paris, France
- University of Goettingen, Goettingen, Germany
- University of Pavia, Pavia, Italy
- University of Seoul, Seoul, Korea

SHRINE (Shared Research Informatics Network) = Distributed Queries



Central “aggregator” broadcasts query to local hospital “adaptors”, which return aggregate counts only

SHRINE

i2b2 Query & Analysis Tool

Logout

Navigate Terms

Find Terms

Demographics

Age

Gender

Female

Male

Unknown

Language

Marital-Status

Race

Vital-Status

Diagnoses

drag an item from here

Query Tool

Query Name:

Group 1

Dates

Occurs > 0x

Exclude

Regional enteritis

Ulcerative colitis

one or more of these

AND

Group 2

Dates

Occurs > 0x

Exclude

Male

none of these

AND

Group 3

Dates

Occurs > 0x

Exclude

Neoplasms

one or more of these

Run Query

New Query

3 Groups

New Group

Previous Queries

Noninfe-Female@16:39:03

Nonin-Femal-Neopl@16:39:11

Circulatory sys@16:39:18

Circulatory sys@17:01:03

Circula-Events @17:04:11

Demographics@17:04:21

10-19 years old@17:04:39

Albanian@02:31:25

Query Status

Executing query...

Elapsed time (seconds): 14.0

Query Finished...

Matching patients (hospital 1): 332 (+/-3)

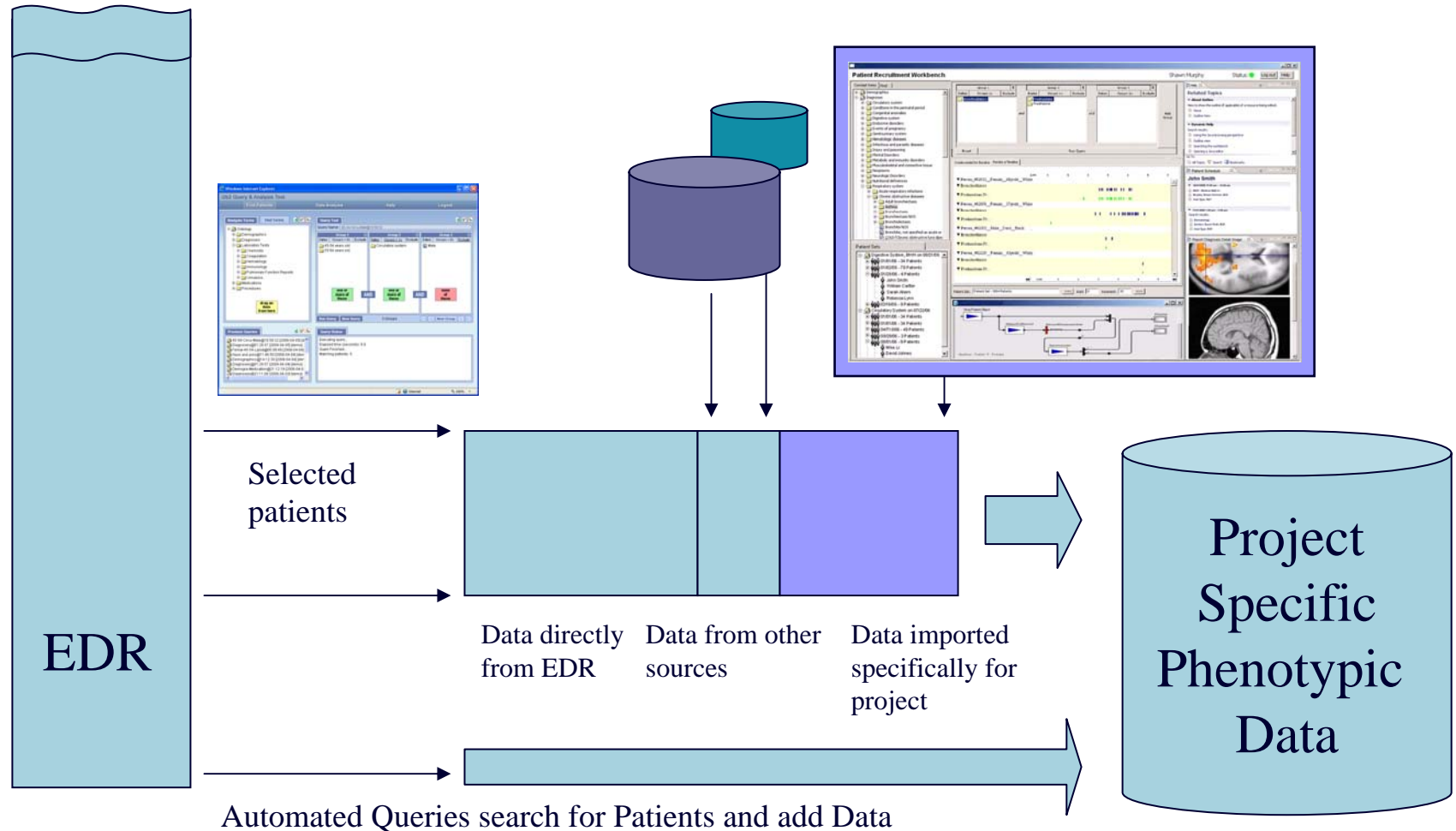
Matching patients (hospital 2): 16 (+/-3)

Matching patients (hospital 3): 151 (+/-3)

High Throughput Methods for supporting Research at Partners Healthcare

- Set of patients is selected from medical record data in a high throughput fashion
- Investigators work with the data of these patients using new i2b2 tools and a specialized team, both developed to work specifically with medical record data
 - NIH/NCBC - Isaac Kohane
- Using the Crimson system, tissues of these patients can be made available for genomic and biochemical analysis
- Automated discovery can be created from these projects to support further hypothesis-driven research

Set of patients is selected through Enterprise Repository and data is gathered into a data mart



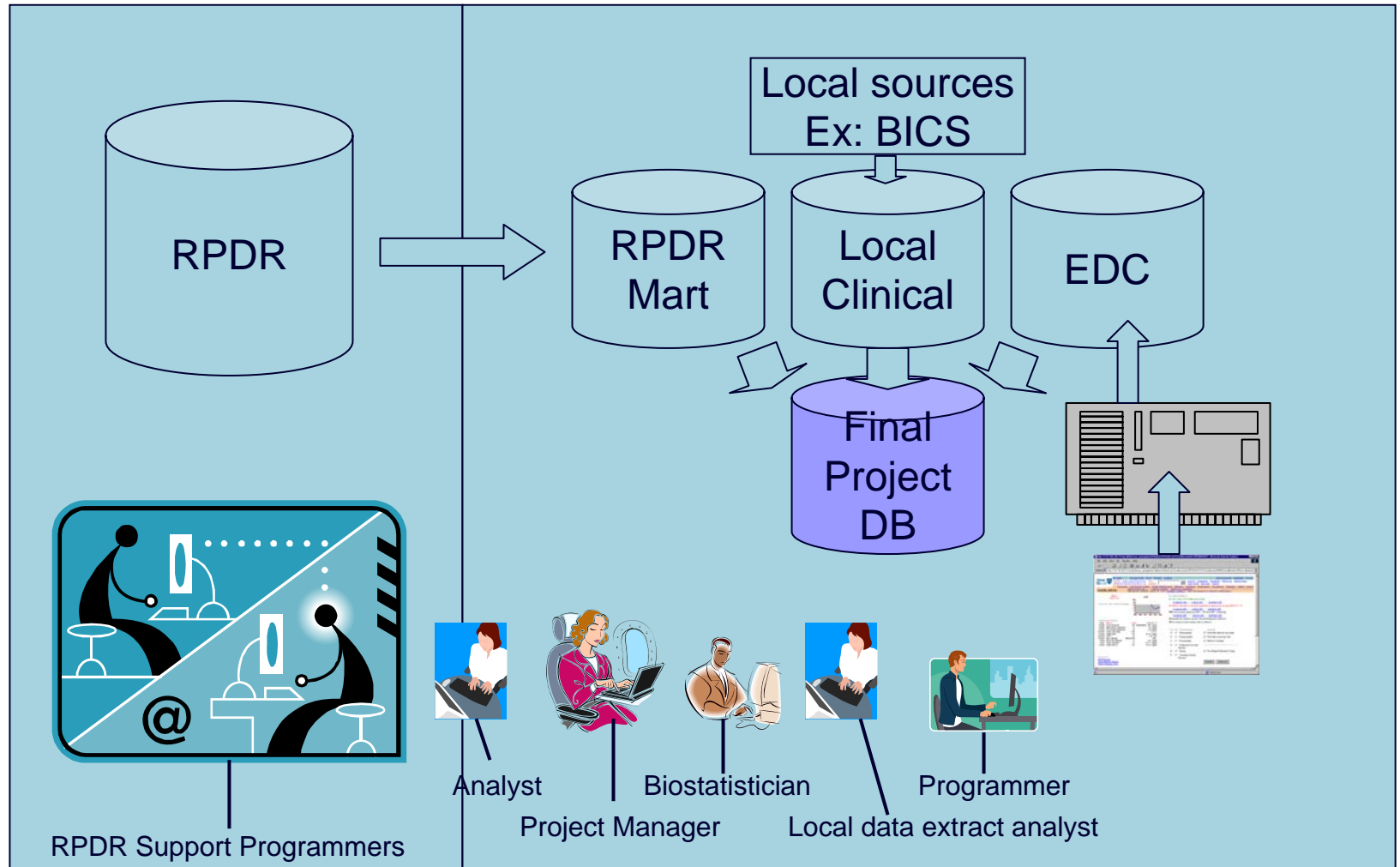
Data is available through the i2b2 Workbench

The screenshot displays the i2b2 Patient Recruitment Workbench interface, showing a complex data management and analysis environment. The interface is divided into several panes:

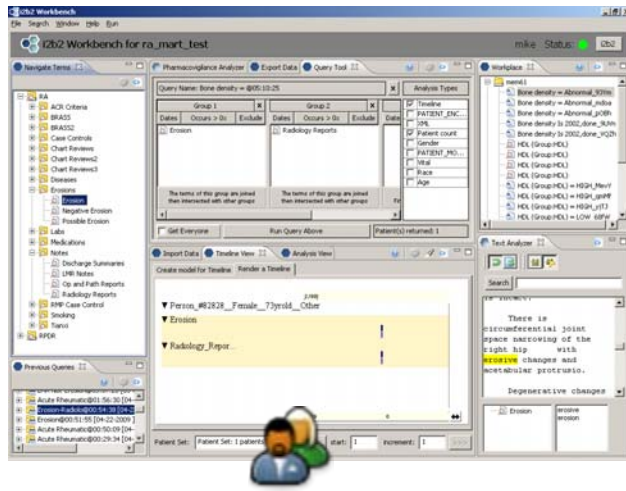
- Concept Trees:** Located on the left, it shows a hierarchical structure of medical concepts. The 'Diagnoses' tree is expanded, showing categories like 'Respiratory system' (including Acute respiratory infections, Chronic obstructive diseases, Adult bronchiectasis, Asthma, Bronchiectasis NOS, Bronchiectasis, Bronchioleclasis, Bronchitis NOS, and Bronchitis, not specified as acute or chronic) and 'Digestive System, BWH on 08/21/06' (including various dates and patient counts).
- Patient Sets:** Below the concept trees, a list of patient sets is shown, including 'Digestive System, BWH on 08/21/06' and 'Circulatory System on 07/22/06', each with associated dates and patient counts.
- Query Groups:** The top right pane shows three groups (Group 1, Group 2, Group 3) with columns for 'Dates', 'Occurs 1x', and 'Exclude'. Group 1 contains 'Bronchodilators', Group 2 contains 'Prednisolone' and 'Prednisone', and Group 3 is empty. A 'Run Query' button is visible.
- Timeline View:** The central pane displays a timeline view for several patients, including 'Person_#82032_Female_68yroid_White', 'Person_#82008_Female_37yroid_White', 'Person_#82202_Male_Deaf_Black', and 'Person_#82229_Female_63yroid_White'. The timeline shows events (represented by vertical bars) for 'Bronchodilators' and 'Prednisolone-Pr...'.
- Related Topics:** The bottom right pane shows a 'Related Topics' section with links to 'About Outline', 'Dynamic Help', and 'Patient Schedule'. It also includes a 'Report Diagnosis Detail Image' section showing a brain scan image.
- Workflow Diagram:** The bottom left pane shows a workflow diagram with components like 'Drop Patient Object', 'ObjectToRecord', 'RecordDisassembler', 'Synchronizer', and 'Display'.

The interface is titled 'Patient Recruitment Workbench' and includes a user name 'Shawn Murphy' and a status bar with 'Log out' and 'Help' buttons.

Team support for Projects



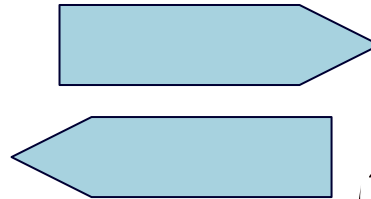
NLP Workflow



I2b2 Project Investigators

Results Delivery

Communication



NLP Specialists

NLP (and comedy) is not pretty

Programmer's File Editor - [050210_1629;MinDem1.txt]

File Edit Options Template Execute Macro Window Help

SOCIAL HISTORY: The patient is married with four grown daughters, **uses tobacco**, has wine with dinner. **Smoker**

PRINCIPAL DIAGNOSIS: LEFT LOWER LOBE PNEUMONIA

SECONDARY: 1. CHD 2. HEENT

SOCIAL HISTORY: The patient is a **nonsmoker**. No alcohol. **Non-Smoker**

SOCIAL HISTORY: **Negative for tobacco**, alcohol, and IV drug abuse.

PAST MEDICAL HISTORY: (1) Hip fracture. (2) Bronchiectasis.

BRIEF RESUME OF HOSPITAL COURSE: 63 yo woman with COPD, **50 pack-yr tobacco (quit 3 wks ago)**, **Past Smoker**

ALLERGIES: (1) Aspirin. (2) Ciprofloxacin. (3) Penicillin.

SOCIAL HISTORY: The patient lives alone and denies tobacco or alcohol use. **Unclear smoking history** **???**

PHYSICAL EXAMINATION: Temperature 97.2, pulse 60, respirations 20, blood pressure 160/63, oxygen saturation 95% on room air. HEENT: Normocephalic and atraumatic. Pupils equal and reactive.

LABORATORY DATA: Sodium 148, potassium 3.4, chloride 97, bicarbonate 24, creatinine 1.2, glucose 100, hemoglobin 12.5, hematocrit 38, platelets 250,000.

HOSPITAL COURSE: ... It was recommended that she receive ... We also added Lactinax, oral form of **Lactobacillus acidophilus** population of her gut. **Hard to pick**

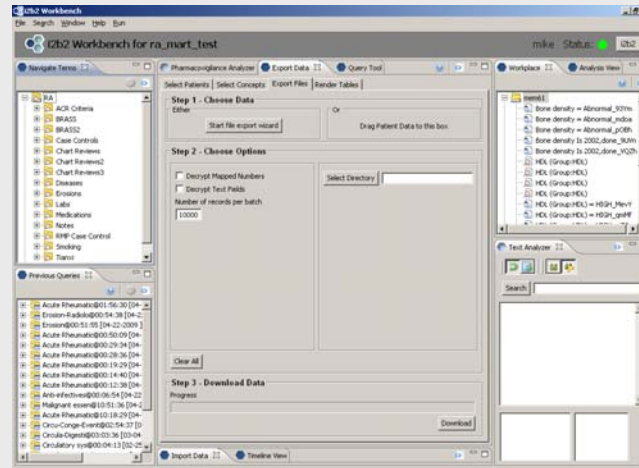
HOSPITAL COURSE: The patient was seen and evaluated by the physician on 10/10/02. The patient was seen and evaluated by the physician on 10/10/02. The patient was seen and evaluated by the physician on 10/10/02.

SH: widow, lives alone, 2 children, no **tob/alcohol**. **Hard to pick**

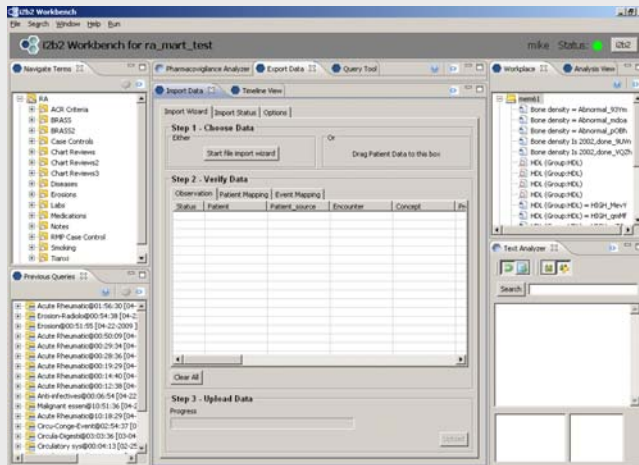
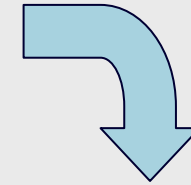
The patient was seen and evaluated by the physician on 10/10/02. The patient was seen and evaluated by the physician on 10/10/02. The patient was seen and evaluated by the physician on 10/10/02.

STANDARD PRESENTATION: 1. 2. 3. 4. 5. 6. 7. 8. 9. 10. 11. 12. 13. 14. 15. 16. 17. 18. 19. 20. 21. 22. 23. 24. 25. 26. 27. 28. 29. 30. 31. 32. 33. 34. 35. 36. 37. 38. 39. 40. 41. 42. 43. 44. 45. 46. 47. 48. 49. 50. 51. 52. 53. 54. 55. 56. 57. 58. 59. 60. 61. 62. 63. 64. 65. 66. 67. 68. 69. 70. 71. 72. 73. 74. 75. 76. 77. 78. 79. 80. 81. 82. 83. 84. 85. 86. 87. 88. 89. 90. 91. 92. 93. 94. 95. 96. 97. 98. 99. 100. 101. 102. 103. 104. 105. 106. 107. 108. 109. 110. 111. 112. 113. 114. 115. 116. 117. 118. 119. 120. 121. 122. 123. 124. 125. 126. 127. 128. 129. 130. 131. 132. 133. 134. 135. 136. 137. 138. 139. 140. 141. 142. 143. 144. 145. 146. 147. 148. 149. 150. 151. 152. 153. 154. 155. 156. 157. 158. 159. 160. 161. 162. 163. 164. 165. 166. 167. 168. 169. 170. 171. 172. 173. 174. 175. 176. 177. 178. 179. 180. 181. 182. 183. 184. 185. 186. 187. 188. 189. 190. 191. 192. 193. 194. 195. 196. 197. 198. 199. 200. 201. 202. 203. 204. 205. 206. 207. 208. 209. 210. 211. 212. 213. 214. 215. 216. 217. 218. 219. 220. 221. 222. 223. 224. 225. 226. 227. 228. 229. 230. 231. 232. 233. 234. 235. 236. 237. 238. 239. 240. 241. 242. 243. 244. 245. 246. 247. 248. 249. 250. 251. 252. 253. 254. 255. 256. 257. 258. 259. 260. 261. 262. 263. 264. 265. 266. 267. 268. 269. 270. 271. 272. 273. 274. 275. 276. 277. 278. 279. 280. 281. 282. 283. 284. 285. 286. 287. 288. 289. 290. 291. 292. 293. 294. 295. 296. 297. 298. 299. 300. 301. 302. 303. 304. 305. 306. 307. 308. 309. 310. 311. 312. 313. 314. 315. 316. 317. 318. 319. 320. 321. 322. 323. 324. 325. 326. 327. 328. 329. 330. 331. 332. 333. 334. 335. 336. 337. 338. 339. 340. 341. 342. 343. 344. 345. 346. 347. 348. 349. 350. 351. 352. 353. 354. 355. 356. 357. 358. 359. 360. 361. 362. 363. 364. 365. 366. 367. 368. 369. 370. 371. 372. 373. 374. 375. 376. 377. 378. 379. 380. 381. 382. 383. 384. 385. 386. 387. 388. 389. 390. 391. 392. 393. 394. 395. 396. 397. 398. 399. 400. 401. 402. 403. 404. 405. 406. 407. 408. 409. 410. 411. 412. 413. 414. 415. 416. 417. 418. 419. 420. 421. 422. 423. 424. 425. 426. 427. 428. 429. 430. 431. 432. 433. 434. 435. 436. 437. 438. 439. 440. 441. 442. 443. 444. 445. 446. 447. 448. 449. 450. 451. 452. 453. 454. 455. 456. 457. 458. 459. 460. 461. 462. 463. 464. 465. 466. 467. 468. 469. 470. 471. 472. 473. 474. 475. 476. 477. 478. 479. 480. 481. 482. 483. 484. 485. 486. 487. 488. 489. 490. 491. 492. 493. 494. 495. 496. 497. 498. 499. 500. 501. 502. 503. 504. 505. 506. 507. 508. 509. 510. 511. 512. 513. 514. 515. 516. 517. 518. 519. 520. 521. 522. 523. 524. 525. 526. 527. 528. 529. 530. 531. 532. 533. 534. 535. 536. 537. 538. 539. 540. 541. 542. 543. 544. 545. 546. 547. 548. 549. 550. 551. 552. 553. 554. 555. 556. 557. 558. 559. 560. 561. 562. 563. 564. 565. 566. 567. 568. 569. 570. 571. 572. 573. 574. 575. 576. 577. 578. 579. 580. 581. 582. 583. 584. 585. 586. 587. 588. 589. 590. 591. 592. 593. 594. 595. 596. 597. 598. 599. 600. 601. 602. 603. 604. 605. 606. 607. 608. 609. 610. 611. 612. 613. 614. 615. 616. 617. 618. 619. 620. 621. 622. 623. 624. 625. 626. 627. 628. 629. 630. 631. 632. 633. 634. 635. 636. 637. 638. 639. 640. 641. 642. 643. 644. 645. 646. 647. 648. 649. 650. 651. 652. 653. 654. 655. 656. 657. 658. 659. 660. 661. 662. 663. 664. 665. 666. 667. 668. 669. 670. 671. 672. 673. 674. 675. 676. 677. 678. 679. 680. 681. 682. 683. 684. 685. 686. 687. 688. 689. 690. 691. 692. 693. 694. 695. 696. 697.

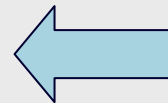
NLP Specialists Workstation



Export Notes



Import
Derived
Codes



NLP Specialists

Investigator Review

The screenshot displays the i2b2 Workbench interface for the 'ra_mart_test' dataset. The top menu bar includes 'File', 'Search', 'Window', 'Help', and 'Run'. The main window is titled 'i2b2 Workbench for ra_mart_test' and shows a user 'mike' with a status of 'online'.

The interface is divided into several panes:

- Navigate Terms:** A tree view on the left showing various medical terms and categories, including 'RA', 'ACR Criteria', 'BRASS', 'BRASS2', 'Case Controls', 'Chart Reviews', 'Diseases', 'Erosions', 'Labs', 'Medications', 'Notes', 'RMP Case Control', 'Smoking', 'Tianxi', and 'RPDR'.
- Pharmacovigilance Analyzer:** A central pane for creating and running queries. It shows a query named 'Bone density = @05:10:25'. The query is structured with two groups: 'Group 1' (Erosion) and 'Group 2' (Radiology Reports). The 'Analysis Types' section on the right lists various analysis options, including 'Timeline', 'PATIENT_ENC...', 'XML', 'Patient count', 'Gender', 'PATIENT_MO...', 'Vital', 'Race', and 'Age'.
- Workplace:** A pane on the right showing a list of results for the query, including 'Bone density = Abnormal_93Ym', 'Bone density = Abnormal_mdoa', 'Bone density = Abnormal_p08h', 'Bone density Is 2002,done_9UVn', 'Bone density Is 2002,done_VQZh', 'HDL (Group:HDL)', 'HDL (Group:HDL)', 'HDL (Group:HDL)', 'HDL (Group:HDL) = HIGH_MevY', 'HDL (Group:HDL) = HIGH_qmMf', 'HDL (Group:HDL) = HIGH_yjTJ', and 'HDL (Group:HDL) = LOW_68FW'.
- Text Analyzer:** A pane at the bottom right showing a search for 'erosive' and displaying a text snippet: 'There is circumferential joint space narrowing of the right hip with erosive changes and acetabular protrusio. Degenerative changes'.
- Previous Queries:** A pane at the bottom left showing a list of previous queries, including 'Acute Rheumatic@01:56:30 [04-2009]', 'Erosion-Radiolo@00:54:38 [04-2009]', 'Erosion@00:51:55 [04-22-2009]', 'Acute Rheumatic@00:50:09 [04-2009]', and 'Acute Rheumatic@00:29:34 [04-2009]'.

The main query results pane shows a list of results for the query 'Bone density = @05:10:25'. The results are displayed in a table with columns for 'Person', 'Erosion', and 'Radiology_Report'. The first result is 'Person_#82828_Female_73yroid_Other'. The 'Erosion' column shows 'Erosion' and the 'Radiology_Report' column shows 'Radiology_Report'.

Select patients for clinical trials

The screenshot displays the BIRN Workbench software interface, which is used for managing and viewing medical data. The main window is titled "BIRN Workbench for Demo Group" and includes a "Query Tool" tab. The "Image View" tab is active, showing a list of patients and their associated images.

Table 1: Patient Selection Data

Row #	Decision	PSet #	Patient ID	Patient Name	Gender	Race	Date of Birth	Age	MRNs
1	Yes	1	1000000001	xxxxxx, xxxxxx	F	black		74	
2	Yes	1	1000000002	xxxxxx, xxxxxx	F	white		55	
3	Yes	1	1000000003	xxxxxx, xxxxxx	M	asian		73	
4	Und	1	1000000004	xxxxxx, xxxxxx	M	black		28	

Table 2: Image Selection Data

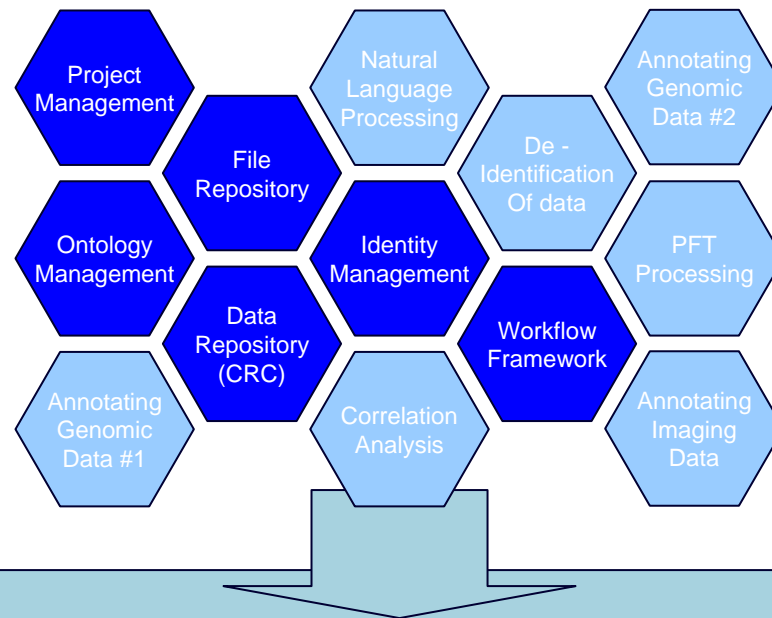
Index	Name	Time
1	Sagittal	2003-05-03T
2	Sagittal	2003-05-03T
3	Sagittal	2003-05-03T
4	Sagittal	2003-05-03T

The interface also includes a "Previous Queries" panel on the left, a "Patient Set" panel in the center, and a "Query Tool" panel on the right. The "Query Tool" panel shows a list of queries and a "Patient Set" panel with a "Patient Set - 416 Patients" button. The "Query Tool" panel also includes a "Subject ID" field and a "List" button.

The "Image View" tab displays a list of images and a "Patient Set" panel. The "Patient Set" panel shows a list of patients and a "Patient Set - 416 Patients" button. The "Image View" tab also includes a "Subject ID" field and a "List" button.

The "Image View" tab displays a list of images and a "Patient Set" panel. The "Patient Set" panel shows a list of patients and a "Patient Set - 416 Patients" button. The "Image View" tab also includes a "Subject ID" field and a "List" button.

- Repurpose medical record information for research studies
 - I2b2 Workbench
 - Natural language processing
- Enable genomic studies
 - Tissue/blood selection
 - Genetic data integration

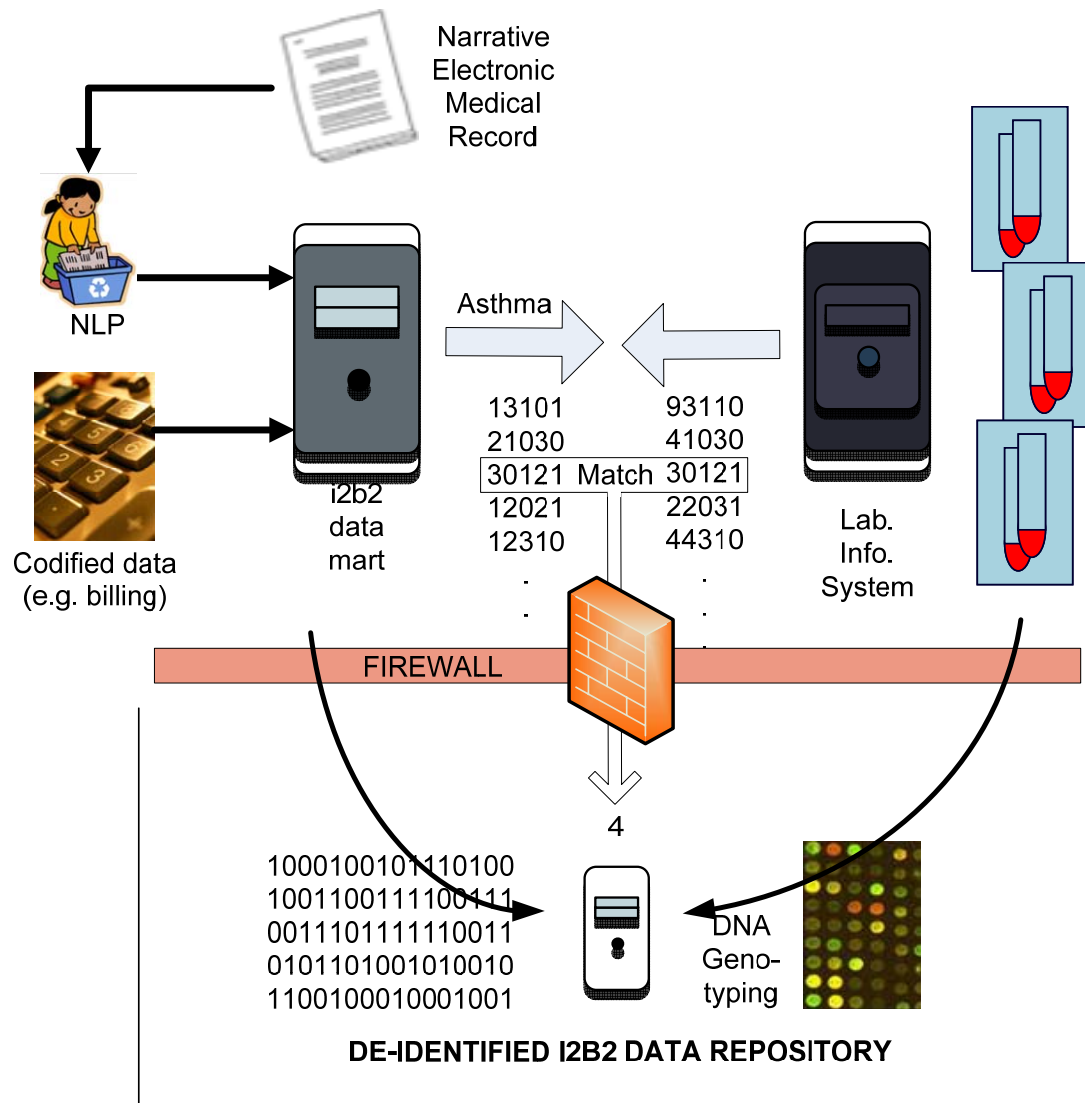


Use of medical record data in clinical studies
focused upon teamwork and workflow

High Throughput Methods for supporting Research at Partners Healthcare

- Set of patients is selected from medical record data in a high throughput fashion
- Investigators work with the data of these patients using new i2b2 tools and a specialized team, both developed to work specifically with medical record data
- Using the BETR/Crimson system, tissues of these patients can be made available for genomic and biochemical analysis
 - BWH Pathology –Lynn Bry
- Automated discovery can be created from these projects to support further hypothesis-driven research

Genotype samples and compare to controls



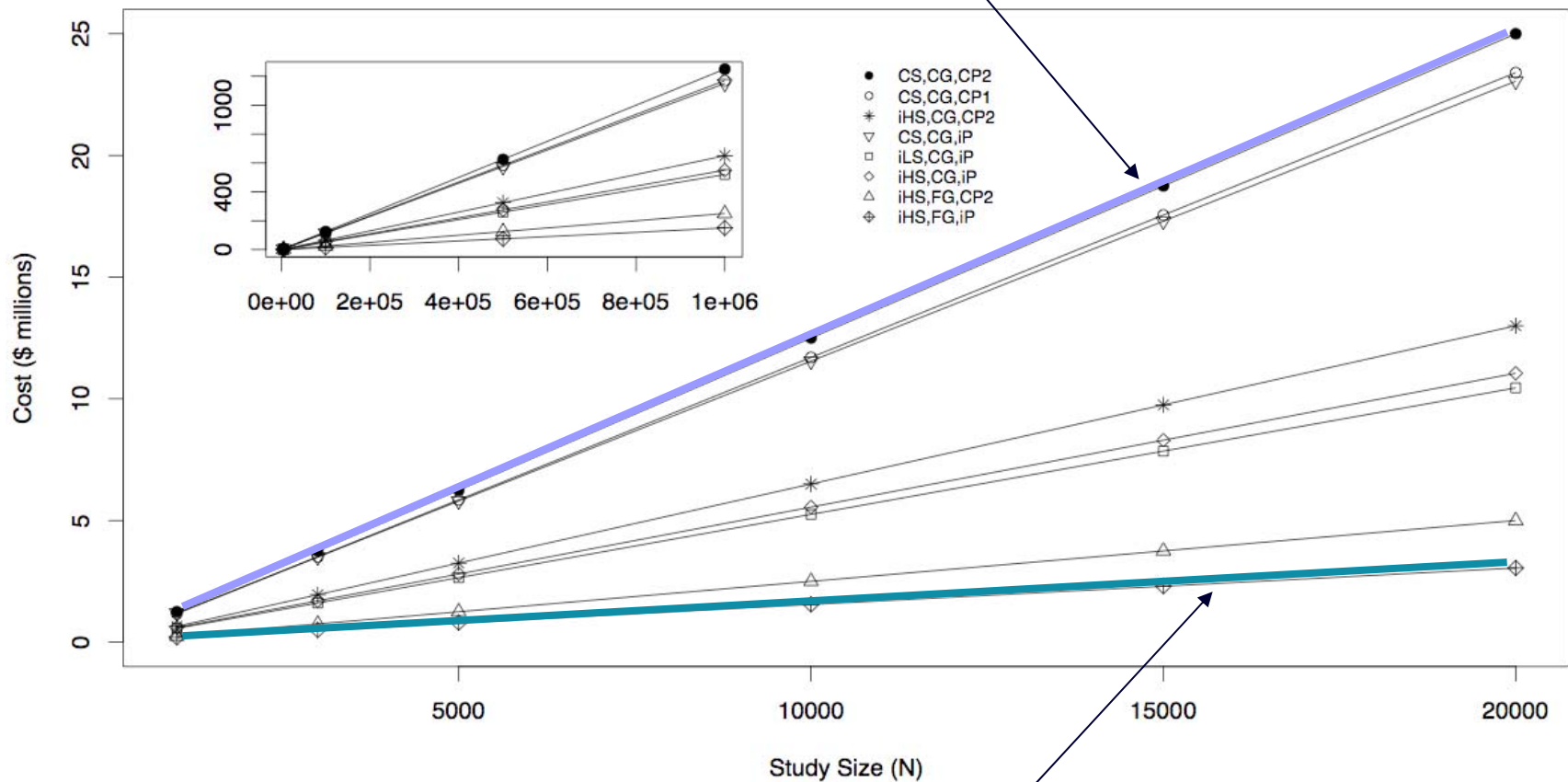
Cost and time benefit of Instrumenting with Sample Collection for Modest-size Study with 10,000 subjects (cases + controls)

Old vs. New	Cost (\$)	Time
1 chart review per patient (CP1)	\$20	15 minutes/subject
High-throughput phenotyping (iP) through RPDR and i2b2	\$50K Total	1 month total (conservative high estimate)
Sample acquisition through primary care provider (CP)	\$650	3-5 subjects/week ¹
High-throughput sample acquisition through RPDR and BETR/Crimson.	\$20	50-200 subjects /week²

= \$6.7 million/study vs. \$250 thousand/study

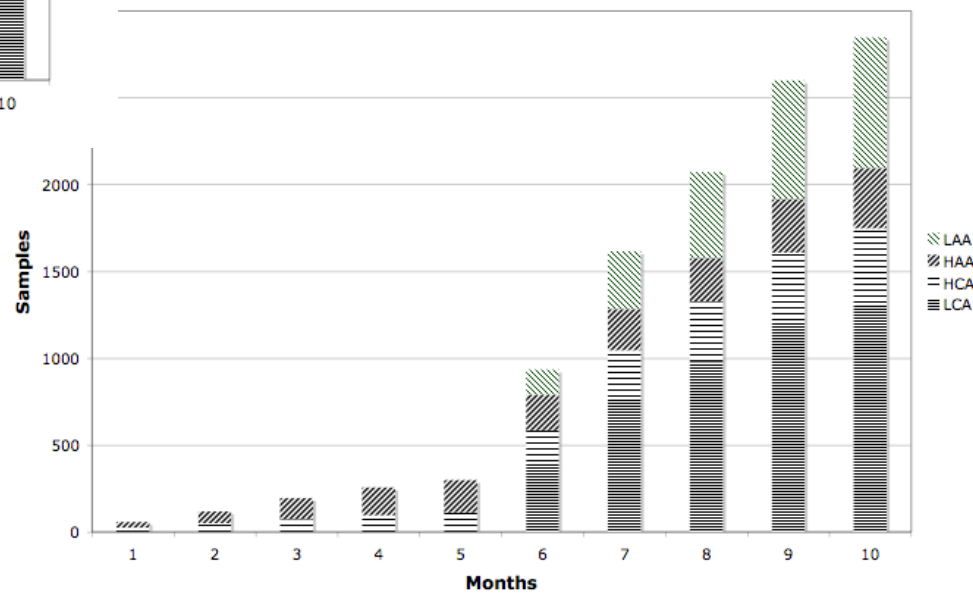
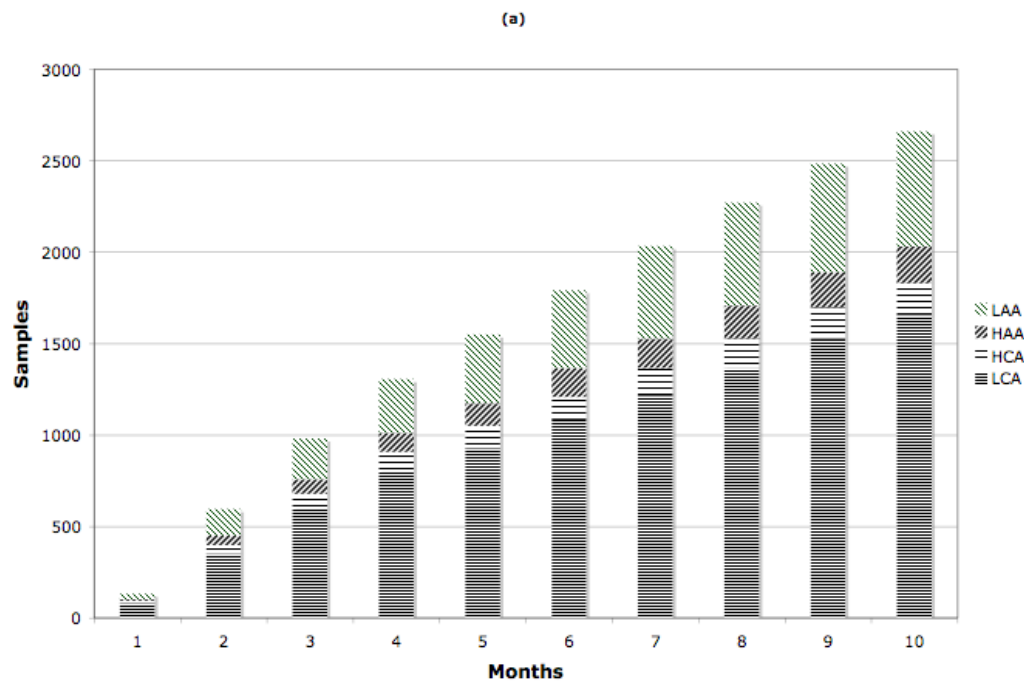
Escalating cost and time benefit of Instrumenting with Sample Collection

Previous model for collecting specimens

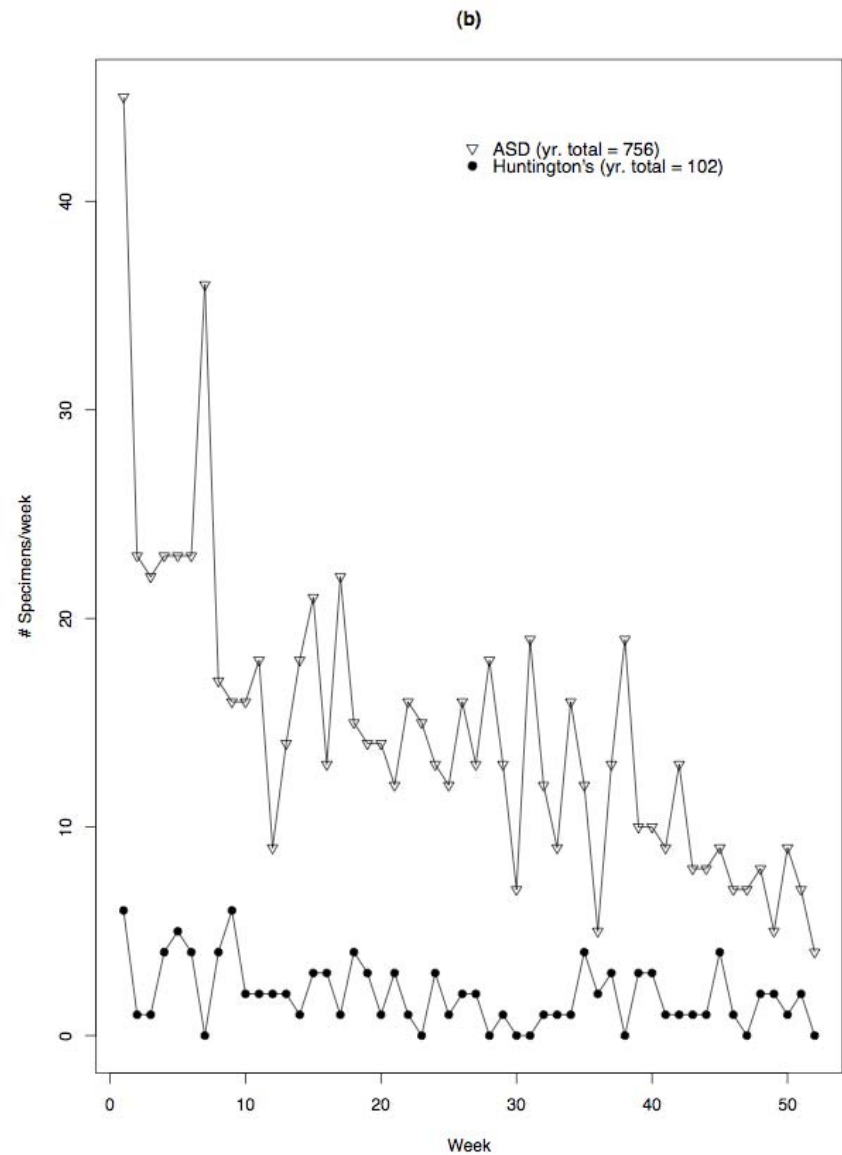
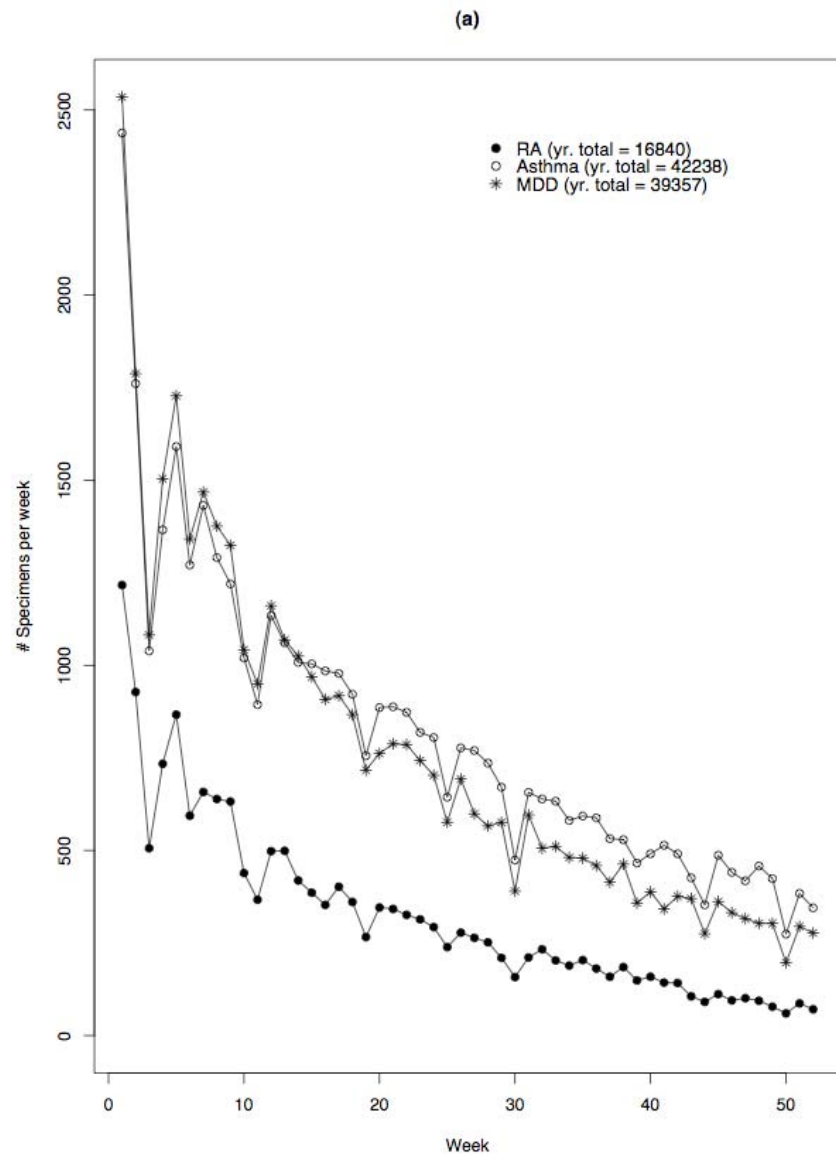


New model for collecting specimens

Meeting Expectations



Accrual Rates



High Throughput Methods for supporting Research at Partners Healthcare

- Set of patients is selected from medical record data in a high throughput fashion
- Investigators work with the data of these patients using new i2b2 tools and a specialized team, both developed to work specifically with medical record data
- Using the Crimson system, tissues of these patients can be made available for genomic and biochemical analysis
- Automated discovery can be created from these projects to support further hypothesis-driven research
 - Pharmacovigilance – John Brownstein and Judy Colecchi

Centrally supported Automated Discovery Projects

Clinical trials performed in-silico

- Performing an observational, phase IV study is an expensive and complex process that can be potentially modeled in a retrospective database using groups of patients available in the large amounts of highly organized medical data.
- Fundamental problems complicate this approach:
 - Patients drift in and out of the system. Sophisticated statistical models using adequate control populations are necessary to compensate.
 - Confounding variables are not found in the database. Sophisticated natural language processing is needed to extract the confounders from textual reports to allow these confounders to be controlled.
 - Missing data disrupts typical statistical approaches

Navigate Terms

- Avandia_NLP
 - PV
 - Avandia
 - Cardiovascular Event
 - Enrolled
 - No Rosi And CV
 - No Rosi And No CV
 - Rosi And CV
 - Rosi And No CV
- RPDR
 - Demographics
 - Diagnoses
 - DRG
 - Health History
 - Laboratory Tests
 - Medications
 - Alternative medicines
 - Anti-infectives
 - Anthyperlipidemic agents
 - Antineoplastics
 - Biologicals
 - Cardiovascular agents
 - Central nervous system agents
 - Coagulation modifiers
 - CPT Devices
 - CPT Medications
 - Devices
 - Gastrointestinal agents
 - Hormones
 - Adrenal cortical steroids
 - Antidiabetic agents
 - Alpha-glucosidase inhibitors
 - Byetta (exenatide) - LMR 5176
 - Insulin
 - Meglitinides
 - Non-sulfonylureas
 - Sulfonylureas
 - Thiazolidinediones
 - Metformin-rosiglitazone
 - Pioglitazone
 - Rosiglitazone

Previous Queries

- Avandia-Cardiov@09:20:31 [11-24-2008] [snm0]
- Enrolled@10:55:15 [11-13-2008] [snm0]
- Actos@09:39:18 [11-03-2008] [snm0]
- Actos@09:37:00 [11-03-2008] [snm0]
- Actos@09:36:48 [11-03-2008] [snm0]
- Acute Rheumatic@11:07:49 [07-02-2008] [snm0]
- Circulatory sys@11:05:19 [07-02-2008] [snm0]
- Circulatory sys@11:00:50 [07-02-2008] [snm0]
- Medications@11:00:31 [07-02-2008] [snm0]

Query Tool

Query Name:

Reset Groups

Group 1

Group 2

Group 3

Add Group

Run Query

Patient(s) returned:

Timeline View

Create model for Timeline

Render a Timeline

Query Name: Avandia-Cardiov@09:20:31 [11-24-2008] [snm0]

Row #	Name of Terms	Value	Value Text	Height	Color
1	Enrolled	N/A	N/A	Medium	
2	Rosiglitazone event	N/A	N/A	Medium	
3	Cardiovascular event	N/A	N/A	Medium	
4	No Rosi And CV	N/A	N/A	Medium	
4	No Rosi And No CV	N/A	N/A	Medium	
4	Rosi And CV	N/A	N/A	Medium	
4	Rosi And No CV	N/A	N/A	Medium	
5	Avandia - NLP	N/A	N/A	Medium	
5	rosiglitazone - NLP	N/A	N/A	Medium	

Delete From List

Delete All

Put In Order

Move Up

Move Down

☐ Display concepts with no data ☒ Display patient demographics

Patient Set: Patient Set: 954 patients <<< start: 11 increment: 10 >>>

Help

Workplace

- SHARED
 - snm0
 - Avandia-Cardiov@09:20:31 [11-24-2008] [snm0]

Avandia_NLP

PV

Avandia

Cardiovascular Event

Enrolled

No Rosi And CV

No Rosi And No CV

Rosi And CV

Rosi And No CV

RPDR

Demographics

Diagnoses

DRG

Health History

Laboratory Tests

Medications

Alternative medicines

Anti-infectives

Antihyperlipidemic agents

Antineoplastic

Biologicals

Cardiovascular agents

Central nervous system agents

Coagulation modifiers

CPT Devices

CPT Medications

Devices

Gastrointestinal agents

Hormones

Adrenal cortical steroids

Antidiabetic agents

Alpha-glucosidase inhibitors

Byetta (exenatide) - LMR 5176

Insulin

Meglitinides

Non-sulfonylureas

Sulfonylureas

Thiazolidinediones

Metformin-rosiglitazone

Pioglitazone

Rosiglitazone

Previous Queries

Avandia-Cardiov@09:20:31 [11-24-2008] [snm0]

Enrolled@10:55:15 [11-13-2008] [snm0]

Actos@09:39:18 [11-03-2008] [snm0]

Actos@09:37:00 [11-03-2008] [snm0]

Actos@09:36:48 [11-03-2008] [snm0]

Acute Rheumatic@11:07:49 [07-02-2008] [snm0]

Circulatory sys@11:05:19 [07-02-2008] [snm0]

Circulatory sys@11:00:50 [07-02-2008] [snm0]

Medications@11:00:31 [07-02-2008] [snm0]

Query Tool

Query Name:

Reset Groups

Group 1

Dates Occurs > 0x Jude

Group 2

Dates Occurs > 0x Jude

Group 3

Dates Occurs > 0x Jude

Add Group

Run Query

Patient(s) returned:

Timeline View

Export Data

KTreeMap View

Create model for Timeline

Render a Timeline

2/99 0 1 2 3 4 5 6 7

▼ Person_#240407_Female_Dead_White

▼ Enrolled

▼ Rosiglitazone_e...

▼ Cardiovascular_...

▼ No_Rosi_And_CVN...

▼ Avandiarosigit...

▼ Person_#240419_Female_53yrold_Black

▼ Enrolled

▼ Rosiglitazone_e...

▼ Cardiovascular_...

▼ No_Rosi_And_CVN...

▼ Avandiarosigit...

▼ Person_#240606_Female_73yrold_White

▼ Enrolled

▼ Rosiglitazone_e...

2/99 0 1 2 3 4 5 6 7 8

Patient Set: Patient Set: 954 patients <<< start: 1 increment: 10 >>>

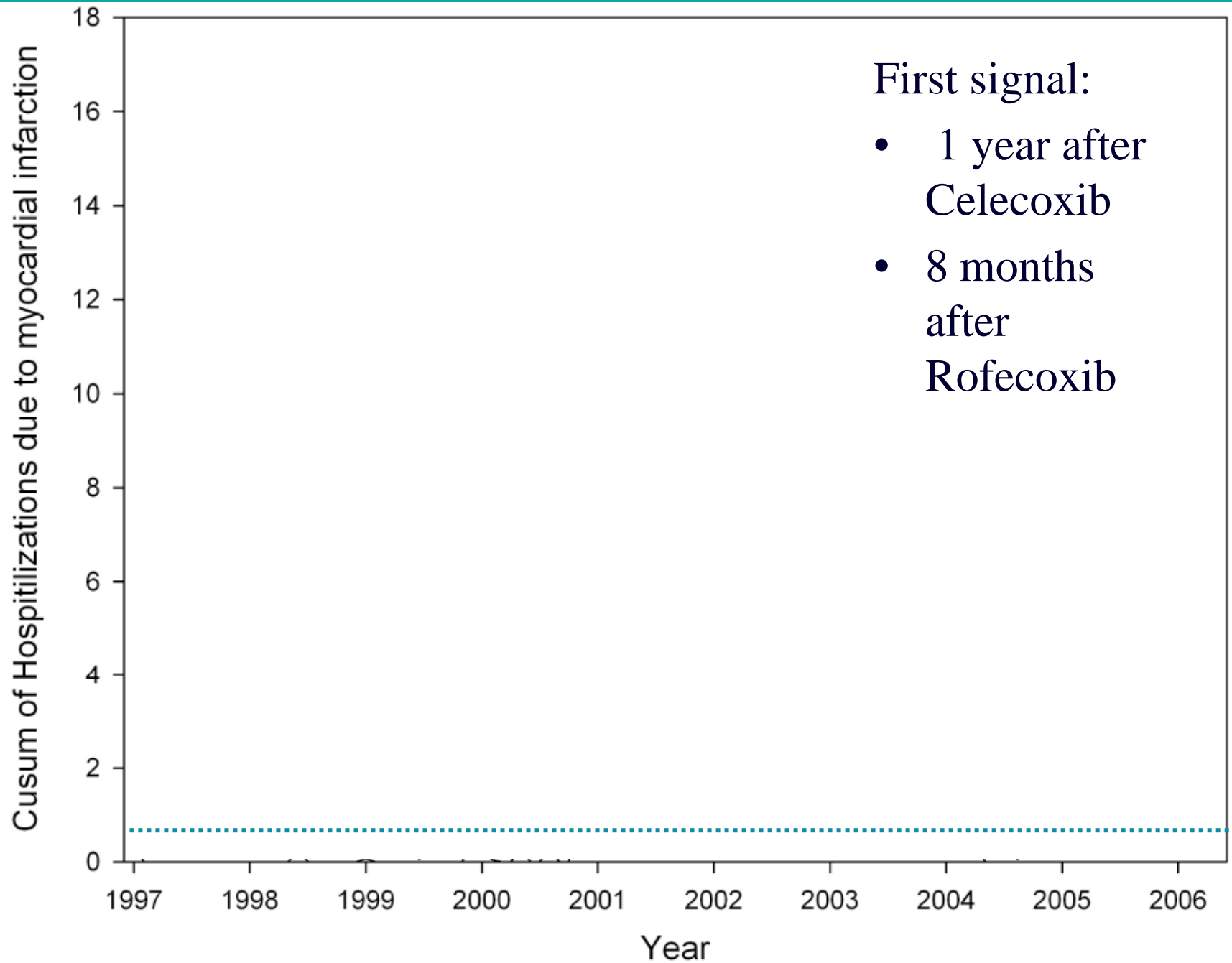
Help

Workplace

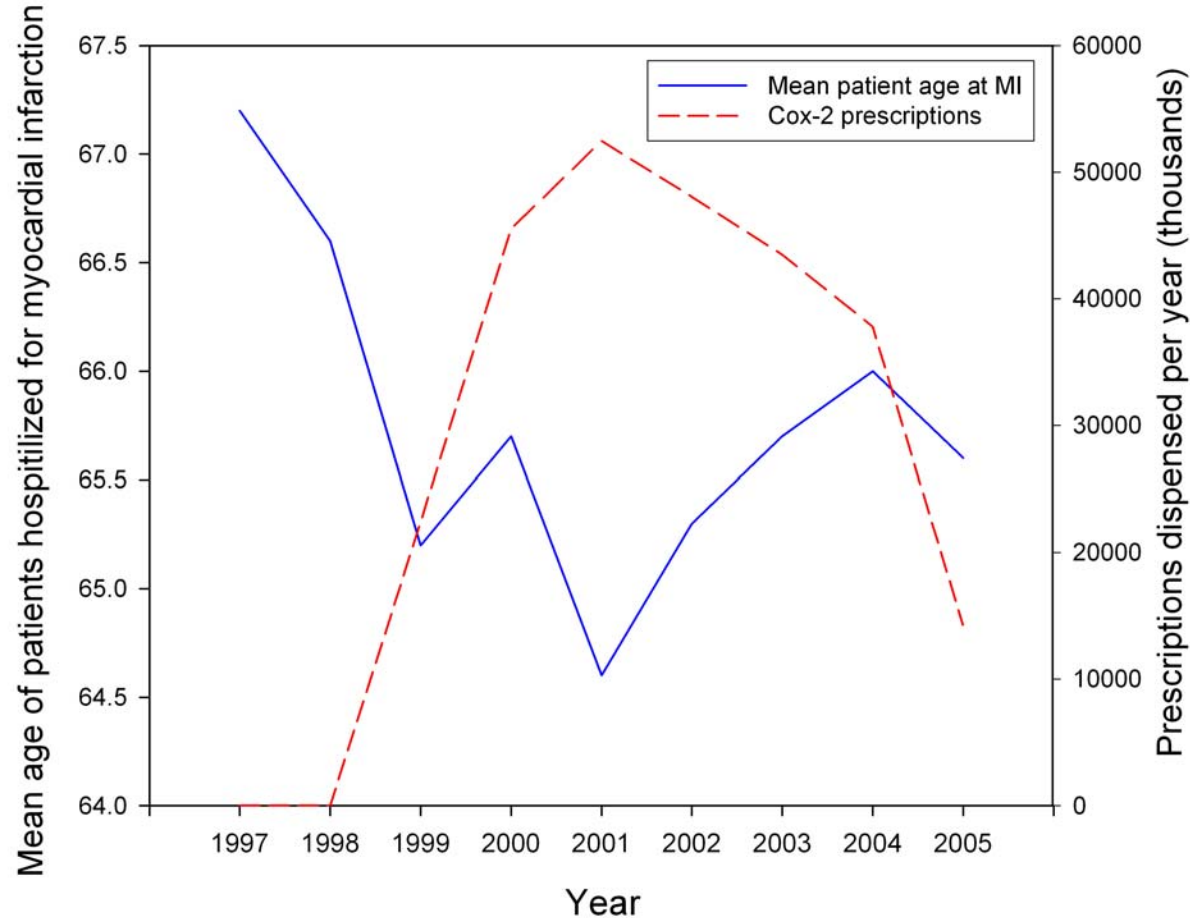
SHARED

snm0

Avandia-Cardiov@09:20:31 [11-24-2008] [snm0]

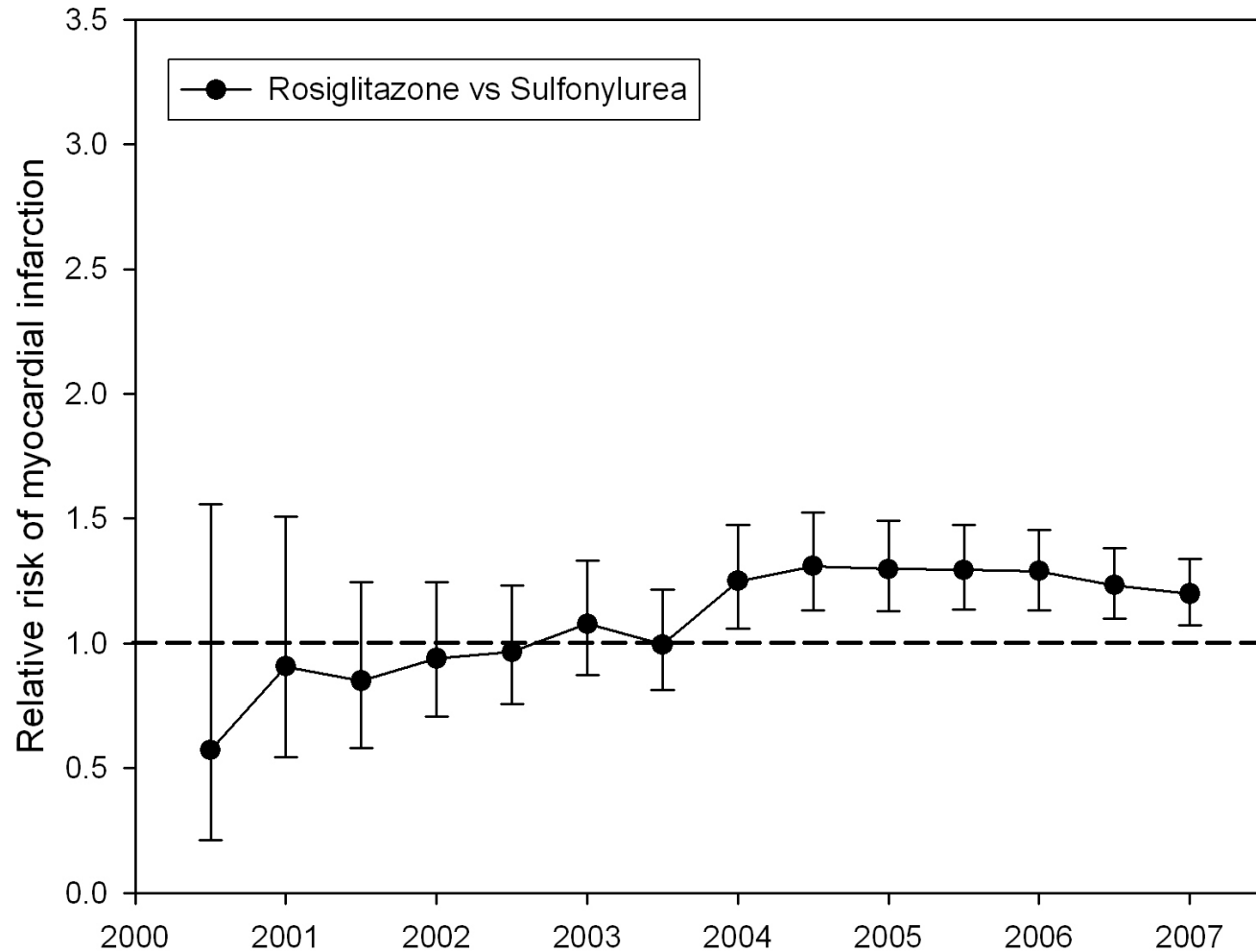


Effect on patient age



- Negative association between mean age at MI and prescription volume
- Spearman correlation -0.67, $P < 0.05$

Prospective Surveillance of Cardiovascular Events while taking Avandia



Odds Ratios for Diseases expressed in comparing Rosiglitazone vs. Pioglitazone

i2b2 Workbench for Demo Group

File Search Window Help Run

Demo User Status: ● i2b2

Navigate Terms

- Laboratory Tests
- Medications
 - Alternative medicines
 - Anti-infectives
 - Antihyperlipidemic agents
 - Antineoplastic
 - Biologicals
 - Cardiovascular agents
 - Central nervous system agent
 - Coagulation modifiers
 - Gastrointestinal agents
 - Hormones
 - Adrenal cortical steroids
 - Antidiabetic agents
 - Alpha-glucosidase inhibitors
 - Insulin
 - Meglitinides
 - Non-sulfonylureas
 - Sulfonylureas
 - Thiazolidinediones
 - Bisphosphonates

Query Tool

Query Name:

Reset Groups

Group 1			Group 2			Group 3		
Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude
Thiazolidinediones								

The terms of this group are joined then intersected with other groups

Drag terms from Navigate, Find and Workplace into this group

Drag terms from N Find and Workplace into this group

Run Query Patient(s) returned:

Workplace

- SHARED
- demo
 - Cardiac Tests
 - Cardiovascular agents
 - Definitions (Queries)
 - Definitions (Query Groups)
 - Demogra-Diagnos@02:31:36 [1-11-
 - Diabetes mellit@22:57:00 [1-15-20C
 - Diagnoses
 - IRON (LOINC:2498-4)
 - Iron (G-ALT (SG@09:08:27 [12-15-2
 - Observations
 - PATIENT:1000000025
 - PATIENTSET_8971_Ischemic
 - Patient Sets

Pharmacovigilance Analysis

Set up analysis Treemap Breakdown Timeline Export

Analysis Title: Avandia vs. Pioglitazone Calculation for SNM0 on 3-04-2009

Strategy: Squarified

Color Provider: HSB linear

TREX params:

TotalPatients:

Open File .. Save File .. Open in Excel

10 Lessons from i2b2

- 1) Power is Numbers of patients recruited
- 2) Enable Enterprise use of patients for research
- 3) Extensible Architecture for developers
- 4) Enable Scientist Workflow
- 5) Enable Team Work with informaticians and researchers
- 6) Enable Natural Language Processing
- 7) Importance of Visualizations
- 8) High throughput Tissue acquisition for Genomic Research
- 9) Enable Health Surveillance
- 10) Enable Data Sharing