Overview of EHR phenotyping – successes, challenges, and examples

Josh Denny, MD, MS
josh.denny@vanderbilt.edu
Vanderbilt University, Nashville, Tennessee, USA

AMIA 2015 NLP Tutorial
Disclaimers

I receive funding from:

• NIH: NLM, NHGRI, NIGMS, NCI, NCATS
• Reynolds Foundation (Geriatrics Education)
• National Board of Medical Examiners
Outline for my block

• Intro to use of the EHR for genomic research
• NLP for EHR phenotyping
  • diseases/traits
  • Drug-response phenotypes
• NLP for medical education
EHR adoption increasing faster than anticipated

![Graph showing EHR adoption over time compared to predictions.](image-url)
Genomic achievements since the Human Genome Project

2004: Publication of finished human genome sequence

2005: First genome-wide association study

2010: 500th genome-wide association study

2010: First EMR-based genetic studies
The Synthetic Derivative
A de-identified and continuously-updated image of the EMR: 2,358,760 subjects
Using clinical notes to enhance knowledge

Clinical Notes, test reports, etc

Billing codes

Retrained cTAKES Smoking module (an example of a routine procedure)

Structured Output
DrugName: atenolol
Strength: 50 mg
Frequency: daily

EMR’ Research Derivative

Find Biomedical Concepts and Qualifiers (KnowledgeMap)

CC: SOB
HPI: This is a 65yo w/ h/o CHF, ... no dm2... on atenolol 50mg daily... Mother had RA.

Medication Extraction (MedEx)

Structured Output
certainty (positive, negated)
Who experienced it? (patient or family member?)

Chief_complaint:
C0392680: Shortness of Breath
history_present_illness:
Congestive Heart Failure
Type 2 diabetes, negated
mother_medical_history:
rheumatoid arthritis

Billing codes

Research Derivative

VANDERBILT UNIVERSITY MEDICAL CENTER
Extracted detailed smoking history from clinical narratives

Travis Osterman
AMIA
Weds 10:30-12 session
CC: SOB
HPI: 71 yo woman h/o DM, HTN, Dilated CM/CHF, Afib s/p embolic event, chronic diarrhea, admitted with SOB. CXR pulm edema. Rx’d Lasix.
All: none
Meds Lasix 40mg IV bid, ASA, Coumadin 5, Prinivil 10, glucophage 850 bid, glipizide 10 bid, immodium prn
A/P:
- increase lasix to 80 bid
- maintain sao2 > 92% per RT protocol
- no fever and wbc wnl, so cont immodium prn
- dm2: ccm
Resources for EMR-based research at Vanderbilt

The Synthetic Derivative
A de-identified and continuously-updated image of the EMR: 2,358,760 subjects

BioVU
Subjects with DNA: 197,330

- Dense (GWAS-level) genotypes: ~20,000
- Exome chip data: ~36,000
EHR Phenotyping
"Simple" Example: Who has hypertension?
Definition: SBP > 140 or DBP > 90
Our “simple” example: Hypertension

Multiple components are better (and blood pressure is the worst)
The “demonstration project”

- Are genotype-phenotype relations replicated in BioVU?
- Genotype “high-value” SNPs in the first 10,000 samples accrued.
  - 21 established loci (>1 SNP for some)
  - in 5 diseases with known associations:
    - Atrial fibrillation
    - Crohn’s disease
    - Multiple Sclerosis
    - Rheumatoid arthritis
    - Type II Diabetes
- Develop “electronic phenotype algorithms” to identify cases and controls
# RA – Case Definition Evolution

<table>
<thead>
<tr>
<th>#</th>
<th>Definition</th>
<th># Cases</th>
<th>Problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ICD9 codes for RA + Medications (only in problem list)</td>
<td>371</td>
<td>Found incomplete problem lists</td>
</tr>
<tr>
<td>2</td>
<td>Same as above but searched notes</td>
<td>411</td>
<td>Patients billed as RA but actually other conditions, overlap syndromes such as psoriatic arthritis, juvenile RA</td>
</tr>
<tr>
<td>3</td>
<td>Above + require text “rheumatoid arthritis” and small list of exclusions</td>
<td>358</td>
<td>Overlap syndromes with other autoimmune conditions, conditions in which physicians did not agree</td>
</tr>
<tr>
<td>4</td>
<td>Above + exclusion of other inflammatory arthritides</td>
<td>255</td>
<td>PPV = 97%; a few “possible RA” or family history items remained</td>
</tr>
</tbody>
</table>
Final RA case definition

ICD 9 codes (any of the below)

- 714  Rheumatoid arthritis and other inflammatory polyarthropathies
- 714.0  Rheumatoid arthritis
- 714.1  Felty’s syndrome
- 714.2  Other rheumatoid arthritis with visceral or systemic involvement

AND

Medications (any of the below)

methotrexate [MTX][amethopterin] sulfasalazine [azulfidine]; Minocycline [minocin][solodyn]; hydroxychloroquine [Plaquinil]; adalimumab [Humira]; etanercept [Enbrel] infliximab [Remicade]; Gold [myochrysine]; azathioprine [Imuran]; rituximab [Rituxan] [MabThera]; anakinra [Kineret]; abatacept [Orencia]; leflunomide [Arava]

AND

Keywords (any of the below)

rheumatoid [rheum] [reumatoid] arthritis [arthritides] [arthritis] [arthristis] [arthritus] [arthrtis] [arthritis]

= Narrative components
Final RA case definition - 2

**AND NOT**

ICD 9 codes (any of the below)

- 714.30 Polyarticular juvenile rheumatoid arthritis, chronic or unspecified
- 714.31 Polyarticular juvenile rheumatoid arthritis, acute
- 714.32 Pauciarticular juvenile rheumatoid arthritis
- 714.33 Monoarticular juvenile rheumatoid arthritis
- 695.4 Lupus erythematosus
- 710.0 Systemic lupus erythematosus
- 373.34 Discoid lupus erythematosus of eyelid
- 710.2 Sjogren’s disease
- 710.3 Dermatomyositis
- 710.4 Polymyositis
- 555 Regional enteritis
- 555.0 Regional enteritis of small intestine
- 555.1 Regional enteritis of large intestine
- 555.2 Regional enteritis of small/large intestine
- 555.9 Regional enteritis of unspecified site
- 564.1 Irritable Bowel Syndrome
- 135 Sarcoidosis
- 696 Psoriasis and similar disorders
- 696.0 Psoriatic arthropathy
- 696.1 Other psoriasis and similar disorders excluding psoriatic arthropathy
- 696.8 Other psoriasis and similar disorders
- 099.3 Reiter’s disease
- 716.8 Arthropathy, unspecified
- 274.0 Gouty arthropathy
- 358.0 myasthenia gravis
- 358.00 myasthenia gravis without acute exacerbation
- 358.01 myasthenia gravis with acute exacerbation
- 775.2 neonatal myasthenia gravis
- 719.3 Palindromic rheumatism
- 719.30 Palindromic rheumatism, site unspecified
- 719.31 Palindromic rheumatism involving shoulder region
- 719.32 Palindromic rheumatism involving upper arm
- 719.33 Palindromic rheumatism involving forearm
- 719.34 Palindromic rheumatism involving hand
- 719.35 Palindromic rheumatism involving pelvic region and thigh
- 719.36 Palindromic rheumatism involving lower leg
- 719.37 Palindromic rheumatism involving ankle and foot
- 719.38 Palindromic rheumatism involving other specified sites
- 719.39 Palindromic rheumatism involving multiple sites
- 720 Ankylosing spondylitis and other inflammatory spondylopathies
- 720.0 Ankylosing spondylitis
- 720.8 Other inflammatory spondylopathies
- 720.81 Inflammatory spondylopathies in diseases classified elsewhere
- 720.89 Other inflammatory spondylopathies
- 720.9 Unspecified inflammatory spondylopathy
- 721.2 Thoracic spondylosis without myelopathy
- 721.3 Lumbosacral spondylosis without myelopathy
- 729.0 Rheumatism, unspecified and fibrositis
- 340 Multiple sclerosis
- 341.9 Demyelinating disease of the central nervous system unspecified
- 323.9 transverse myelitis
- 710.1 Systemic sclerosis
- 245.2 Hashimoto’s thyroiditis
- 242.0 Toxic diffuse goiter
- 443.0 Raynaud’s syndrome

**AND NOT**

Keywords (any of the below)

- juvenile [juv] rheumatoid [rheum] [reumatoid] [rhumatoid] arthritis [arthritides] [arthritis] [arthristis] [arthritus] [arthritis] [arthritis]
- juvenile [juv] arthritis arthritis [arthritides] [arthritis] [arthristis] [arthritus] [arthritis] [arthritis]
- juvenile chronic arthritis [arthritides] [arthritis] [arthristis] [arthritus] [arthritis] [arthritis]
- juvenile [juv] RA; JRA
- Inflammatory [inflammatory] [inflam] osteoarthritis [osteoarthrosis] [OA]
- Reactive [psoriatic] arthritis [arthropathy] [arthritides] [arthritis] [arthristis] [arthritus] [arthritis] [arthritis]
What we learned - Finding phenotypes in the EMR

Algorithm Development and Implementation

- Identify phenotype of interest
- Case & control algorithm development and refinement
- Manual review; assess precision
- Deploy in BioVU
- Genetic association tests

Billing codes
ICD9 & CPT

Clinical Notes
(NLP - natural language processing)

Labs & test results
NLP

Medications
ePrescribing & NLP

True cases

Identify phenotype of interest

<95%

≥95%
A common general approach to phenotyping

![Diagram showing the process of phenotyping with EHR, definite cases, possible cases, excluded cases, and controls.](image)

- **EHR**: Application of electronic phenotype selection logic
- **Definite cases**
- **Possible cases**
- **Excluded cases** (insufficient evidence)
- **Controls**

Manual review (typically rare diseases)

Case-control genotypic analysis

Validating EMR phenotype algorithms
(Using first 10,000 patients in BioVU)

<table>
<thead>
<tr>
<th>Disease</th>
<th>Methods</th>
<th>Definite Cases</th>
<th>Controls</th>
<th>Case PPV</th>
<th>Control PPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atrial fibrillation</td>
<td>NLP of ECG impressions</td>
<td>168</td>
<td>1695</td>
<td>98%</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>ICD9 codes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CPT codes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crohn’s Disease</td>
<td>ICD9 codes</td>
<td></td>
<td></td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>Medications (NLP)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type 2 Diabetes</td>
<td>ICD9 codes</td>
<td></td>
<td></td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>Medications (NLP)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NLP exclusions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Labs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple Sclerosis</td>
<td>ICD9 codes or text diagnosis</td>
<td>66</td>
<td>1857</td>
<td>87%</td>
<td>100%</td>
</tr>
<tr>
<td>Rheumatoid Arthritis</td>
<td>ICD9 codes</td>
<td>170</td>
<td>701</td>
<td>97%</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>Medications (NLP)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NLP exclusions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Common themes:
Billing codes – 5/5
NLP – 5/5
Meds – 4/5
Labs – 2/5

NLP = Natural language processing
### Results

<table>
<thead>
<tr>
<th>Disease</th>
<th>Marker</th>
<th>Gene / Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atrial fibrillation</td>
<td>rs2200733</td>
<td>Chr. 4q25</td>
</tr>
<tr>
<td></td>
<td>rs10033464</td>
<td>Chr. 4q25</td>
</tr>
<tr>
<td>Crohn's disease</td>
<td>rs11805303</td>
<td>IL23R</td>
</tr>
<tr>
<td></td>
<td>rs17234657</td>
<td>Chr. 5</td>
</tr>
<tr>
<td></td>
<td>rs1000113</td>
<td>Chr. 5</td>
</tr>
<tr>
<td></td>
<td>rs17221417</td>
<td>NOD2</td>
</tr>
<tr>
<td></td>
<td>rs2542151</td>
<td>PTPN22</td>
</tr>
<tr>
<td>Multiple sclerosis</td>
<td>rs3135388</td>
<td>DRB1*1501</td>
</tr>
<tr>
<td></td>
<td>rs2104286</td>
<td>IL2RA</td>
</tr>
<tr>
<td></td>
<td>rs6897932</td>
<td>IL7RA</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>rs6457617</td>
<td>Chr. 6</td>
</tr>
<tr>
<td></td>
<td>rs6679677</td>
<td>RSBN1</td>
</tr>
<tr>
<td></td>
<td>rs2476601</td>
<td>PTPN22</td>
</tr>
<tr>
<td>Type 2 diabetes</td>
<td>rs4506565</td>
<td>TCF7L2</td>
</tr>
<tr>
<td></td>
<td>rs12255372</td>
<td>TCF7L2</td>
</tr>
<tr>
<td></td>
<td>rs12243326</td>
<td>TCF7L2</td>
</tr>
<tr>
<td></td>
<td>rs10811661</td>
<td>CDKN2B</td>
</tr>
<tr>
<td></td>
<td>rs8050136</td>
<td>FTO</td>
</tr>
<tr>
<td></td>
<td>rs5219</td>
<td>KCNJ11</td>
</tr>
<tr>
<td></td>
<td>rs5215</td>
<td>KCNJ11</td>
</tr>
<tr>
<td></td>
<td>rs4402960</td>
<td>IGF2BP2</td>
</tr>
</tbody>
</table>

Ritchie et al., AJHG 2010
Identifying cases with precision from the EMR requires structured and unstructured information.

<table>
<thead>
<tr>
<th></th>
<th>RA</th>
<th>MS</th>
<th>CD</th>
<th>T2D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Had ICD-9 codes:</td>
<td>3.9%</td>
<td>1.8%</td>
<td>1.8%</td>
<td>17.3%</td>
</tr>
<tr>
<td>Met algorithm definition:</td>
<td>2.7%</td>
<td>1.2%</td>
<td>1.6%</td>
<td>9.7%</td>
</tr>
<tr>
<td>Accuracy of ICD9 codes:</td>
<td>69%</td>
<td>66%</td>
<td>89%</td>
<td>56%</td>
</tr>
</tbody>
</table>
Hypothyroidism algorithm

Case medications
levothyroxine, synthroid, levoxyl unithroid, armour thyroid, desicated thyroid, cytome, triostat, liothryronine, synthetic triiodothyronine, liotrix, thyrolar

ICD-9 codes for hypothyroidism
244, 244.8, 244.9, 245, 245.2, 245.8, 245.9

ICD-9 codes for secondary causes of hypothyroidism
244.0, 244.1, 244.2, 244.3

Abnormal lab values
TSH > 5 OR FT4 < 0.5

ICD-9 codes for post surgical or post radiation hypothyroidism
193*, 242.0, 242.1, 242.2, 242.3, 242.9, 244.0, 244.1, 244.2, 244.3, 258*

CPT codes for post radiation hypothyroidism
77261, 77262, 77263, 77280, 77285, 77290, 77295, 77299, 77300, 77301, 77305, 77310, etc.

Case Definition
All three conditions required:
- ICD-9 code for hypothyroidism OR abnormal TSH/FT4
- Thyroid replacement medication use
- Require at least 2 instances of either medication or lab with at least 3 months between the first and last instance of medication and lab

Case Exclusions
Exclude if the following information occurs at any time in the record:
- Secondary causes of hypothyroidism
- Post surgical or post radiation hypothyroidism
- Other thyroid diseases
- Thyroid altering medication

Case Exclusions
Time dependent case exclusions:
- Recent pregnancy TSH/FT4
- Recent contrast exposure

Exclusion keywords
multiple endocrine neoplasia, MEN I, MEN II, thyroid cancer, thyroid carcinoma

Thyroid-altering medications
Phenytoin, Dilantin, Infatabs, Dilantin Kapsels, Dilantin-125, Phenytex, Amiodarone Pacerone, Cordarone, Lithium, Eskalith, Lithobid, Methimazole, Tapazole, Northix, Propylthiouracil, PTU

Pregnancy exclusion ICD 9 codes
Any pregnancy billing code or lab test if all Case Definition codes, labs, or medications fall within 6 months before pregnancy to one 1 year after pregnancy.

Exclusion keywords
optiray, radiocontrast, iodine, omnipaque, visipaque, hypaque, ioversol, diatrizoate, iodixanol, isovue, ipamidol, conray, iothalamate, renografin, sinografin, cystografin, conray, iodipamide
## Hypothyroidism Validation

<table>
<thead>
<tr>
<th>Site</th>
<th>Case PPV (%)</th>
<th>Control PPV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group Health</td>
<td>98</td>
<td>100</td>
</tr>
<tr>
<td>Marshfield</td>
<td>91</td>
<td>100</td>
</tr>
<tr>
<td>Mayo Clinic</td>
<td>82</td>
<td>96</td>
</tr>
<tr>
<td>Northwestern</td>
<td>98</td>
<td>100</td>
</tr>
<tr>
<td>Vanderbilt</td>
<td>98</td>
<td>100</td>
</tr>
<tr>
<td>All sites (weighted)</td>
<td>92.4</td>
<td>98.5</td>
</tr>
</tbody>
</table>

Same algorithm, deployed at five sites

More examples of patients with secondary hypothyroidism without code evidence - only findable via more complex NLP not in algorithm
Hypothyroidism: “No-Genotyping” GWAS

Denny et al., AJHG 2011
Sharing algorithms: PheKB.org
This is not just genetics!

- eMERGE, PCORnet, NIH Collaboratory, PGRN, PGPop
- 381 active users, 48 institutions
- 21 publicly available phenotypes, 67 phenotypes in development
- There are 165 implementations
- Social networking features; versioning; etc.
- Data dictionary and data set validation
What EHR data do we need for research?

Evaluation of EHR data types used in 92 phenotypes posted on PheKB.org (median positive predictive value = 96%)

<table>
<thead>
<tr>
<th>Data modalities or methods</th>
<th>Number of phenotypes utilizing these features</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Public (N=30)</td>
<td>Non-Public (N=62)</td>
</tr>
<tr>
<td>ICD-9 Codes</td>
<td>27</td>
<td>37</td>
</tr>
<tr>
<td>Medications</td>
<td>25</td>
<td>32</td>
</tr>
<tr>
<td>Natural Language Processing</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>CPT Codes</td>
<td>14</td>
<td>24</td>
</tr>
<tr>
<td>Laboratory/test results</td>
<td>14</td>
<td>21</td>
</tr>
</tbody>
</table>
ICD, Meds, and (nonnegated) Text mentions identify different counts of possible cases for different diseases

Wei-WQ et al. JAMIA 2015
ROC curves for ICD-9, primary notes, and specific medications for 10 diseases (1750 reviewed charts)

- ICD codes (AUC: 0.68)
- Primary Notes (0.73)
- Medications (0.54)

Two or more components
How to find cases: general maxims

- ICD9s (and CPT) very useful but not sufficient
  - nearly all algorithms use them
  - good for sensitivity but not necessarily PPV
- Labs
- Meds – marker of disease and severity – not usually that helpful by itself
- NLP – often confirmatory and refining, unless dealing with rare diseases
PPV and sensitivity over 10 diseases by counts of occurrence
**Completed eMERGE GWAS**

All of these are on PheKB and will eventually be public

<table>
<thead>
<tr>
<th>Diseases</th>
<th>Endophenotypes</th>
<th>Selected consortia contributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dementia</td>
<td>PR Duration</td>
<td>Height</td>
</tr>
<tr>
<td>Cataracts</td>
<td>QRS Duration</td>
<td>QTc</td>
</tr>
<tr>
<td>Autoimmune Hypothyroidism</td>
<td>HDL/LDL</td>
<td>Rheum. Arthritis</td>
</tr>
<tr>
<td>Diverticulosis/diverticulitis</td>
<td>height</td>
<td>Myocardial Infarction Genetics Consortium</td>
</tr>
<tr>
<td>Type 2 Diabetes</td>
<td>white blood cell counts</td>
<td>Intl. Mult Sclerosis Genet. Consort.</td>
</tr>
<tr>
<td>Diabetic retinopathy</td>
<td>red blood cell counts</td>
<td></td>
</tr>
<tr>
<td>Herpes zoster</td>
<td>Cardiorespiratory Fitness</td>
<td></td>
</tr>
<tr>
<td>PheWAS</td>
<td>ESR levels</td>
<td>Genomic Investigation of Statin Therapy</td>
</tr>
<tr>
<td>Peripheral Arterial Disease</td>
<td>Platelet levels</td>
<td></td>
</tr>
<tr>
<td>Venous Thromboembolism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glaucoma</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ocular hypertension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abdominal Aortic Aneurysm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colon polyps</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Pharmacogenomic phenotypes**

- ACE inhibitor cough
- Heparin induced thrombocytopenia
- Resistant hypertension
- Drug Induced Liver Injury
- *C. difficile* colitis

**Bold** = GWAS completed with significant results
Are algorithms portable, part II: Rheumatoid Arthritis

- Previously published logistic regression model
  - Developed at Partners Healthcare
  - ICD9, Labs, Meds, NLP
  - Tested at Northwestern and Vanderbilt

- Is the signature the same across
  - Differing healthcare environments and EHR systems?
  - Differing NLP systems (regular expression vs. out-of-the-bag KnowledgeMap)?

<table>
<thead>
<tr>
<th>Site</th>
<th>n</th>
<th>PPV</th>
<th>Sensitivity</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original (Partners)</td>
<td>400</td>
<td>94%</td>
<td>63%</td>
<td>95%</td>
</tr>
<tr>
<td>Northwestern</td>
<td>390</td>
<td>87%</td>
<td>60%</td>
<td>92%</td>
</tr>
<tr>
<td>Vanderbilt</td>
<td>376</td>
<td>95%</td>
<td>57%</td>
<td>95%</td>
</tr>
</tbody>
</table>

Liao et al. Arth Care Res 2010
Carroll et al. JAMIA 2012
Using Machine Learning for Phenotyping

Carroll et al, AMIA 2011
A Demonstration: clopidogrel response

Figure 1: stent placement procedure (by manual review, we identified...)

Phenotype algorithm:
MI defined by ICD-9/lab values/NLP or Intracoronary stent defined by CPT/NLP + Clopidogrel on discharge

Possible case
Between 2 and 730 days from initial events with documented clopidogrel use and evidence of:
MI (ICD9/lab values/NLP) or Revascularization (CPT/NLP) or Stroke (ICD9/NLP) or Death

N = 591
Manual review
Definite case
N = 260
Inadequate DNA sample or patient opt-out
N = 35
Genotyping
N = 225

Possible control
Between 2 and 365–730 days from initial events with documented clopidogrel use and no evidence of:
MI (ICD9/lab values/NLP) or Revascularization (CPT/NLP) or Stroke (ICD9/NLP) or Death

N = 811
Manual review
Definite control
N = 547
Inadequate DNA sample or patient opt-out
N = 79
Genotyping
N = 468

Between 2 and 730 days from initial events with documented clopidogrel use and evidence of:
MI (ICD9/lab values/NLP) or Revascularization (CPT/NLP) or Stroke (ICD9/NLP) or Death

N = 591
Manual review
Definite case
N = 260
Inadequate DNA sample or patient opt-out
N = 35
Genotyping
N = 225

Between 2 and 365–730 days from initial events with documented clopidogrel use and no evidence of:
MI (ICD9/lab values/NLP) or Revascularization (CPT/NLP) or Stroke (ICD9/NLP) or Death

N = 811
Manual review
Definite control
N = 547
Inadequate DNA sample or patient opt-out
N = 79
Genotyping
N = 468
GWAS of ACEI-cough (no prior GWAS)

ACEI-cough
(NLP of allergy sections)

eMERGE phenotype – document ACEI cough allergies in the EMR

Our GWAS:
OR=1.3 [95%CI: 1.2-1.4]

eMERGE replication (same algorithm):
OR=1.32 [1.01-1.70]

European replication – people changed from ACEI to an ARB (reason unknown)

European replication
OR=1.15 [1.01-1.30]
Part 3 – Using NLP to assess and improve Medical Education
Part #1: Assessing Curricula

- LCME and ACGME require increasing documentation of curriculum objectives, coverage, and student patient experiences.
- Accreditation standards specific content, competencies, amount of training, etc. for periodic reviews:
  - Patient case mix
  - Topics taught
  - Response to certain trans-course “hot topics” over time – women’s health, substance abuse, etc.
Traditional Medical Curriculum

First year

Anatomy  Physiology  Histology  Microbiology  …etc
Traditional Medical Education Model

First year

Patients

Second year

Clinical education
Guiding questions

- **Faculty:** “I am teaching about congestive heart disease – what have students already learned about this?”

- **Students:** Studying immunoglobulins, need to find relevant prior concepts like splicing

- **Administrators:** Where do we cover large concepts, like geriatrics?
Courses grouped by years, with calendars specific for each year

Supports multiple “programs” (Med school, residencies, etc)
# 2007/2008 Structure, Function & Development Schedule

[View previous year]

**Course Links**
- [edit]: Gross Anatomy sessions | Cell Biology sessions

**Labels:** Exam or Quiz | Physiology | Gross Anatomy | Cell Biology

## Unit #1

<table>
<thead>
<tr>
<th>Date</th>
<th>Time</th>
<th>Location</th>
<th>Title</th>
<th>Lecturer</th>
<th>Document(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10/25</td>
<td>08:00 AM - 08:30 AM</td>
<td>LH 202</td>
<td>Introduction</td>
<td>Arthur F Dailey</td>
<td></td>
</tr>
<tr>
<td>10/25</td>
<td>08:30 AM - 10:30 AM</td>
<td>LH 202</td>
<td>Introduction to Anatomical Donations Program and In-Lab Memorial svc. Intro to Gross Anatomy Lab, Safety and Technology</td>
<td>Arthur F Dailey</td>
<td></td>
</tr>
<tr>
<td>10/25</td>
<td>10:30 AM - 12:00 PM</td>
<td>LH 202</td>
<td>Layered &amp; Segmented Structure of body; Intro to Nerves &amp; Nerve Classification; Simple Spinal n.</td>
<td>Arthur F Dailey</td>
<td></td>
</tr>
<tr>
<td>10/25</td>
<td>01:00 PM - 02:00 PM</td>
<td>LH 202</td>
<td>(Embryo) Neuromuscular Development 1</td>
<td>Lillian B Nannya</td>
<td></td>
</tr>
<tr>
<td>10/25</td>
<td>02:00 PM - 05:00 PM</td>
<td>11S 10th Floor</td>
<td>GA Lab: Removal of Skin and Subcutaneous Tissue of Back; Superficial Muscles of the Back and Cervical Nerve XI</td>
<td>Arthur F Dailey</td>
<td></td>
</tr>
<tr>
<td>10/26</td>
<td>08:00 AM - 09:00 AM</td>
<td>LH 202</td>
<td>(Embryo) Neuromuscular Development 2</td>
<td>Lillian B Nannya</td>
<td></td>
</tr>
<tr>
<td>10/26</td>
<td>09:00 AM - 10:00 AM</td>
<td>LH 202</td>
<td>Types of Muscle Action and Movements; Freely Moveable Joints</td>
<td>John S Halle</td>
<td></td>
</tr>
<tr>
<td>10/26</td>
<td>10:00 AM - 12:00 PM</td>
<td>11S 10th Floor</td>
<td>GA Lab: Scapular Region</td>
<td>Arthur F Dailey</td>
<td></td>
</tr>
<tr>
<td>10/29</td>
<td>08:00 AM - 09:00 AM</td>
<td>LH 202</td>
<td>Vertebral Column; Postural Muscles; Spinal Cord and Its Environment</td>
<td>Arthur F Dailey</td>
<td></td>
</tr>
<tr>
<td>10/29</td>
<td>09:00 AM - 10:00 AM</td>
<td>LH 202</td>
<td>Neuromuscular Phys #1: Membrane Transport; Fluid Compartments; Osmosis</td>
<td>Al George</td>
<td></td>
</tr>
<tr>
<td>10/29</td>
<td>01:00 PM - 02:00 PM</td>
<td>11S 10th Floor</td>
<td>GA Lab: Deep Back (Perform Laminectomy)</td>
<td>Arthur F Dailey</td>
<td></td>
</tr>
<tr>
<td>10/30</td>
<td>08:00 AM - 09:00 AM</td>
<td>LH 202</td>
<td>Overview of Lymphatic System; Principles of Collateral Circulation</td>
<td>Lillian B Nannya</td>
<td></td>
</tr>
<tr>
<td>10/30</td>
<td>09:00 AM - 11:00 AM</td>
<td>11S 10th Floor</td>
<td>GA Lab: Complete Dissection of Deep Back; Pectoral Region, Including Removal of Skin from Arm (Excluding Female Breast)</td>
<td>Arthur F Dailey</td>
<td></td>
</tr>
</tbody>
</table>
Faculty upload native formats (e.g., PowerPoint) and KM creates other formats automatically.

Objectives

Get general idea of cell organization
- Describe the major cellular components

Understand what changes in organelles tell you about dynamics of cells

Develop global perspective on cell organelles
- Structure: molecular organization and appearance
- How structure relates to function

Organization of Eucaryotic Cells
- Cathleen Pettipher, Ph.D.
- Molecular Foundations of Medicine
- August 20, 2007

Common Ancestor for all Living Organisms

Shared Mechanisms of Biological Function
- Share a common genetic code and store the information in the form of DNA
- Information from DNA to proteins: one gene, one protein
Document Processing

Document uploaded by lecturer, placed in queue

Document Conversion Server pulls next document off queue, converts to HTML and Text

HTML & PDF versions

Text version placed in queue

KM Concept Identifier

Web server

PDF, MS Word, WordPerfect, HTML, PowerPoint, etc

CUIS
Searching for "urinary incontinence" (ex: heart failure and digoxin)

Analysis of first ~60,000 searches showed that 85% were medical concepts
Documents containing the concept
“urinary incontinence”

VMS I
- Biochemistry (1 document)
- Gross Anatomy (3 documents)
- Microbiology (1 document)

VMS II
- Lab Diagnosis (3 documents)
- Medical Neuroscience (2 documents)
  - Normal Pressure Hydrocephalus [first hit] 4 hits
- 1 more...
- Pathology (1 document)
- Pharmacology (4 documents)
- Physical Diagnosis (1 document)

VMS III
- Pediatrics (1 document)
- Psychiatry (1 document)
- Surgery (1 document)
  - Urology; PSA; scrotal pain; urinary complaints [first hit] Harriette Scarpero 3 hits

VMS IV
- Emergency Medicine (2 documents)
  - Neurologic Emergencies [first hit] Susan Marlow 2 hits
  - 1 more...
Slide #92: "Urinary Incontinence"

- Stress incontinence
- Urge incontinence
- Mixed incontinence

Slide #93: "Incontinence"

The involuntary loss of urine

- Stress incontinence (SUI) - occurs with increases in intraabdominal pressure
- Urge incontinence (UUI) - leakage that is preceded by an intense uncontrolable urge to void
- Mixed incontinence - both SUI and UUI

Slide #94: "Components of a Thorough Evaluation"

- Detailed history*
- Questionnaire*
- Physical examination*
- Other testing: UA, voiding diaries*, pad test*, uroflow, postvoid residual (PVR), urodynamics*

* Recommended minimal standards for pretx eval by AUA
Three definitions for “CHF” – the system disambiguates each occurrence of “CHF” into one of these three matches when in documents

Definitions for "chf"

KnowledgeMap Definitions:

**Congenital hepatic fibrosis** [Approx. 6 documents in KM]: no definition

**Congestive heart failure** [Approx. 281 documents in KM]: Weakness of the heart muscle that leads to a buildup of fluid in body tissues.

**Hemorrhagic Fever, Crimean** [no documents]: A severe, often fatal disease in humans caused by the Crimean-Congo hemorrhagic fever virus (HEMORRHAGIC FEVER VIRUS, CRIMEAN-CONGO).
CONGESTIVE HEART FAILURE

Heart failure exists when the heart does not provide adequate blood flow to the tissues. This state causes congestion of the tissues, leading to swelling and shortness of breath.

It produces the clinical syndrome of dyspnea, peripheral edema, and pulmonary edema. CHF is the clinical state in which an abnormality of myocardial function is responsible for the failure of the heart to pump at a rate commensurate with the requirements of metabolizing tissues. CHF is encountered frequently by the clinician; it accounts for 2% of all hospital admissions and carries a 5-year survival rate of <50%.

1. CAUSES OF CONGESTIVE HEART FAILURE

A. Myocardial Disease (pump defect)
1. Coronary heart disease -- myocardium is impaired by ischemia
2. Cardiomyopathy -- intrinsic myocardial defect
3. Infiltrative diseases:
   amyloidosis
   sarcoidosis
   myocarditis

B. Excessive Workload due to:
1. Increased resistance to ejection which can be due to pressure overload, hypertension, aortic or pulmonary stenosis, or hypertrophic cardiomyopathy.
2. Increased stroke volume, volume overload which can be due to aortic, mitral or tricuspid insufficiency or congenital left-to-right shunts.
3. Increased body demands (high output failure), can occur with thyrotoxicosis, anemia, pregnancy, or arteriovenous fistulas (abnormal shunt between an artery and vein which increases venous return and decreases oxygen to shunted area)
Infantile Polycystic Kidney Disease (Autosomal Recessive Polycystic Kidney Disease, ARPKD)

I. General and clinical features:

A. Incidence and relationship to congenital hepatic fibrosis: Infantile polycystic kidney disease is closely associated with congenital hepatic fibrosis (CHF). CHF is an important cause of portal hypertension in children and adolescents. In general, in patients who present as neonates and infants, the clinical picture is dominated by renal failure. Patients who present later tend to have liver problems as the major clinical feature. Although at one time thought to represent distinct disorders, different affected members in the same family may present at different ages with either predominant renal or liver abnormalities, attesting to the underlying genetic relationship of these diseases.

Infantile polycystic kidney disease is inherited in an autosomal recessive manner (i.e., parents are not affected), with the reported incidence varying from 1:6000 to 1:40,000.

B. Clinical presentation: Can present at any age but is usually diagnosed in the first 6 months of life or shortly after birth.

- Can be suspected during prenatal ultrasound and is often associated with oligohydramnios, which is a condition where the amniotic fluid is reduced.

- Presentation shortly after birth can be due to dehydration, polyuria, and polydipsia.

- Presentation later in infancy can be due to dehydration, polyuria, and polydipsia.

- The kidneys may be sufficiently enlarged to result in a palpable mass. Deteriorating renal function is inevitable, and patients may present later with renal failure/uremia/hypertension.

As implied above, patients presenting in later childhood and early adulthood usually present with signs of liver involvement, particularly portal hypertension, which may result in hepatosplenomegaly and bleeding esophageal varices.

II. Pathogenesis:

Recent data has mapped a gene for ARPKD to the short arm of chromosome 6. The specific gene has not been identified.
How do we find broad concepts like “geriatrics” or “women’s health”?
180 concepts related to "Geriatrics"

1. Deselect Any Incorrect Terms:
   - Geriatrics [Biomedical Occupation or Discipline] (130 expanded concepts)
   - Elderly [Temporal Concept] (89 expanded concepts)

2. Type Any Additional Terms (one per line):

3. Expand Top 15 Lectures from 2006/2007 Course Year Documents

4. Submit:
   - Expand: See more concepts
   - Finalize: Get Document Matches

Options: Show MeSH Concepts Only
Concepts related to “Geriatrics” (using UMLS MRREL relationships)

After building, you can save, export to Excel, periodically use to run reports as CSV/Excel
Searching for documents matching those 180 Geriatric concepts.
Lecture: Aging and Alzheimer's Disease I

NORMAL AGING OF THE BRAIN/ALZHEIMER'S DISEASE

Insight into aging of the brain and Alzheimer's disease is the result of a massive research effort; work on Alzheimer's disease, in particular, represents a rapidly evolving area of research; new and often conflicting data are generated daily; the purpose of this set of lectures is to provide you with the requisite background information necessary to allow you to continue to assimilate new findings as they are generated, and to provide you with a synopsis of both an historical and current (albeit changing) understanding of Alzheimer's disease.

* Recommended reading:

"DeBaggio, Thomas. Losing My Mind. The Free Press, 2002"

"Bayley, John. Elegy for Iris. Picador Press, 1999"

"Jozefowicz and Holloway. Case 26 "The Gopher Hunter""

I. Maturational events which occur normally in the nervous system
How well does KM find metaconcepts?

- Identified gold standard set of 380 documents as containing high, medium or low relevance to each topic
- Used KM to generate a variable number of subconcepts for each broad concept and calculated a relevance score for each document.

<table>
<thead>
<tr>
<th>Topic</th>
<th>ROC area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetics</td>
<td>0.98</td>
</tr>
<tr>
<td>Women’s Health</td>
<td>0.93</td>
</tr>
<tr>
<td>Dermatology</td>
<td>0.95</td>
</tr>
<tr>
<td>Radiology</td>
<td>0.97</td>
</tr>
</tbody>
</table>
Finding broad curricular topics

- Used for LCME, creating/rearranging courses, revising curriculum

Using to infuse Geriatrics in the curriculum:

![Graph showing # Lectures with Geriatrics over different years]
Part #2: Assessment in Clinical Years

- Testing based: USMLE, NCLEX, Residency Board Exams
- Experience Based:
  - ACGME and RRC
  - Nursing requirements
- Both current methods tend to aggregate at high levels
- Experience is an important part of competency
Learning Portfolio – leveraging EMR to capture experience

Trainee creates note in the EMR

Electronic Medical Record

Clinical Portfolio

Appropriate mentors emailed (students)

Mentors evaluate and:
1. Give feedback
2. Assign learning objectives

Automatically Log Procedures

Use Natural Language Processing to find note content

Database of concepts
Teachers have students/trainees they "mentor". What documents/patients they see (for patient confidentiality) is driven by:

- Mentorship type
- Timeframe of mentorship

<table>
<thead>
<tr>
<th>Student</th>
<th>Mentor Type</th>
<th>Date Range</th>
<th>Patients</th>
<th>Procedures</th>
<th>Notes</th>
<th>Reflections</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cronk</td>
<td>Master Clinical Teacher</td>
<td>6/1/2005 - 5/15/2007</td>
<td>56</td>
<td>0</td>
<td>105</td>
<td>0</td>
</tr>
<tr>
<td>Davis</td>
<td>Master Clinical Teacher</td>
<td>6/1/2005 - 5/15/2007</td>
<td>56</td>
<td>0</td>
<td>105</td>
<td>0</td>
</tr>
<tr>
<td>Eubank</td>
<td>Master Clinical Teacher</td>
<td>6/1/2005 - 5/15/2007</td>
<td>56</td>
<td>0</td>
<td>105</td>
<td>0</td>
</tr>
<tr>
<td>Frazier</td>
<td>Master Clinical Teacher</td>
<td>6/1/2005 - 5/15/2007</td>
<td>56</td>
<td>0</td>
<td>105</td>
<td>0</td>
</tr>
<tr>
<td>Goss</td>
<td>Master Clinical Teacher</td>
<td>6/1/2005 - 5/15/2007</td>
<td>56</td>
<td>0</td>
<td>105</td>
<td>0</td>
</tr>
<tr>
<td>Hensley</td>
<td>Master Clinical Teacher</td>
<td>6/1/2005 - 5/15/2007</td>
<td>56</td>
<td>0</td>
<td>105</td>
<td>0</td>
</tr>
<tr>
<td>Ivey</td>
<td>Master Clinical Teacher</td>
<td>6/1/2005 - 5/15/2007</td>
<td>56</td>
<td>0</td>
<td>105</td>
<td>0</td>
</tr>
<tr>
<td>Jansen</td>
<td>Master Clinical Teacher</td>
<td>6/1/2005 - 5/15/2007</td>
<td>56</td>
<td>0</td>
<td>105</td>
<td>0</td>
</tr>
<tr>
<td>Kropf</td>
<td>Clerkship Director</td>
<td>5/4/2006 - 5/15/2006</td>
<td>245</td>
<td>0</td>
<td>657</td>
<td>N/A</td>
</tr>
<tr>
<td>Lambert</td>
<td>Clerkship Director</td>
<td>8/15/2006 - 11/20/2006</td>
<td>126</td>
<td>44</td>
<td>207</td>
<td>N/A</td>
</tr>
<tr>
<td>Lubeck</td>
<td>Clerkship Director</td>
<td>7/7/2006 - 12/16/2006</td>
<td>125</td>
<td>43</td>
<td>266</td>
<td>N/A</td>
</tr>
<tr>
<td>Moberg</td>
<td>Inpatient Attending</td>
<td>7/15/2007 - 8/1/2007</td>
<td>20</td>
<td>0</td>
<td>55</td>
<td>N/A</td>
</tr>
<tr>
<td>Nowak</td>
<td>Resident/Fellow</td>
<td>8/26/2005 - 10/21/2005</td>
<td>45</td>
<td>0</td>
<td>76</td>
<td>N/A</td>
</tr>
<tr>
<td>Petkovic</td>
<td>Resident/Fellow</td>
<td>3/1/2006 - 5/24/2006</td>
<td>70</td>
<td>2</td>
<td>108</td>
<td>N/A</td>
</tr>
</tbody>
</table>
### Existing Notes

You have **653** notes.

<table>
<thead>
<tr>
<th>Date of Service</th>
<th>Patient</th>
<th>Note Type</th>
<th>Submit Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007-05-01</td>
<td></td>
<td>Medical Student Brief Operative Note</td>
<td>2007-05-29</td>
</tr>
<tr>
<td>2007-05-03</td>
<td></td>
<td>Internal Medicine Clinic Visit</td>
<td>2007-05-23</td>
</tr>
<tr>
<td>2007-05-04</td>
<td></td>
<td>Outpatient History and Physical</td>
<td>2007-05-23</td>
</tr>
<tr>
<td>2007-05-05</td>
<td></td>
<td>Internal Medicine Clinic Visit</td>
<td>2007-05-23</td>
</tr>
<tr>
<td>2007-05-06</td>
<td></td>
<td>Internal Medicine Clinic Visit</td>
<td>2007-05-23</td>
</tr>
<tr>
<td>2007-05-07</td>
<td></td>
<td>Internal Medicine Clinic Resident Acute Clinic Visit</td>
<td>2007-05-23</td>
</tr>
<tr>
<td>2007-05-08</td>
<td></td>
<td>Speech and Language Treatment Plan Report</td>
<td>2007-05-21</td>
</tr>
<tr>
<td>2007-05-09</td>
<td></td>
<td>Internal Medicine Clinic Visit</td>
<td>2007-05-16</td>
</tr>
<tr>
<td>2007-05-10</td>
<td></td>
<td>Internal Medicine Clinic Visit</td>
<td>2007-05-16</td>
</tr>
<tr>
<td>2007-05-11</td>
<td></td>
<td>Internal Medicine Clinic Visit</td>
<td>2007-05-16</td>
</tr>
<tr>
<td>2007-05-12</td>
<td></td>
<td>Internal Medicine Clinic Resident Acute Clinic Visit</td>
<td>2007-05-16</td>
</tr>
<tr>
<td>2007-05-13</td>
<td></td>
<td>History and physical</td>
<td>2007-05-16</td>
</tr>
<tr>
<td>2007-05-16</td>
<td></td>
<td>Internal Medicine Clinic Visit</td>
<td>2007-05-09</td>
</tr>
<tr>
<td>2007-05-17</td>
<td></td>
<td>Internal Medicine Clinic Visit</td>
<td>2007-05-09</td>
</tr>
<tr>
<td>2007-05-18</td>
<td></td>
<td>Internal Medicine Clinic Visit</td>
<td>2007-05-09</td>
</tr>
<tr>
<td>2007-05-19</td>
<td></td>
<td>Internal Medicine Clinic Visit</td>
<td>2007-05-09</td>
</tr>
<tr>
<td>2007-05-20</td>
<td></td>
<td>Internal Medicine Clinic Patient Acute Clinic Visit</td>
<td>2007-05-09</td>
</tr>
<tr>
<td>2007-05-21</td>
<td></td>
<td>Procedure Note (Laceration Repair)</td>
<td>2007-05-04</td>
</tr>
<tr>
<td>2007-05-23</td>
<td></td>
<td>Outpatient History and Physical</td>
<td>2007-05-02</td>
</tr>
<tr>
<td>2007-05-24</td>
<td></td>
<td>Internal Medicine Clinic Resident Acute Clinic Visit</td>
<td>2007-05-02</td>
</tr>
<tr>
<td>2007-05-26</td>
<td></td>
<td>Internal Medicine Clinic Visit</td>
<td>2007-04-25</td>
</tr>
<tr>
<td>2007-05-28</td>
<td></td>
<td>Internal Medicine Clinic Visit</td>
<td>2007-04-25</td>
</tr>
<tr>
<td>2007-05-29</td>
<td></td>
<td>Internal Medicine Clinic Visit</td>
<td>2007-04-25</td>
</tr>
</tbody>
</table>
Provides “one-stop shopping” for:
- Term Definitions
- Drug information (Epocrates, Lexi-Comp)
- Medical references (UpToDate, PubMed, MedlinePlus)
- All curriculum content (VUSM, MPH, residencies, etc)
- Additional local resources (POGOe, Geriatric Review Syllabus)
- Other EBM resources and Google
# 25 Core Clinical Problems (CCP)

<table>
<thead>
<tr>
<th>Abdominal pain</th>
<th>Headache</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abnormal uterine bleeding</td>
<td>Jaundice</td>
</tr>
<tr>
<td>Abnormal vaginal discharge</td>
<td>Loss of consciousness</td>
</tr>
<tr>
<td>Abnormalities of mood</td>
<td>Obesity</td>
</tr>
<tr>
<td>Altered mental status</td>
<td>Pelvic pain</td>
</tr>
<tr>
<td>Back pain</td>
<td>Pharyngitis</td>
</tr>
<tr>
<td>Breast disease</td>
<td>Rash</td>
</tr>
<tr>
<td>Chest pain</td>
<td>Seizures</td>
</tr>
<tr>
<td>Cough</td>
<td>Shock</td>
</tr>
<tr>
<td>Dysuria</td>
<td>Shortness of breath</td>
</tr>
<tr>
<td>Fever</td>
<td>Substance abuse</td>
</tr>
<tr>
<td>GI bleeding</td>
<td>Trauma</td>
</tr>
<tr>
<td></td>
<td>Weight loss</td>
</tr>
</tbody>
</table>
Student view of how many VC3 topics they’ve completed. (Teachers can see this also.)

Mapping of a note to a VC3 topic happens manually and automatically for high scoring documents.

<table>
<thead>
<tr>
<th>Learning Objective</th>
<th>Date Recorded</th>
<th>Event Recorded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abdominal Pain</td>
<td>10/17/2008</td>
<td>Pediatric Surgery Consultation Note</td>
</tr>
<tr>
<td></td>
<td>11/1/2008</td>
<td>Medical Student Admission History and Physical</td>
</tr>
<tr>
<td>Altered Mental Status</td>
<td></td>
<td>None recorded</td>
</tr>
<tr>
<td>Back Pain</td>
<td>10/17/2008</td>
<td>Return Clinic Visit Progress Note</td>
</tr>
<tr>
<td>Chest Pain</td>
<td>5/9/2008</td>
<td>Medical Student Admission History and Physical</td>
</tr>
<tr>
<td></td>
<td>5/23/2008</td>
<td>Medical Student Admission History and Physical</td>
</tr>
<tr>
<td></td>
<td>10/19/2008</td>
<td>Medical Student Admission History and Physical</td>
</tr>
<tr>
<td>Breast Disease</td>
<td></td>
<td>None recorded</td>
</tr>
<tr>
<td>Coma</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cough</td>
<td>10/20/2008</td>
<td>Medical Student Admission History and Physical</td>
</tr>
<tr>
<td></td>
<td>10/20/2008</td>
<td>Medical Student Admission History and Physical</td>
</tr>
<tr>
<td>Depression</td>
<td>11/2/2008</td>
<td>Medical Student Progress Note History and physical</td>
</tr>
<tr>
<td></td>
<td>11/2/2008</td>
<td></td>
</tr>
<tr>
<td>Dysuria</td>
<td>11/2/2008</td>
<td>Pediatric Infectious Disease Initial Consultation</td>
</tr>
<tr>
<td></td>
<td>11/2/2008</td>
<td>Medical Student Admission History and Physical</td>
</tr>
<tr>
<td></td>
<td>11/2/2008</td>
<td>Clinic Visit</td>
</tr>
<tr>
<td>Fever</td>
<td>10/17/2008</td>
<td>Medical Student Admission History and Physical</td>
</tr>
<tr>
<td></td>
<td>10/17/2008</td>
<td>Progress Note Daily Progress Note</td>
</tr>
<tr>
<td>Gastrointestinal Bleeding</td>
<td></td>
<td>None recorded</td>
</tr>
<tr>
<td>Heart Murmurs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jaundice</td>
<td>11/2/2008</td>
<td>History and physical</td>
</tr>
<tr>
<td></td>
<td>11/2/2008</td>
<td>Progress Note Daily Progress Note</td>
</tr>
<tr>
<td>Menstrual abnormalities</td>
<td></td>
<td>None recorded</td>
</tr>
<tr>
<td>Mood Disorder</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pelvic Pain</td>
<td></td>
<td>None recorded</td>
</tr>
<tr>
<td>Pharyngitis</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Searching for relevant notes matching core objective “Back Pain”

He discussed these concepts

...in these sections
Part #3: Evaluating and integrating

Study 1: Curriculum vs. Notes

Learning Portfolio (clinical notes)

Concept database

KnowledgeMap (curriculum documents)

Natural language processing

Clinical content filters
- 25 VC3 problems (CCPs)
- 7 types of infections

Compare content, identify discrepancies

CC: abdominal pain
This is a 78yo with PMH significant for who presents with back pain after a fall…
Coverage of VC3 Topics

- Abdominal Pain
- Abnl Uterine Bleeding
- Abnl Mood
- Abnl Weight Loss
- Altered Mental Status
- Back Pain
- Breast Disease
- Cough
- Chest Pain
- Dysuria
- Fever
- GI Bleed

Concept density

- Clinical Notes
- Curriculum documents

- Jaundice
- Loss of Consciousness
- Observe
- Obesity
- Pharyngitis
- Pelvic Pain
- Seizures
- Rash
- Shortness of Breath
- Shock
- Substance Abuse
- Trauma
- Vaginal Discharge

Denny et al. AMIA 2010

300k student notes
15k lecture documents
Coverage of Infectious Diseases

Curriculum Documents
- Bacteria: 38%
- Virii: 27%
- Fungi: 11%
- Tick-borne: 7%
- Protozoan: 6%
- Helminths: 5%
- Prions: 4%

Clinical Notes
- Bacteria: 60.8%
- Virii: 22.2%
- Fungi: 11.8%
- Tick-borne: 2.1%
- Helminths: 1.3%
- Protozoans: 1.7%
- Prions: 0.1%

Denny et al. AMIA 2010
**Study 2: Automated Education Advisors**

- **Student types a note in the EMR**
- **Database of Note Concepts**
- **NLP**
- **NLP Rules Engine**
- **Customized Emails**
  - Key facts about the diagnosis
  - Whether or not they met some criteria
  - Links to key references about the topic (on KM)
  - Links to documents most relevant to their note

- **Current Email Advisors:**
  - Advanced directives (pts > 65, if they don’t mention them)
  - Altered mental status (must say AMS concept in key section of note)

Denny et al. JBI 2015
CHIEF COMPLAINT: confusion, weakness, and lethargy
HISTORY OF PRESENT ILLNESS: Mrs. X is a 70 year old female with metastatic undifferentiated carcinoma, likely lung in origin, who was recently discharged from the hospital s/p left femoral fracture and biopsy due to a fracture who now presents with increasing confusion, weakness, and lethargy.

PHYSICAL EXAMINATION: General: waxing and waning alertness,

SUMMARY: This is a 72 year old female with metastatic lung carcinoma admitted for delirium most likely secondary to hypercalcemia.

ASSESSMENT AND PLAN:
1. Hypercalcemia Hyperparathyroidism... malignancy...
6. Disp -Will keep hospitalized until altered mental status improves...
CHIEF COMPLAINT: confusion, weakness, and lethargy

HISTORY OF PRESENT ILLNESS: Mrs. X is a 70 year old female with metastatic undifferentiated carcinoma, likely lung in origin, who was recently discharged from the hospital s/p left femoral fracture and biopsy due to a fracture who now presents with increasing confusion, weakness, and lethargy.

PHYSICAL EXAMINATION: General: waxing and waning alertness,

SUMMARY: This is a 72 year old female with metastatic lung carcinoma admitted for delirium most likely secondary to hypercalcemia.

ASSESSMENT AND PLAN:
1. Hypercalcemia Hyperparathyroidism... malignancy...
6. Disp -Will keep hospitalized until altered mental status improves...
CHIEF COMPLAINT: confusion, weakness, and lethargy

HISTORY OF PRESENT ILLNESS: Mrs. X is a 70 year old female with metastatic undifferentiated carcinoma, likely lung in origin, who was recently discharged from the hospital s/p left femoral fracture and biopsy due to a fracture who now presents with increasing confusion, weakness, and lethargy.

PHYSICAL EXAMINATION: General: waxing and waning alertness

SUMMARY: This is a 72 year old female with metastatic lung carcinoma admitted for delirium most likely secondary to hypercalcemia.

ASSESSMENT AND PLAN:
1. Hypercalcemia
2. Hyperparathyroidism
3. Malignancy
4. Diabetes Mellitus
5. Chronic Kidney Disease
6. Disp -Will keep hospitalized until altered mental status improves

You are getting this email as part of a project to improve your understanding of altered mental status. This email is generated based on your note: Medical Student Admission History and Physical, written on 2011-01-15 19:42:15.

Key facts about Altered Mental Status:
• The differential diagnosis of altered mental status is extensive including dementia, delirium, substance induced, drug side effects, infection, intracranial lesions or strokes, trauma, and metabolic entities such as liver disease or hypoglycemia.
• Alzheimer’s disease, vascular dementia, and dementia with Lewy bodies are the most common forms of degenerative dementias seen in late life.

KM documents most like yours:
• Typical Laboratory Results in the Differential Diagnosis of Hypercalcemia | Joshua Charles Denny | Geriatrics Review Syllabus (Geriatrics)
• Hypercalcemia | Natasha Janelle Schneider | Outpatient Medicine Curriculum (Core Lecture Series)
• Fluid Management for Students | Kyle Bertram Brothers | Pediatrics (VMS III)
• Pharmacological Concepts | Joseph A Awad | Pharmacology (VMS II)

Other searches that may be relevant to this patient:
• Differential diagnosis of metabolic (liver ds, electrolytes, glucose abnormalities) as causes of AMS. (4 overlapping concepts)
• Differential diagnosis of delerium as a cause of AMS. (2 overlapping concepts)
• Signs and symptoms of AMS (2 overlapping concepts)
• Evaluation of AMS (1 overlapping concepts)
### How does it actually work?

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>name</td>
<td>Altered Mental Status</td>
</tr>
<tr>
<td>classifications</td>
<td>VMS3, VMS4 (type of people to evaluate)</td>
</tr>
<tr>
<td>eval_saved_search_num</td>
<td>446 (list of CUIs related to AMS, built on website)</td>
</tr>
<tr>
<td>eval_wordsearch</td>
<td>AMS, altered mental status (extra words to search, nonnegated)</td>
</tr>
<tr>
<td>eval_limit_note_sections</td>
<td>1 (searches only “high value” note sections)</td>
</tr>
<tr>
<td>feedback_saved_searches</td>
<td>426,426,428,429,430,431...</td>
</tr>
<tr>
<td>feedback_saved_searches_programs</td>
<td>1,2,4 (what curricula to use; documents are ranked via TF-IDF from very broadly defined AMS CUIs in the saved searches)</td>
</tr>
<tr>
<td>email_matching_concepts</td>
<td>TRUE (tell the student what we found in their note)</td>
</tr>
<tr>
<td>min_age</td>
<td>18</td>
</tr>
<tr>
<td>max_age</td>
<td>9999</td>
</tr>
<tr>
<td>min_score_to_keep</td>
<td>1</td>
</tr>
<tr>
<td>once_per_patient</td>
<td>1</td>
</tr>
<tr>
<td>km_docs_to_send</td>
<td>74 (a list of highly relevant AMS resources)</td>
</tr>
<tr>
<td>email_subjectline</td>
<td>Altered Mental Status</td>
</tr>
<tr>
<td>email_header</td>
<td>You are getting this email as part of a project to improve your learning on &lt;b&gt;altered mental status&lt;/b&gt;.</td>
</tr>
<tr>
<td>email_instruction</td>
<td>Key facts about Altered Mental Status:&lt;ul&gt; ...</td>
</tr>
</tbody>
</table>
Summary

• EHR-linked DNA biobanks can be used for genomic and pharmacogenomic discovery. They can be cost efficient and fast.

• Best algorithms to find phenotypes include codes, labs, meds, and/or NLP through combination of Boolean and/or machine learning approaches – these algorithms are placed on PheKB

• NLP is often confirmatory for phenotypes

• NLP can also be used to improve cataloging of medical education content and tracking of trainee experiences