NLP in Medical reports

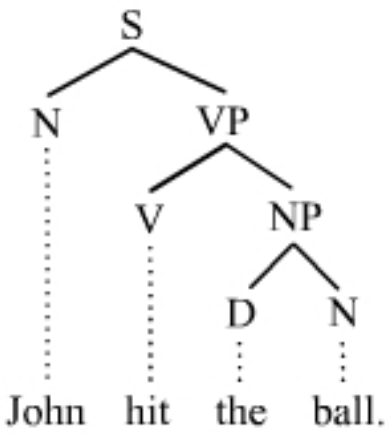
George Cernile, Manager A.I. Technology Group

Progression of A.I.

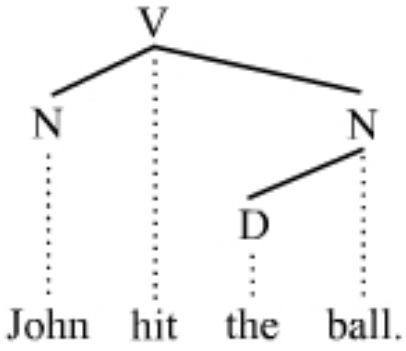
- Natural language processing
  - 1970s Formal grammars (Chomsky) – hand coded rules
- Direct coding of human expertise
  - Heuristics
  - Rule based systems
  - Semantic analysis and networks
  - Knowledge acquisition and representation
- Machine learning approaches
  - Association (clustering)
  - Supervised and unsupervised parts of speech tagging (by induction)
  - Discovery of patterns - Neural nets, Deep learning

Assigning structure to language

- Formal grammars – the study of semantics or interpretations
  - Idealizations of natural languages



Constituency-based parse tree



Dependency-based parse tree

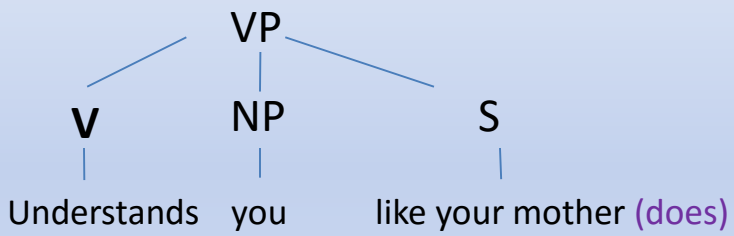
Many models exist – choose one that works for your problem

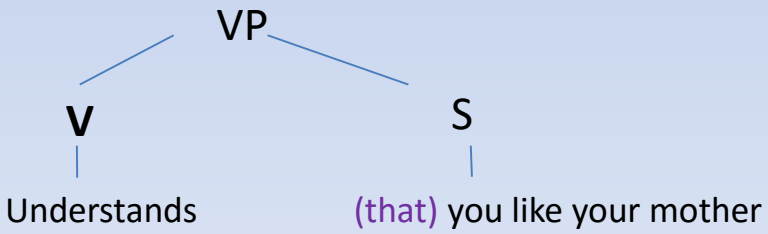
Many formal techniques

Grammars and Parsers
Statistical Methods
Probabilistic parsing
N- Grams
Context sensitive
Context free parsing
Lexicalized parsing
Named entities (and relationships)
Context determination
Parts of speech tagging
Word sense disambiguation
Vector space representation

Example: Phrase Structure Tree

``At last, a computer that understands you like your mother”





Different structures lead to different interpretations

NLP Difficulties

Computer processing of language is a difficult task.

- Ambiguity

"flying planes can be dangerous"

- Need background knowledge to determine context
  - Pilot at risk?
  - Danger to people on ground?
  - Plane is an airplane? A geometric object? A woodworking tool?
  - Is "can" a verb or a noun?

Semantic Analysis in Medical reports

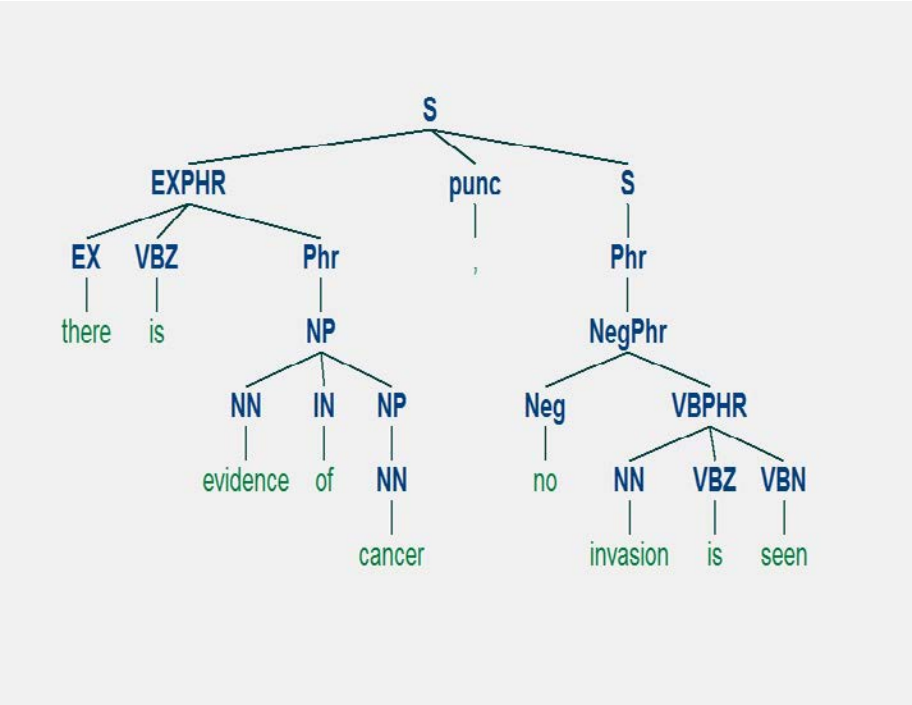
"there is evidence of cancer, no invasion is seen"

In reality:

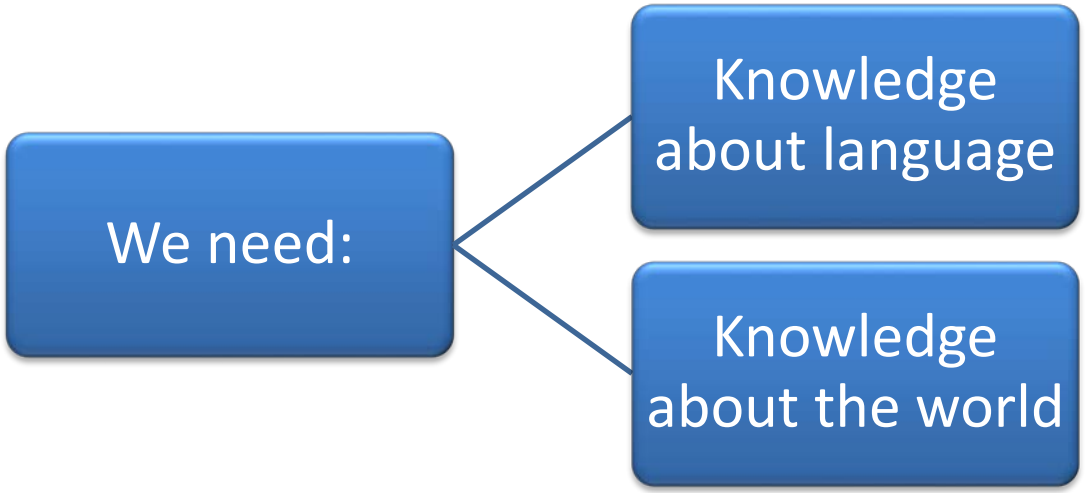
- The use of grammar is inconsistent
- Data is often contained in ad-hoc lists or tables or other creative structures
- Uses specialized acronyms that require context understanding such as "ALL", or "CA"
- Uses data groupings such as in TNM values or Gleason score

Context is difficult for machines when grammar is not well formed!

What about poorly structured sentences? - These approaches are very fragile.



NLP



There is a Knowledge Bottleneck in NLP

Deep analysis is often traded for robust approximations

Empirical or heuristic approach seems to be best for real world applications.

We tested several NLP toolkits

- Apache Open NLP
- Stanford Parser & CoreNLP
- UIMA: Unstructured Information Management Application
- Python NLTK (Natural Language ToolKit)
- Link Grammar - Carnegie Mellon University
- Word2vec *vector('king') - vector('man') + vector('woman')* is close to *vector('queen')*

- Grammar parsers tend to make more mistakes when parsing more complex sentences.
- Unstructured sentences are a hit-and-miss
- Still require specialized lexicons (medical)

Machine learning approaches

Machine learning approaches tackle different types of problems

- Large data sets now available (eg: the internet)
- Statistical modeling is based on observation, not first principles
- New algorithms and training approaches have fueled renewed interest

- Supervised vs non-supervised
- Regression techniques for classification
- Perceptron algorithms
- Generate trees to minimize information entropy (disorder)
- Clustering (class induction)
- Feature detectors
- Probabilistic context-free grammars
- Hidden Markov models
- Word vector representation

- Deep learning is a new way to train many-layer neural networks
- Vector representations of words and associations with its neighbors allows for input into deep learning networks
- BUT! Don't throw out the power of heuristics quite yet!

The reality of NLP in medical reports

Moving from case-selection to discrete data extraction required new approaches – The problem is much more difficult

Confounding data

- Extra numbers and values - sometimes conflicting
- Needs knowledge on how to interpret

Ad hoc tables and data aggregates

- Formatted with tabs, spaces, characters
- Requires pattern recognition

Lots of variability in data representation

When grammar does not help to bound the search for concepts and relations, the matching of keywords to concepts becomes somewhat combinatorial.

What do we do with NLP?

- Are we trying to "understand" medical reports? – Not really.
  - ✓ We are looking for the correct values for specific attributes.
  - ✓ Locate the correct concept in a string of words – which may or may not be in sentences.
- So the problem is translated as a "Search" problem
- We reduce it to the "physics" level – how do we perform this search?
  - Reduce the search space
  - Implement fast search algorithms
  - Add heuristics to improve accuracy
  - Exploit other NLP techniques to augment heuristics

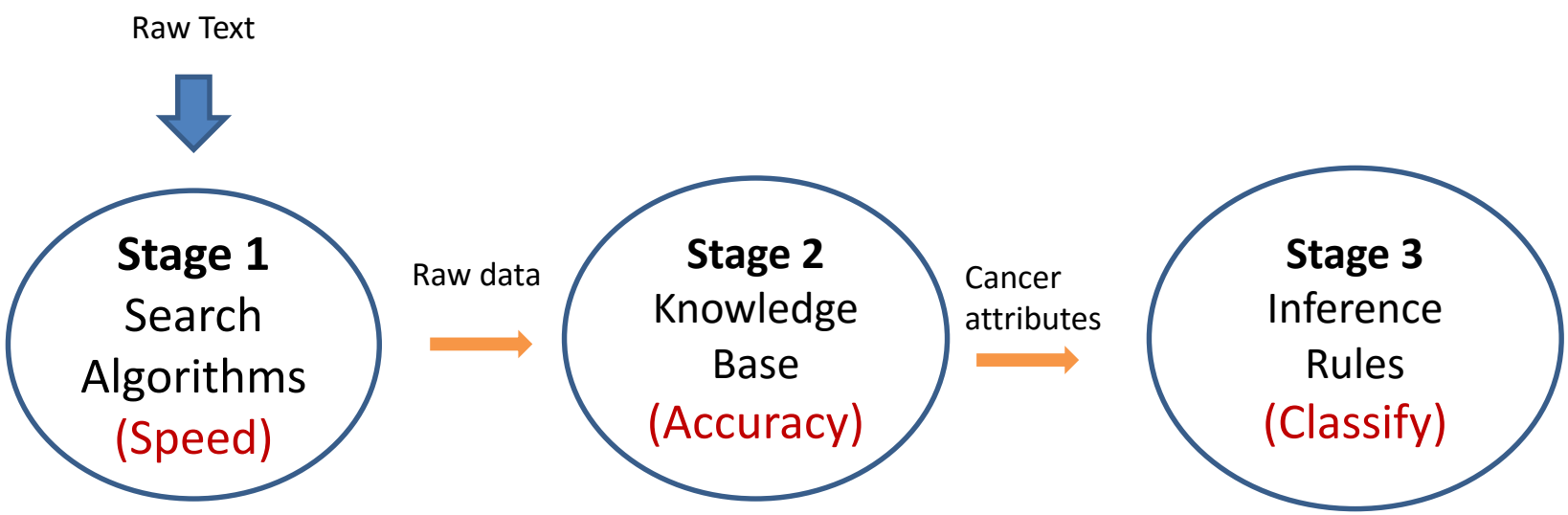
The system goes beyond grammar parsing algorithms to identify patterns and data aggregates thereby increasing accuracy

AIM's AI Engine

- Attribute value identification
- Heuristic segmentation *
- Table detection
- Numeric processing
- Data formatting (general methods)
- Scoring heuristics

- Current speed tests with extraction of 90 possible data items per report including case-finding classification rules 60 – 70 reports per second.

3 Main Stages of the AI Engine



Generate lots of data from the raw text. Uses heuristics to find concepts and numeric values

Test and refine this data against a knowledgebase by removing unwanted information

Infer properties such as cancer site, reportability or CNS tumor classification

Stage 1 - Efficient Search

- When sentences are not well formed, concept matching becomes a combinatorial problem
- A method that uses with fast keyword search with efficient concept identification is required to achieve high accuracy and speed.

Five main ideas

- Smart segmentation and distance mapping
- Key word search and concept tracking
- Concept resolution by simultaneous growing and pruning
- Negation detection
- Heuristics scoring

Heuristic segmentation

this technique helps to bound the concepts as well as assign negation context.

A Fast (one pass) search algorithm that use heuristics to reduce the search space

- The AI engine implements what we call "heuristic segmentation" where the boundaries of where to search for concepts are defined by heuristics rather than grammar trees.

- Concept trackers
  - Grow concepts when potential keywords are found
  - Prune concepts when keywords go out of scope as defined by the heuristic segmentation

This stage "generates" what can be considered the "RAW" data - all the candidate concepts, numbers and things that can be found within the text, but by design - most of this stuff is not the desired data.

Stage 2 - Test and Refine (Accuracy)

Heuristics that constrain data specific properties

- Numeric units and ranges, allowed values
- Conflict resolution, context determination.
- Question/answer mapping

This stage examines the volume of data produced by the stage 1 algorithms and refines or removes errant data.

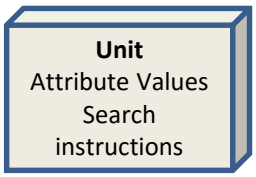
- Each data item has it's own properties and rules – independent of the general NLP methods. We call these pieces of knowledge "UNITS"
- Units are the direct coding of knowledge about how to identify and interpret the correct information for that data item

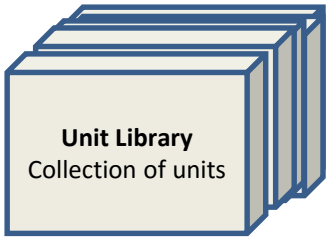
Stage 2 - Test and Refine (Accuracy)

The Units Knowledge base

A unit is the representation of an Attribute and its possible values.

- A unit is self contained (contains all the information required to identify and extract the information)
- The processing of a unit is done by the AI search engine
- Units do not care about any particular domain or application
- Units are developed, tested and stored in a library
- Performance measurement is done on the individual unit





Example: Interpretation and standardization

Synoptic Element	Data Value
Specimen Type	
Lymph Node Sampling	
Lymph Nodes Examined	
Lymph Nodes Positive	
Lymph Nodes Negative	
Extranodal Extension	
Specimen Size	
Tumor Laterality	
Tumor Site	
Tumor Size **	5.0x3.0x3.0 cm
Histologic Type	
Grading System	
Histologic Grade	
Tubule Formation	
Nuclear Pleomorphism	
Mitotic Count	
Necrosis	
Pathologic Staging (pTNM)	
HER2 Status	
ER Status	
PR Status	
Margins **	
Distance of Tumor from Margin Uninvolved ...	
Distance of Tumor from Deep Margin **	3.5 cm

PID: 03:SG1798
Breast.
This **tumour measures** approximately **5.0 cm superior to inferior, 3.0 cm anterior to posterior and 3.0 cm** medial to lateral. This tumour is approximately 2.0 cm from the closest resection margin which is the anterior margin. It is 4.5 cm from the superior, 3.5 cm from the deep, 9.9 cm from the inferior, 8.0 cm from the lateral and 6.5 cm from the medial margins.

Example: Pattern recognition

Synoptic Element	Data Value
Specimen Type	Radical Prostatectomy
Prostatectomy Weight **	
Specimen Size **	
Histologic Type	Adenocarcinoma
Histologic Grade **	Gleason Pattern
Primary Grade **	4
Secondary Grade **	3
Tertiary Pattern	
Total Gleason Grade **	7/10
Tumor Quantitation,% of pro...	25 %
PSA	
Tumor Size	
Tumor Greatest Dimension	
Pathological Staging (pTNM)	
Regional Lymph Nodes	Lymph Node - Negative for malignancy
	Lymph Nodes mentioned in Report
Regional Lymph Nodes Exa...	
Regional Lymph Nodes Inv...	
Regional Lymph Nodes Ne...	2
Distant Metastasis	
Margin Involvement	Margin(s) uninvolved by invasive carcinoma
Apical Margin **	Apical Margin uninvolved by invasive carci...
Bladder Neck Margin	
Anterior Margin	
Lateral Margin	
Posterior Margin	
Radial/Circumferential margin	
Distal margin	

```
PID: 505522250
PROSTATEctomy.

(a) Moderately differentiated adenocarcinoma of prostate involving
        approximately 25% of surface area of prostate gland, Gleason score
        4 + 3 = 7 out of 10, perineural and focal lymphovascular invasion
        noted - prostate gland.
    (b) The tumour focally involves the inked resection margin of the left
        lobe.
    (c) The base margin is clear.
    (d) The apical margin is clear (focally the tumour reaches very close
        to the inked margin ).
    (e) Negative for malignancy - bilateral seminal vesicles.
    (d) Negative for malignancy - (#1 \T\ 2) right and left pelvic obturator
        lymph nodes
```

Example: Table detection

Synoptic Element	Data Value
Specimen Type	Segmental Resection
Specimen Length **	
Tumor Site	Sigmoid colon
Tumor Configuration	
Tumor Size **	0.6x0.5x0.2 cm
Tumor Greatest Dimension	
Intactness of Mesorectum	
CEA	
Histologic Type	Adenocarcinoma
Histologic Grade	Low-grade
Pathologic Staging	pT2: Tumor invades muscularis prop
Regional Lymph Nodes **	pN0: No regional lymph node metast
Distant Metastasis	
Lymph Nodes Examined **	8
Lymph Nodes Involved **	
Lymph Nodes Negative **	
Margins of Resection	There are Uninvolved Margins
Distance of Invasion from M...	
Proximal Margin	
Distance from closest Proxim...	
Distal Margin	
Distance from closest Distal ...	
Circumferential/Radial Margin	
Distance from closest Circum...	
Mesenteric Margin	
Distance from closest Mesen...	
Lymphatic Invasion	Lymphatic Invasion Absent
Venous Invasion	Venous Invasion Absent
Perineural Invasion **	Perineural Invasion Absent
Tumor Border Configuration	
Lymphocytic Response To T	

```
TUMOUR SUMMARY:
--------------
SPECIMEN:   -  Segmental resection.
TYPE:       -  Invasive adenocarcinoma, NOS.
LOCATION:   -  Sigmoid colon.
GRADE:      -  Well differentiated (G1).
INVASION:   -  Intraepithelial tumor with invasion of the lamina propria
                  with no extension through muscularis mucosa into
     submucosa (pT2).
TOTAL NUMBER OF LYMPH NODES: -  8.
LNs:          No regional lymph node metastasis (pN0).
PROXIMITY TO NEAREST MARGIN: -  7.5 cm.
SIZE:    -  0.6 x 0.5 x 0.2 cm.
LYMPHATIC/VASCULAR INVASION: -  Negative.
PERINEURAL INVASION:  -  Negative.
MARGINS:     -  Negative.
```

Stage 3 - Inference engine

- This is where the higher level operations take place
- Production rules based system using inference engine to examine the resulting information and make decisions.
- Used to develop expert systems - rules can address any level of logic complexity.
- Examples
  - Case-selection
  - CNS classification,
  - Site identification,
  - Recurrence **
  - Automated coding **

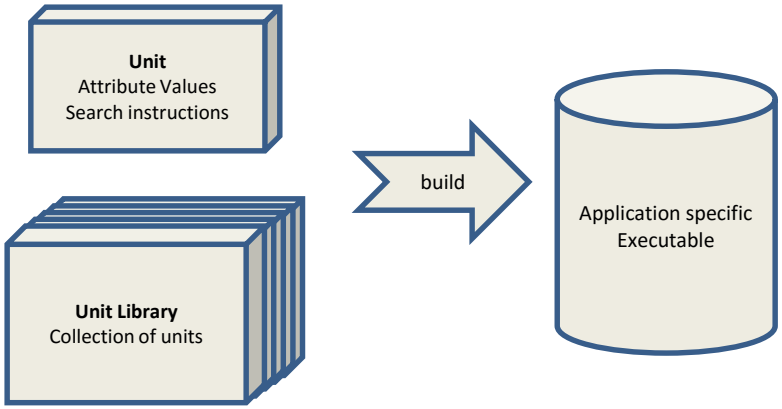
Example rule – found metastatic condition

```
(defrule CNS!Conditions!Metastatic
        (agenda conditions $?)
        (codes (code-type C)(wordSense positive)(code ?c &:(match ?c 70.0 70.1 70.9 72.8 72.9 75.1 75.2 75.3)
                                                          |:(in ?c 71.0 71.9)
                                                          |:(in ?c 72.0 72.5))
                (sectionName ?section)(sentenceId ?sentenceId))

        (codes (code-type M) (code ?mCode)(b 6)(n ?num)(wordSense positive)
                (sectionName ?section))
                 =>
       (say "Condition Metastatic "  ?mCode)
       (assert (Condition (name metastatic) (code ?mCode)(sentenceId ?sentenceId)))
)
```

Building a working system

Select the units needed for a specific application



- Architecture separates NLP processing algorithms from knowledge base definitions
- Units allow many improvements in how we define, manage, test and improve the AI knowledge bases.
- Units can be built for any type of information in any document type.
- Inference engine capabilities allows complex inference rules.

Units so far..

Unit Name
Additional Pathologic Findings
AFP
ALK
Arterial Invasion
b-hCG
Bloom Richardson Grade
Bloom Richardson Score
BRAF
Breslow's depth
CA-125
Calcification
Calcitonin
CEA
Clark's Staging
Depth of Invasion
Distance of Tumor from Anterior Margin
Distance of Tumor from Closest margin
Distance of Tumor from Deep Margin
Distance of Tumor from Inferior Margin
Distance of Tumor from Lateral Margin
Distance of Tumor from Medial Margin
Distance of Tumor from Posterior Margin
Distance of Tumor from Superior Margin
Distant Metastasis (pM)
EGFR
ER - Allred Score
ER Status
Examination type
Extranodal Extension

FISH Resu+B31:B60lts
Focality
Gleason Grade - Primary Pattern
Gleason Grade - Secondary Pattern
Gleason Grade - Tertiary Pattern
Gleason Score
Grade of dysplasia
Grading System
HER2 % cells stained
HER2 gene copy number
HER2 Result
HER2:CEP 17 ratio
Histologic Grade
Histologic Type
HX
ICD-O-3 Morphology AIM Inc.
ICD-O-3 Topography AIM Inc.
Implants
Ki-67
KRAS
Laterality
LDH
Lymph Nodes Examined
Lymph Nodes Negative
Lymph Nodes Positive
Lymphatic Invasion
Lymphovascular (LV) Invasion
Miscellaneous Terms
Mitotic Count
Nottingham Grade

Nottingham Score
Nuclear Pleomorphism
Organ(s) Included
Pathologic Staging (FIGO)
Perineural Invasion
Periprostatic Fat Invasion
PK
Pleural Invasion
Positive Cancer Terms
PR - Allred Score
PR Status
Primary Tumor (pT)
Procedure
PSA
Regional Lymph Nodes (pN)
S-100
Seminal Vesicle Invasion
Site ID
Specimen Greatest dimension
Specimen Size
Specimen Type
Specimen Weight
Stage
Treatment
Tubule Formation
Tumor Configuration
Tumor Site and Extent
Tumor Size - Greatest dimension (cm)
Tumor Weight
Venous (Large Vessel) Invasion

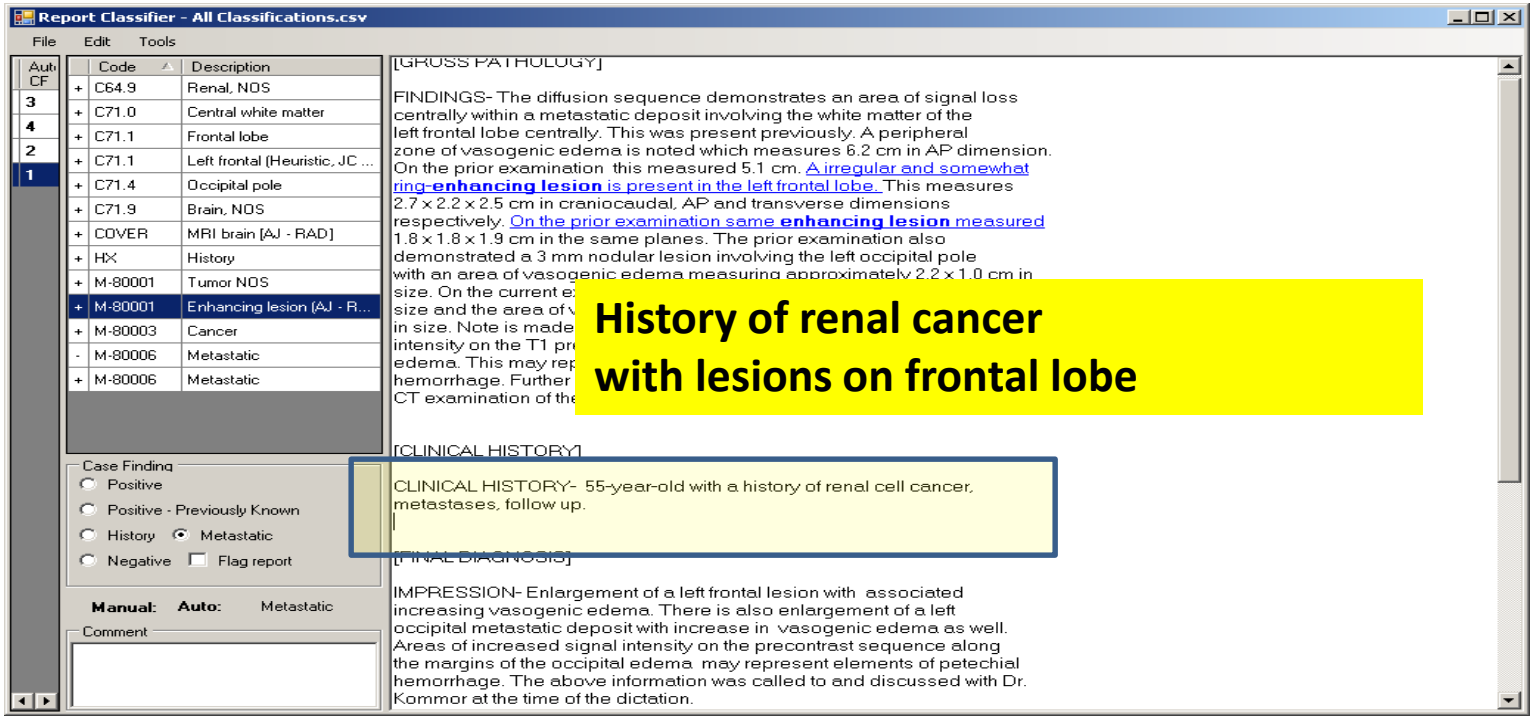
Units so far..

Unit Name	FISH Resu+B31:B60lts	Nottingham Score
Additional Pathologic Findings	Focality	Nuclear Pleomorphism
AFP	Gleason Grade - Primary Pattern	Organ(s) Included
ALK	Gleason Grade - Secondary Pattern	Pathologic Staging (FIGO)
Arterial Invasion	Gleason Grade - Tertiary Pattern	Perineural Invasion
b-hCG	Gleason Score	Periprostatic Fat Invasion
Bloom Richardson Grade	Grade of dysplasia	PK
Bloom Richardson Score	Grading System	Pleural Invasion
BRAF	HER2 % cells stained	Positive Cancer Terms
Breslow's depth	HER2 gene copy number	PR - Allred Score
CA-125	HER2 Result	PR Status
Calcification	HER2:CEP 17 ratio	Primary Tumor (pT)
Calcitonin	Histologic Grade	Procedure
CEA	Histologic Type	PSA
Clark's Staging	HX	Regional Lymph Nodes (pN)
Depth of Invasion	ICD-O-3 Morphology AIM Inc.	S-100
Distance of Tumor from Anterior Margin	ICD-O-3 Topography AIM Inc.	Seminal Vesicle Invasion
Distance of Tumor from Closest margin	Implants	Site ID
Distance of Tumor from Deep Margin	Ki-67	Specimen Greatest dimension
Distance of Tumor from Inferior Margin	KRAS	Specimen Size
Distance of Tumor from Lateral Margin	Laterality	Specimen Type
Distance of Tumor from Medial Margin	LDH	Specimen Weight
Distance of Tumor from Posterior Margin	Lymph Nodes Examined	Stage
Distance of Tumor from Superior Margin	Lymph Nodes Negative	Treatment
Distant Metastasis (pM)	Lymph Nodes Positive	Tubule Formation
EGFR	Lymphatic Invasion	Tumor Configuration
ER - Allred Score	Lymphovascular (LV) Invasion	Tumor Site and Extent
ER Status	Miscellaneous Terms	Tumor Size - Greatest dimension (cm)
Examination type	Mitotic Count	Tumor Weight
Extranodal Extension	Nottingham Grade	Venous (Large Vessel) Invasion

Successful applications

1. E-Path-Pathology and diagnostic imaging
2. Abrevio document abstracting and RCA
3. Information from Radiotherapy reports
4. The coding problem for primary tumors *

Diagnostic imaging -Metastatic



Positive – Previously Known

Report Classifier - All Classifications.csv

File Edit Tools

Auto CF		Code	Description
3	-	AIM-930-...	recurrence
4	+	C09.9	Tonsillar
2	+	C41.2	Spine
1	+	C53.9	Cervical
	+	C71.0	Cerebral [JC - RAD]
	+	C71.3	Parietal lobe
	+	C71.5	Lateral ventricle, NOS
	+	C71.5	Ventricle, NOS
	+	C71.7	Fourth ventricle, NOS
	+	C71.8	Corpus callosum
	+	C71.9	Brain, NOS
	+	C76.0	Head
	+	COVER	MRI brain [AJ - RAD]
	+	HX	History
	-	M-80001	Mass, NOS [JC - RAD]
	+	M-80001	Tumor NOS
	+	M-94703	Medulloblastoma, NOS (...

Case Finding
- Positive
- Positive - Previously Known
- History
- Metastatic
- Negative
- Flag report

Comment
Manual: Positive (Previous

[GROSS PATHOLOGY]

Note is made of a right parietal approach ventricular shunt catheter with the catheter passing through the right lateral ventricle from a posterior to anterior direction, the tip terminating in the region of the anterior aspect of the corpus callosum. There has been interval ventricular dilatation. The ventricles were mildly dilated on the prior study and are currently of normal caliber with the exception of postsurgical widening of the fourth ventricle.
Within the supratentorial space, there has been the interval decrease in degree of edema within the subcortical white matter in the right parietal lobe in the inferior parietal lobule at the site of ventricular shunt placement. There is no evidence of acute cerebral infarction. In this patient with right-sided weakness, there is no evidence of left-sided white matter edema or cortical infarction in the supratentorial space. There is no midline shift.
Within the infratentorial space, note is made of postsurgical change following suboccipital craniectomy for resection of a **medulloblastoma**. They are secondary postoperative widening of the fourth ventricle. There is no evidence of a brainstem infarction. There is cerebellar tonsillar ectopia with secondary narrowing of the CSF spaces of the foramen magnum.

[CLINICAL HISTOR

UNSPEC SURGEF
ADMIT DATE/TIME
Brain tumor remove
MRI of brain perfor

Resection medulloblastoma
Post surgical change

[FINAL DIAGNOSIS]

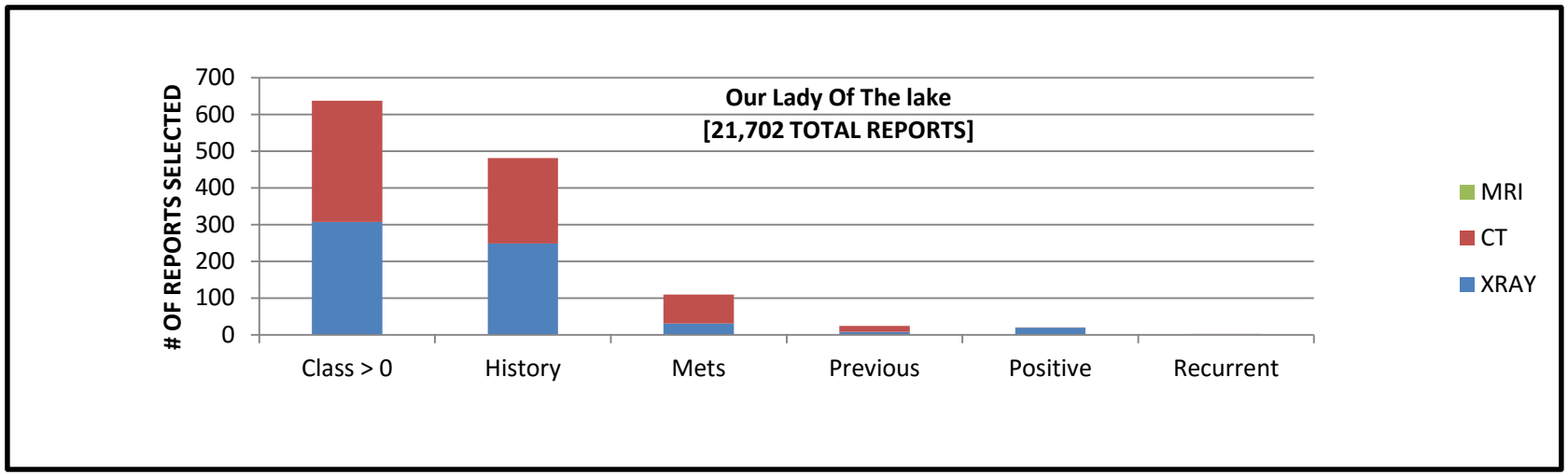
1. Interval resolution of pneumocephalus following surgery with interval ventricular decompression following placement of a ventricular shunt catheter. The ventricles are now of normal caliber.
2. Postoperative change following resection of a **medulloblastoma** at the level the fourth ventricle.
3. Soft tissue at the foramen magnum compatible with moderate cerebellar tonsillar ectopia. Given the patient's new onset right-sided weakness and absence of CT evidence of infarction, recurrent or residual mass or progressive hemorrhage in the brain, further evaluation with an MRI of the cervical spine to evaluate for the degree of cerebellar tonsillar ectopia and for the possibility of cervical syrinx or compression of the cervicomedullary junction may be of benefit.
Dictated by: Robert J Kadner, M.D.
Images and Report reviewed and interpreted by: Robert J Kadner, M.D.
<PS><Electronically signed by: Robert J Kadner, M.D.>
05/25/2010 1252

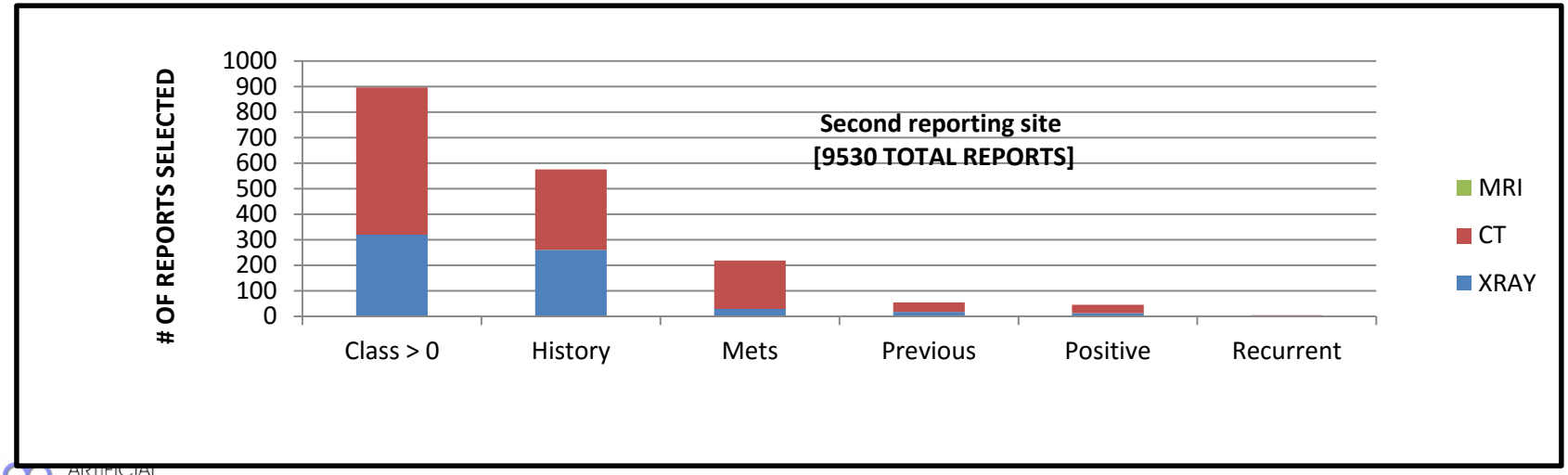
Diagnostic imaging of the Chest

Hospital A	XRAY	CT	MRI	TOTAL
Reports sample	21,284	416	2	21,702
Classification > 0	307	330	0	637
1 History of cancer	249	232	-	481
2 Metastases	31	79	-	110
3 Previously known	8	16	-	24
4 Positive (new)	19	1	-	20
5 Recurrent	-	2	-	2

Hospital B (Mar-Apr/16)	XRAY	CT	MRI	TOTAL
Reports sample	1877	7651	2	9530
Classification > 0	319	577	0	896
1 History of cancer	260	315	-	575
2 Metastases	29	189	-	218
3 Previously known	17	37	-	54
4 Positive (new)	13	32	-	45
5 Recurrent	-	4	-	4

Diagnostic imaging of the Chest





The Coding problem

Can an automated system provide the SEER Site coding designations?

- Use defined SEER Rules
- Compare against large reference set and identify discrepancies.
- Create new rules and re-test
- Measure confidence - Identify high confidence codes
- Provide explanation for codes

Automated coding wizard

GROSS DESCRIPTION:
Specimen is received in two parts

Parts 1: Specimen is labeled "left lobe and left isthmus". Specimen is received fresh for intraoperative consultation and consists of a 6 gm, 3.9 x 2.5 x 2 cm lobe of thyroid gland. Specimen is inked and serially sectioned to reveal a 1.7 cm tan-white mass abutting the surrounding surface focally. A fine needle aspiration is performed. Representative section of the mass is submitted for frozen. The remnant of the frozen is submitted in cassette 1 FS. The remainder of the thyroid parenchyma is dark red and homogeneous. There is no additional nodule grossly identified. The remainder of the specimen is submitted: Cassette A-D - mass, Cassette E-G - the remainder of the thyroid gland, in a total of eight cassettes.

Parts 2: Specimen is labeled "paratracheal lymph node". Specimen is received fresh and consists of a 0.4 x 0.2 x 0.2 cm ovoid tan nodule. Entire specimen is submitted in one cassette. ZZ/FA/pds
#

DIAGNOSIS:
1. THYROID, LEFT LOBE AND ISTHMUS (RESECTION):
CARCINOMA OF THE THYROID, 1.7 CM, WITH INVASION OF THE THYROID CAPSULE.
TUMOR APPROACHES TO WITHIN FRACTIONS OF A MILLIMETER OF THE EXTERNAL
SURFACE.
NO INVASION OF THE SURROUNDING SKELETAL MUSCLE SEEN.

2. PARATRACHEAL LYMPH NODE (BIOPSY):
METASTATIC CARCINOMA, MICROSCOPIC FOCI, SUBCAPSULAR SINUS OF LYMPH
NODE, SEE COMMENT.

M-80103

M-82603 95%

M-80503

C73.9 - 95%

C77.9

Explanation:
If have
C73.9 (Thyroid NOS) + M80503 (Papillary carcinoma)
Then add
M82603 (Papillary Carcinoma of thyroid)

Coding: What if we can use this data?

Unit Name
Additional Pathologic Findings
AFP
ALK
Arterial Invasion
b-hCG
Bloom Richardson Grade
Bloom Richardson Score
BRAF
Breslow's depth
CA-125
Calcification
Calcitonin
CEA
Clark's Staging
Depth of Invasion
Distance of Tumor from Anterior Margin
Distance of Tumor from Closest margin
Distance of Tumor from Deep Margin
Distance of Tumor from Inferior Margin
Distance of Tumor from Lateral Margin
Distance of Tumor from Medial Margin
Distance of Tumor from Posterior Margin
Distance of Tumor from Superior Margin
Distant Metastasis (pM)
EGFR
ER - Allred Score
ER Status
Examination type
Extranodal Extension

FISH Resu+B31:B60lts
Focality
Gleason Grade - Primary Pattern
Gleason Grade - Secondary Pattern
Gleason Grade - Tertiary Pattern
Gleason Score
Grade of dysplasia
Grading System
HER2 % cells stained
HER2 gene copy number
HER2 Result
HER2:CEP 17 ratio
Histologic Grade
Histologic Type
HX
ICD-O-3 Morphology AIM Inc.
ICD-O-3 Topography AIM Inc.
Implants
Ki-67
KRAS
Laterality
LDH
Lymph Nodes Examined
Lymph Nodes Negative
Lymph Nodes Positive
Lymphatic Invasion
Lymphovascular (LV) Invasion
Miscellaneous Terms
Mitotic Count
Nottingham Grade

Nottingham Score
Nuclear Pleomorphism
Organ(s) Included
Pathologic Staging (FIGO)
Perineural Invasion
Periprostatic Fat Invasion
PK
Pleural Invasion
Positive Cancer Terms
PR - Allred Score
PR Status
Primary Tumor (pT)
Procedure
PSA
Regional Lymph Nodes (pN)
S-100
Seminal Vesicle Invasion
Site ID
Specimen Greatest dimension
Specimen Size
Specimen Type
Specimen Weight
Stage
Treatment
Tubule Formation
Tumor Configuration
Tumor Site and Extent
Tumor Size - Greatest dimension (cm)
Tumor Weight
Venous (Large Vessel) Invasion

Machine learning approach – preliminary results

- Run experiments using previously coded reports as the reference set.
- Extract UNITS data from each report and combine with reference data to create a training set.
- Run experiments to auto generate coding rules and test for accuracy.
- Use those rules in a production system as a Coding Wizard.

Experiments – machine generated rules

Input vector generated by UNITS Engine for each report ~ 3000 reports

	Histologic type1	Histologic type2	C-code1	C-code2	M-code1	M-code2	M-code3	Reference M-code

Rules produced by machine learning with estimated confidence values

1. HistologicType_1=Adenocarcinoma NOS, TopographyCode_1=C61.9 49 ==> M_CODE=M-81403 49 acc:(0.98136)
2. TopographyCode_1=C61.9 Morphology, Code_1=M-81403 44 ==> M_CODE=M-81403 44 acc:(0.97935)
3. HistologicType_1=Invasive Adenocarcinoma Morphology, Code_1=M-81403 26 ==> M_CODE=M-81403 26 acc:(0.96584)
4. MorphologyCode_1=M-80001 Morphology, Code_2=M-81403 21 ==> M_CODE=M-81403 21 acc:(0.95821)
5. HistologicType_1=Adenocarcinoma NOS, HistologicType_2=Adenocarcinoma NOS 19 ==> M_CODE=M-81403 19 acc:(0.95412)
6. HistologicType_1=Adenocarcinoma NOS, TopographyCode_2=C77.5 17 ==> M_CODE=M-81403 17 acc:(0.94916)
7. HistologicType_1=Adenocarcinoma NOS, MorphologyCode_2=M-80003 17 ==> M_CODE=M-81403 17 acc:(0.94916)
8. HistologicType_1=Invasive Adenocarcinoma, TopographyCode_1=C61.9 16 ==> M_CODE=M-81403 16 acc:(0.94626)
9. TopographyCode_1=C61.9, MorphologyCode_1=M-80003 16 ==> M_CODE=M-81403 16 acc:(0.94626)

Testing the rules on new data

Site	Training set	classifier	Accuracy
Georgia	90%	NaiveBase	68%
		BayesNet	77%
		J48	69%
	100%	NaiveBayes	74%
		BayesNet	81%
Kentucky	90%	NaiveBayes	65%
		BayesNet	65%
		J48	67%

Some issues

Preliminary experiments with small data set – need more reference data

We are not 100% confident of the accuracy of the coding because of age of data set.

What's Next?

- Continue development of prototype using live reference data.
- The approach can be imbedded in a production system to build training data from user input.
- It can monitor its existing rules based on ongoing user input and compare.
- Flag when a rule change is found and may automatically update the rule set if needed.
- The availability of UNITS data for machine learning can open up many possibilities.

Thank you

Heuristic segmentation – simple example

In **word co-ordinates only**

The brown fox jumped over the small pond and landed on the green grass, however he missed the quick rabbit.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

*Introduce some partial segments with example scores (**and** = 20 **,** = 8 **However** = 20 **Period** = 100)*

Super imposed co-ordinates:

The brown fox jumped over the small pond **and** landed on the green grass**, however** he missed the quick rabbit**.** Continue..

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

20 **8 20** **100**

Cumulative Co-ordinates

The brown fox jumped over the small pond and landed on the green grass, however he missed the quick rabbit. Continue…

1 2 3 4 5 6 7 8 **29** 30 31 32 33 34 **42** 62 63 64 65 66 **167** ….

The brown fox jumped over the small puddle and landed on the green grass, however he missed the quick rabbit. Continue…

Hills increase the distance and hence reduce the probability of concept matching between keywords. Segmentation (hill sizes) can be concept dependent - the landscape changes as the search progresses and with the concept being tracked.

Simple example

The brown fox jumped over the small puddle and landed on the green grass, however he missed the quick rabbit. Continue…

1 2 3 4 5 6 7 8 **29** 30 31 32 33 34 **42** 62 63 64 65 66 **167**

Say we have these concepts to track

- The quick brown fox
- The quick rabbit
- The small rabbit

Follow the trace

N:found

Place-holder

A-[removed]

A-[Complete]

Search engine found the concept "the quick rabbit"

Found keyword **1:the**

A-[1: **the quick brown fox**]
B-[1: the quick rabbit]
C-[1: the small rabbit]

Found keyword **2:brown**

A-[1: **the quick 2:brown fox**]

Found keyword **6:the**

D-[6: **the quick brown fox**]
E-[6: the quick rabbit]
F-[6: the small rabbit]

Found keyword **7:small**

C-[1: the 7:small rabbit]
F-[6: the 7:small rabbit]

…….

……..

Found keyword **65:quick**

A-[1: **the quick brown fox]**
B-[1: the quick rabbit]
C-[1: the small rabbit]
D-[6: **the quick brown fox]**
E-[6: the quick rabbit]
F-[6: the small rabbit]
G-[64:the 65:quick **brown fox**]
H-[64:the 65:quick **rabbit**]

Found keyword **66:rabbit**

H-[64:the 65:quick rabbit]

RCA – Rapid Case Ascertainment

- Convert free text path reports to standardized data
- Searchable for studies, trials..
- Automated candidate identification and notification
- Works with historical as well as current reports
- Produces statistics on pathology data contained in reports

Pathology Report

CLINICAL HISTORY/MACROSCOPY
Right mastectomy and axillary tissue. A right mastectomy specimen with overlying skin measuring 220mm x 85mm and underlying breast tissue measuring 220mm x 100mm x 70mm. The axillary tail measures 125 x 60mm. The nipple is slightly retracted and located centrally. The superior margin is painted red, the inferior margin painted green and the deep cut margin is painted blue. Cut sections of the underlying breast tissue shows an ill-defined grey white yellow lesion with patchy areas of haemorrhage measuring 35 x 35 x 35mm located immediately below the nipple, 20mm from the inferior margin, 45mm from the deep cut margin, 50mm from the superior margin, 85mm from the medial margin and 100mm from the lateral cut margin. A1 - nipple, B1 - upper outer quadrant, C1 - upper inner quadrant, D1 - lower outer quadrant, E1 - lower inner quadrant, F1, G1 - tumour composite blocks, H1, I1 - tumour composite blocks, J1 - deep cut margin, K1 - superior margin, L1 – inferior margin, M4 - lymph nodes, N4 - lymph nodes, O - 3 serial slices, lymph node, P - 3 lymph nodes.

MICROSCOPY
This right mastectomy specimen demonstrates an invasive ductal carcinoma with the following pathological features:

TUMOUR HISTOLOGY & GRADE
The tumour is of an infiltrating poorly differentiated ductal carcinoma of non-otherwise specified type. The tumour is poorly defined and extremely infiltrative, comprising poorly-formed tubules, nests or strands of cuboidal tumour cells displaying high grade nuclei. The tumour cells are set within fibrotic desmoplastic stroma. Many lactiferous ducts are entrapped within the tumour. Frequent tumour mitoses are seen. Microcalcification is seen in some neoplastic tubules.

Tumour grade (Modified Bloom-Richardson Scoring System):
Tubular formation: 3
Nuclear atypia: 3
Tumour mitoses: 2

Synoptic Report

Synoptic Element	Data Value
Specimen Type	Mastectomy
Lymph Node Sampling **	Axillary dissection
Lymph Nodes Examined **	13
Lymph Nodes Positive **	3
Lymph Nodes Negative	
Extranodal extension	
Specimen Size **	220x85 mm
Laterality	Right
Tumor Site **	Lower inner quadrant
Size of Invasive Component **	35x35x35 mm
Invasive Component Greatest dimension	
Histologic Type	Invasive ductal carcinoma
	Ductal carcinoma in situ
	Paget disease without invasive carcinoma
	Cribriform
	Tubular
Grading System	Bloom Richardson
Histological Grade	Grade III
Tubule Formation **	3
Nuclear Pleomorphism **	3
Mitotic Count **	10
Necrosis	
Pathologic Staging (pTNM)	
HER2 Status **	Negative
ER Status **	Positive
PR Status **	Positive
Margin **	
Distance from Deep Margin **	45 mm
Distance from Lateral Margin **	50 mm
Distance from Superior Margin **	50 mm
Distance from Inferior Margin **	20 mm
Distance from Medial Margin **	50 mm
Distance from Anterior Margin	
Distance from Posterior Margin	
Location of involved margin by invasive ...	
DCIS Margin Involvement	
Distance from uninvolved DCIS closest ...	
Location of uninvolved DCIS margin	
Location of involved DCIS margin	
Extent of Margin Involvement for Invasi...	
Extent of Intraductal Component	
DCIS Extent of Intraductal Component	
Extent of Margin Involvement for DCIS	

Pathology Report

CLINICAL HISTORY/MACROSCOPY

Right mastectomy and axillary tissue. A right mastectomy specimen with overlying skin measuring 220mm x 85mm and underlying breast tissue measuring 220mm x 100mm x 70mm. The axillary tail measures 125 x 60mm. The nipple is slightly retracted and located centrally. The superior margin is painted red, the inferior margin painted green and the deep cut margin is painted blue. Cut sections of the underlying breast tissue shows an ill-defined grey white yellow lesion with patchy areas of haemorrhage measuring 35 x 35 x 35mm located immediately below the nipple, 20mm from the inferior margin, 45mm from the deep cut margin, 50mm from the superior margin, 85mm from the medial margin and 100mm from the lateral cut margin. A1 - nipple, B1 - upper outer quadrant, C1 - upper inner quadrant, D1 - lower outer quadrant, E1 - lower inner quadrant, F1, G1 - tumour composite blocks, H1, I1 - tumour composite blocks, J1 - deep cut margin, K1 - superior margin, L1 – inferior margin, M4 - lymph nodes, N4 - lymph nodes, O - 3 serial slices, lymph node, P - 3 lymph nodes.

MICROSCOPY

This right mastectomy specimen demonstrates an invasive ductal carcinoma with the following pathological features:

TUMOUR HISTOLOGY & GRADE

The tumour is of an infiltrating poorly differentiated ductal carcinoma of non-otherwise specified type. The tumour is poorly defined and extremely infiltrative, comprising poorly-formed tubules, nests or strands of cuboidal tumour cells displaying high grade nuclei. The tumour cells are set within fibrotic desmoplastic stroma. Many lactiferous ducts are entrapped within the tumour. Frequent tumour mitoses are seen. Microcalcification is seen in some neoplastic tubules.

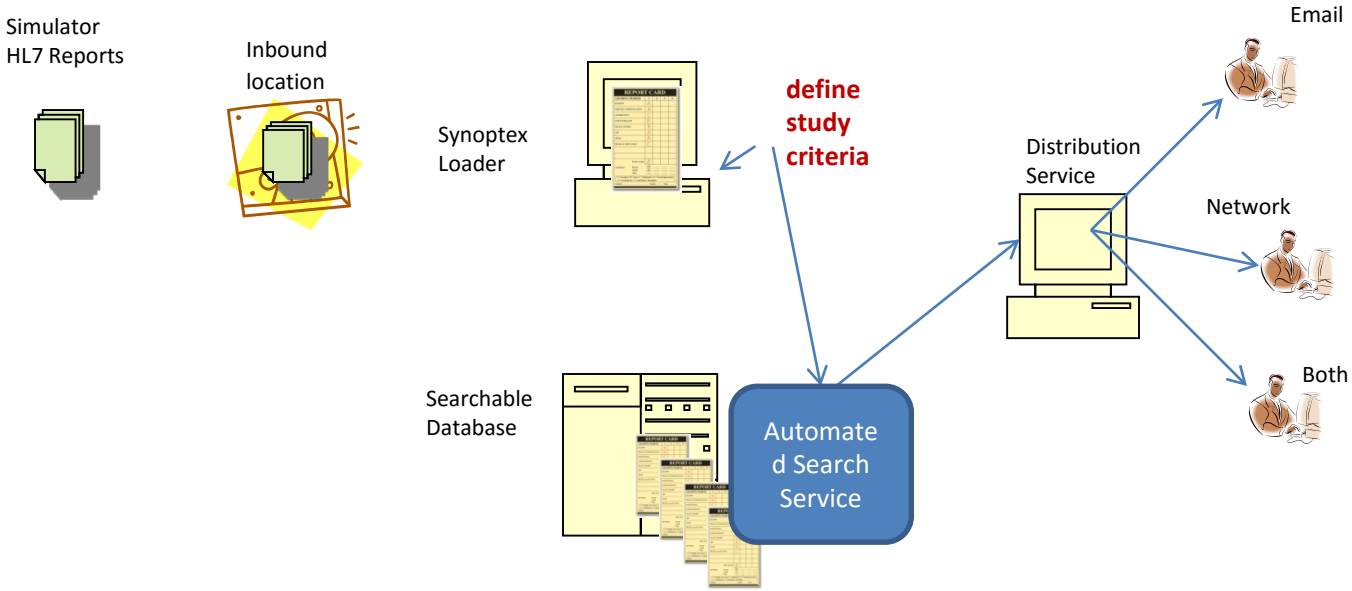
Tumour grade (Modified Bloom-Richardson Scoring System):

Tubular formation:	3
Nuclear atypia:	3
Tumour mitoses:	2

Synoptic Report

Synoptic Element	Data Value
Specimen Type	Mastectomy
Lymph Node Sampling **	Axillary dissection
Lymph Nodes Examined **	13
Lymph Nodes Positive **	3
Lymph Nodes Negative	
Extranodal extension	
Specimen Size **	220x85 mm
Laterality	Right
Tumor Site **	Lower inner quadrant
Size of Invasive Component **	35x35x35 mm
Invasive Component Greatest dimension	
Histologic Type	Invasive ductal carcinoma
	Ductal carcinoma in situ
	Paget disease without invasive carcinoma
	Cribriform
	Tubular
Grading System	Bloom Richardson
Histological Grade	Grade III
Tubule Formation **	3
Nuclear Pleomorphism **	3
Mitotic Count **	10
Necrosis	
Pathologic Staging (pTNM)	
HER2 Status **	Negative
ER Status **	Positive
PR Status **	Positive
Margin **	
Distance from Deep Margin **	45 mm
Distance from Lateral Margin **	50 mm
Distance from Superior Margin **	50 mm
Distance from Inferior Margin **	20 mm
Distance from Medial Margin **	50 mm
Distance from Anterior Margin	
Distance from Posterior Margin	
Location of involved margin by invasive ...	
DCIS Margin Involvement	
Distance from uninvolved DCIS closest ...	
Location of uninvolved DCIS margin	
Location of involved DCIS margin	
Extent of Margin Involvement for Invasi...	
Extent of Intraductal Component	
DCIS Extent of Intraductal Component	
Extent of Margin Involvement for DCIS	

Rapid Case Ascertainment - Data Flow



RCA changes the way you search
Finds the candidates for you in real time

E-Path RCA - [Study Manager]

File View Tools Windows Help

Manage Studies

Studies

New Edit Delete Start View New Reports View Distributed Reports Refresh View Archived Studies

ID	Study Name	Investigator	Status	Distributed Cases	Start Date	Days from Start	Stopped Date	Last Distribution
231	Colorectal	Smith	New	0				
227	George study 1	G Cernile	Running	0	2/6/2012	52		
216	Breast	A. Jashki	Stopped	3	2/1/2012	57	2/1/2012	2/6/2012
204	Breast - Non-Invasive Carcinoma	J. Chen	Stopped	3	12/19/2011	100	12/21/2011	1/4/2012
201	Test 3	Cernile	Running	8	12/15/2011	105		12/15/2011
199	Test1	Cernile	Stopped	182	12/14/2011	106	12/21/2011	12/14/2011
188	Lung - Squamous Cell Carcinoma	Dr. N. Jones	Stopped	0	2/27/2012	30	2/27/2012	
171	Breast - Ductal Carcinoma	Dr. K. Williams	Running	2090	6/30/2011	273		3/26/2012

Search for reports from 3/29/2012 to 3/29/2012

Distribution Cycle Sun Mon Tue Wed Thu Fri Sat

Site Colon and Rectum

Study Constraints

New Edit Delete Query Analyzer ON

Question	Operand	Answer	Enabled
Histologic Type	Equals	Adenocarcinoma	✓
		Mucinous adenocarcinoma (greater than...	✓
		Medullary carcinoma	✓
Primary Tumor (pT)	Equals	pT1: Tumor invades submucosa	✓
		pT2: Tumor invades muscularis propria	✓
		pT3: Tumor invades through the muscul...	✓

Found 9 reports

Report Distribution

New Edit Delete Distribute Now Redistribute

Destination	Format	Security	Enabled
\\YourNetwork\StudyLocatrion	Full Report	Include PHI	✓
gcernile@aim.on.ca	Summary Only	Exclude PHI	✓
DrSmith@hospital.org	Summary Only	Exclude PHI	✓

Destination	DrSmith@hospital.org
Include full report	No
Include patient health information	No
Enabled	Yes

Logged in as 'admin'

E-Path RCA - [Study Manager]

File View Tools Windows Help

Manage Studies | Colorectal - New Reports

Report ID	Status	Temp
04:SU12014	N	Colon
04:SU12014	N	Colon
PHCQC0096	N	Colon
QC-0038	N	Colon
QC-0174	N	Colon
QC-0819	N	Colon
QC-122	N	Colon
QC-224	N	Colon
QC-328	N	Colon

9 Reports

Patient & Report Information

Name: XXXXXXXXXX, XXXXXXXXXX Sex: **Male** Date of Birth: **1/1/2004** Health/Record Number: **A9999999999** Date Received: **2/14/2012** Reporting Facility: **0001**

Report Status: Not Reviewed

Colon and Rectum: Resection

Question	Answer
Report Information	
Report ID	04:SU12014
Reporting Facility	0001
Sex	Male
Date of Birth	2004-01-01
Date of Procedure	2004-09-13
Race	
MACROSCOPIC	
SPECIMEN TYPE	Other (specify): Segmental Resection
Specimen Length (cm) (if applicable)	21
TUMOR SITE	Sigmoid colon
TUMOR CONFIGURATION	Exophytic (polypoid)
TUMOR SIZE	
Greatest dimension (cm)	0.6
Additional dimension (cm)	0.5
Additional dimension (cm)	0.2
Intactness of MESORECTUM	
MICROSCOPIC	
HISTOLOGIC TYPE	Adenocarcinoma
HISTOLOGIC GRADE	Not applicable
Other (specify)	
PATHOLOGIC STAGING (p...	
PRIMARY TUMOR (pT)	pT2: Tumor invades muscularis propria
REGIONAL LYMPH NODES (pN)	pN0: No regional lymph node metastasis
Specify number examined	8
Number involved	
DISTANT METASTASIS (pM)	

Flat dysplastic polyp.

SEGMENTAL RESECTION, COLON:
- WELL DIFFERENTIATED ADENOCARCINOMA, SEE TUMOR SUMMARY FOR DETAILS.

TUMOUR SUMMARY:

SPECIMEN: - Segmental resection.
TYPE: - Invasive adenocarcinoma, NOS.
LOCATION: - Sigmoid colon.
GRADE: - Well differentiated (G1).
INVASION: - Intraepithelial tumor with invasion of the lamina propria with no extension through muscularis mucosa into submucosa (pT2).
TOTAL NUMBER OF LYMPH NODES: - 8.
LNs: No regional lymph node metastasis (pN0).
PROXIMITY TO NEAREST MARGIN: - 7.5 cm.
SIZE: - 0.6 x 0.5 x 0.2 cm.
LYMPHATIC/VASCULAR INVASION: - Negative.
PERINEURAL INVASION: - Negative.
MARGINS: - Negative.

Specimen is received in a container labelled "sigmoid colon" and shows a 21 cm segment of colon with adjacent mesenterial fat. Resection margins are unlabelled. The specimen was previously opened and fixed and on a board and it shows the area of ulceration of the mucosa measuring 0.5 x 0.6 cm located at the distance of 7.5 cm from one resection margin and 12.5 cm from the opposite resection margin. The lesion is located at the mesenterial edge of the colonic wall. The rest of the mucosa is grossly unremarkable. Cut surface of the colonic mucosa through the lesion reveals intramucosal lesion measuring 0.6 x 0.5 x 0.2 cm. Representative sections are submitted as follows:
'a' - one resection margin.
'b' - second resection margin.
'c', 'd' and 'e' - serial sections through the entire lesion.
Multiple small grossly unremarkable lymph nodes were found in the mesenterial fat measuring 0.3 cm in greatest dimension. Eight of them were sampled and submitted in cassettes 'f' through 'j'.
Received in the same container is the stem consisting of circular fragments of the mucosa measuring 2.5 x 1.5 x 1.8 cm. One representative section is submitted in 'k'.
gh

The sections show superficial adenocarcinoma invading the lamina propria. In-situ areas are seen adjacent. The tumor is well differentiated. Lymph nodes are free of tumor.

Logged in as 'admin'

NCI pTNM Study

- Examined 250 ePath reports from each of 5 sites (breast, colon, lung, ovary, and prostate). Reports came from 4 SEER registries (Detroit, Georgia, Kentucky, and Louisiana).

- The randomly selected free text ePath reports were processed through the AIM E-PATH RCA application.

Question	Answer
Additional dimension...	1.0
Additional dimension...	1.3
HISTOLOGIC TYPE (chec...	Invasive ductal carcinoma
HISTOLOGIC GRADE...	
Type of Grading System	
Tubule formation	
Nuclear pleomorphism	
Mitotic count	
For a 25x objective ...	
For a 40x objective ...	
Nottingham System	
Nottingham Total Sc...	
Nottingham Grade	
Bloom Richardson ...	
Bloom Richardson S...	
Bloom Richardson G...	
Number of mitoses p...	
PATHOLOGIC STAGI...	
PRIMARY TUMOR (pT)	pT1mic: Microinvasion 0.1 cm or less in greatest dim...
REGIONAL LYMPH NO...	pN1a: Metastasis in 1-3 axillary lymph nodes (at least...
Number of regional l...	12
Number of regional l...	3
DISTANT METASTASI...	pMX: Cannot be assessed
MARGINS AND INVA...	
MARGINS (check all th...	
Status of margin inv...	Margins uninvolved by invasive carcinoma
Distance from clo...	0.09
Uninvolved margi...	
Status of margin inv...	
VENOUS/LYMPHATIC (L...	Absent

1. Left breast, lumpectomy: Diminutive aggregates of atypical cells consistent with residual invasive ductal carcinoma, grade 3, are identified (tissue section 1E). Immunohistochemistry reveals expression of pan-cytokeratin by these atypical cells, a finding that confirms their epithelial nature. The largest aggregate of atypical cells measures 0.7 mm in greatest dimension. These aggregates are embedded in a fibrotic tumor bed that measures 18 x 13 x 10 mm. The surgical margin is uninvolved by invasive carcinoma; the closest approximation of the residual foci of invasive carcinoma to the surgical margin (posterior) is 0.9 mm. The fibrotic tumor bed contacts the posterior surgical margin. No residual ductal carcinoma in situ is identified. Extensive fibrocystic change and sclerosing adenosis are observed elsewhere in the specimen. 2. Extra posterior margin, resection: No evidence of malignancy. Skeletal muscle is identified. 3. Extra medial margin, resection: No evidence of malignancy. 4. Extra superior margin, resection: No evidence of malignancy. 5. Extra inferior margin, resection: No evidence of malignancy. 6. Extra lateral margin, resection: No evidence of malignancy. 7. Extra anterior margin, resection: No evidence of malignancy. 8. Axillary dissection: Twelve lymph nodes are identified microscopically. Two are involved by tumoral metastasis. The largest continuous focus of tumoral metastasis measures approximately 4 mm in greatest dimension. There is no evidence of extra nodal extension. Surgical Pathology Cancer Case Summary Procedure: Excision without wire guided localization Lymph node sampling: Axillary dissection (partial or complete dissection) Specimen laterality: Left Tumor site: Not specified Histologic type: Invasive ductal carcinoma (no special type) Tumor size: Diminutive aggregates of invasive carcinoma, the largest measures 0.7 mm in greatest dimension Histologic grade: Grade 3. Given the small tumoral size, the Nottingham histologic score cannot be accurately established Tumor focality: (unifocal) few residual aggregates of invasive carcinoma are identified in a single tumoral bed Ductal carcinoma in situ: Not identified Lobular carcinoma in situ: Not identified Macroscopic and microscopic extent of tumor Skin: Skin is not present in the specimen Nipple: The nipple is not present in the specimen Skeletal muscle: The skeletal muscle is uninvolved Margins Margins are uninvolved by invasive carcinoma The distance of invasive carcinoma to the closest margin in the lumpectomy specimen is 0.9 mm (posterior margin); the additional posterior surgical margin obtained as specimen #2 measures 1 cm in width Lymph nodes Total number of lymph nodes examined (sentinel and nonsentinel): 12 Number of lymph nodes with macrometastasis: 2 Number of lymph nodes with micrometastasis: 0 Number of lymph nodes with isolated tumor cells: 0 Extranodal extension: Not identified Treatment effect In breast: Probable or definite response to presurgical therapy in invasive carcinoma In lymph nodes: Probable or definite response to presurgical therapy in metastatic carcinoma Lymphovascular invasion: Not identified Pathologic Staging Primary tumor: ypT1mi Regional lymph nodes: ypN1a Distant metastasis: Not applicable Additional pathologic findings: Fibrocystic change and sclerosing adenosis Ancillary studies: Performed in previous biopsy (see specimen AS13-8050) Microcalcifications: Not identified Clinical history: Not provided BR/sw

M69700	M80103	M85003	P1100	T04030	M49000	M74220	M74320
M80001	T00100	M09400	T13000	M09400	M09400	M09400	

1. In formalin, labeled with the patient's name and medical record number, and designated as "left breast", is a flat, roughly oval segment of tan-yellow fibroadipose tissue, partially fragmented, weighing 72 gm. A short single stitch marks the superior aspect, a long single stitch the lateral and a short double stitch the anterior. The specimen measures 6.5 cm from superior to inferior, 5.4 cm from medial to lateral, 2.8 cm from anterior to posterior. The surface is inked as follows: posterior black, anterior blue, superior green, inferior red, medial orange; lateral yellow. Sectioning reveals extensive areas of fibrosis spanning the entire specimen. A tan-yellow round nodule, 1 x 0.5 x 0.3 cm, is identified on the on the lateral end of the specimen, abutting the

Availability
Cases w/surgical resections only

Site	T	N	M	Sample
Breast	87.2%	88.0%	46.4%	250
Colon	90.0%	89.6%	66.0%	250
Lung	86.4%	88.8%	52.4%	250
Ovary	86.0%	87.6%	56.8%	250
Prostate	95.2%	95.2%	56.8%	250

Sensitivity/Specificity/PPV/NPV

Site	Item	Sensitivity	Specificity	PPV	NPV
Breast	T	98.8%	98.9%	83.8%	99.9%
	N	98.3%	99.3%	88.9%	99.9%
	M	96.6%	99.4%	96.6%	99.4%
Colon	T	99.1%	98.8%	86.7%	99.9%
	N	97.8%	98.5%	88.2%	99.8%
	M	98.7%	98.5%	91.3%	99.8%
Lung	T	99.1%	99.3%	92.9%	99.9%
	N	98.7%	98.8%	94.9%	99.7%
	M	93.0%	97.8%	84.6%	99.1%
Ovary	T	98.3%	99.9%	81.4%	99.9%
	N	96.0%	97.0%	93.5%	98.3%
	M	86.3%	97.8%	91.5%	96.3%
Prostate	T	99.6%	98.5%	84.3%	100.0%
	N	99.6%	97.5%	94.8%	99.8%
	M	88.1%	98.4%	88.6%	98.3%

Sensitivity/Specificity/PPV/NPV Revisited

Site	Item	Sensitivity	Specificity	PPV	NPV	
Breast	T	98.8%	100.0%	99.6%	99.9%	removed 3 cases from 1 lab
	N	98.2%	100.0%	99.6%	99.9%	
	M	96.6%	99.8%	99.1%	99.4%	
Colon	T	98.7%	100.0%	100.0%	99.9%	removed 8 cases from 1 lab
	N	97.8%	100.0%	100.0%	99.7%	
	M	98.8%	100.0%	100.0%	99.8%	
Lung	T	99.1%	100.0%	100.0%	99.9%	removed 2 cases from 2 labs
	N	99.7%	99.6%	98.2%	99.7%	
	M	92.8%	98.1%	85.9%	99.1%	
Ovary	T	98.7%	99.5%	92.6%	99.9%	removed 4 cases from 1 lab
	N	95.9%	97.9%	95.1%	98.3%	
	M	86.3%	99.0%	95.8%	96.3%	
Prostate	T	99.6%	99.9%	98.4%	100.0%	removed 6 cases from 4 labs
	N	99.7%	99.8%	99.6%	99.8%	
	M	88.9%	100.0%	100.0%	98.4%	

Can coding be automated?

CLINICAL HISTORY/MACROSCOPY
Right mastectomy and axillary tissue. A right mastectomy specimen with overlying skin measuring 220mm x 85mm and underlying breast tissue measuring 220mm x 100mm x 70mm. The axillary tail measures 125 x 60mm. The nipple is slightly retracted and located centrally. The superior margin is painted red, the inferior margin painted green and the deep cut margin is painted blue. Cut sections of the underlying breast tissue shows an ill-defined grey white yellow lesion with patchy areas of haemorrhage measuring 35 x 35 x 35mm located immediately below the nipple, 20mm from the inferior margin, 45mm from the deep cut margin, 50mm from the superior margin, 85mm from the medial margin and 100mm from the lateral cut margin. A1 - nipple, B1 - upper outer quadrant, C1 - upper inner quadrant, D1 - lower outer quadrant, E1 - lower inner quadrant, F1, G1 - tumour composite blocks, H1, I1 - tumour composite blocks, J1 - deep cut margin, K1 - superior margin, L1 – inferior margin, M4 - lymph nodes, N4 - lymph nodes, O - 3 serial slices, 1 lymph node, P - 3 lymph nodes.

M-80103
M-85003
M-80003
C50.9
C77.9

MICROSCOPY
This right mastectomy specimen demonstrates an invasive ductal carcinoma with the following pathological features:

TUMOUR HISTOLOGY & GRADE
The tumour is of an infiltrating poorly differentiated ductal carcinoma of non-otherwise specified type. The tumour is poorly defined and extremely infiltrative, comprising poorly-formed tubules, nests or strands of cuboidal tumour cells displaying high grade nuclei. The tumour cells are set within fibrotic desmoplastic stroma. Many lactiferous ducts are entrapped within the tumour. Frequent tumour mitoses are seen. Microcalcification is seen in some neoplastic tubules.

Tumour grade (Modified Bloom-Richardson Scoring System):

Tubular formation:	3
Nuclear atypia:	3
Tumour mitoses:	2
Total score:	8 (Grade III)

TUMOUR LOCATION, SIZE AND EXTENT
The tumour is located 5mm below the nipple and has a macroscopic size of 35mm across. The border of the tumour is poorly circumscribed and infiltrative.

INTRA-LYMPHOVASCULAR OR PERINEURAL TUMOUR PERMEATION
Focal intralymphatic tumour permeation is noted. No

Coding Accuracy Reference Builder - Multiple Files

File Update Reference Tools

☑	600140...	IGNOR
☑	FWDN...	MTLG
☑	FWDNO...	MTLG
☑	CPAC...	MLG
☑	CPACP...	MTLG
☑	KY_AP-...	G
☑	KY_A...	MTLG
☑	KY_A...	TLG
☑	KY_A...	MTLG
☑	KY_AP-...	MTLG
☑	KY_AP-...	MTLG
☑	KY_AP-...	MTLG
☑	KY_AP-...	MTLG
☑	KY_A...	TLG
☑	KY_AP-...	MTLG
☑	KY_A...	LG
☑	KY_A...	MLG
☑	KY_A...	MTLG
☑	KY_AP-...	MTLG
☑	KY_AP-...	MTLG
☑	KY_AP-...	MTLG
☑	KY_AP-...	LG
☑	KY_AP-...	MTLG
☑	KY_AP-...	TLG
☑	KY_AP-...	L
☑	KY_AP-...	MTLG
☑	KY_AP-...	MTLG
☑	KY_AP-...	MTLG

[OTHER]

NONE AVAILABLE.

TEXT DIAGNOSIS FROM PRIOR CASES:
KC08-48 Accsnd 09JAN2008, Signed Out 24JAN2008, Pathologist CTA
AMENDED DIAGNOSIS - SPECIMEN SOURCE

RIGHT BRONCHIAL **WASH**:
-NON-SMALL CELL CARCINOMA.
-INTRADEPARTMENTAL CONSULTATION WITH AGREEMENT IS OBTAINED.

CTA/erm
KC08-30 Accsnd 08JAN2008, Signed Out 09JAN2008, Pathologist SPIRE
RIGHT LUNG, BRONCHIAL BRUSHING, THIN LAYER PREPARATION AND CELL BLOCK:
-POSITIVE FOR MALIGNANCY,
KERATINIZING SQUAMOUS CELL CARCINOMA PRESENT.

COMMENT: Correlation with biopsy is confirmatory of the diagnosis.
A) Bronchial Brushing
Tissue Category source: Respiratory
Right bronchial brushing. Two thin preps. Cell block prepared. Specimen received fixed in Cytolyt. 35 ml. Mucoid specimen. Tissue fragments present. Routine AFB-fungus. C-4-08
jck
The tumor is notable for necrosis, numerous malignant squamous cells present, notable for glassy eosinophilic cytoplasm. Intradepartmental consultation with agreement is obtained.
SES/erm
RIGHT LUNG, BRONCHIAL BRUSHING, THIN LAYER PREPARATION AND CELL BLOCK:
-POSITIVE FOR MALIGNANCY,
KERATINIZING SQUAMOUS CELL CARCINOMA PRESENT.

Automatic Coding

Criteria	Value
Morphology	M-80463
Topography	C34.9
Laterality	1
Grade	9

Manual Coding

Criteria	Value
Morphology	M-80713
Topography	C34.9
Laterality	1 Right origin
Grade	9 Unknown or N/A

☑ Reviewed

Comments

Morphology: upon review coder changed to Keratinizing squamous cell carcinoma 80713. There are actually two specimens noted within this report, one of which has a diagnosis of non-small cell carcinoma (80463), however, the more specific diagnosis of keratinizing squamous cell carcinoma would be used to code the case.

Raw Output

Criteria	Value
Morphology	M-80103
Morphology	M-80413
Morphology	M-80463
Morphology	M-80703
Morphology	M-80713
Morphology	M-80001
Morphology	M-80011
Morphology	M-80013
Topography	C34.9
Laterality	1
Grade	9

C# | CLips

CODEX Explanation

Priority 4: C Code in Associated Link

M-80463 C34.9: Priority 2

Associated Link Information:
****

Step 8: Remove Same Family CCode in order C**.9 which means if Cxy.z and Cxy.9 then remove Cxy.9, Occurence and no Associate Link:

Associated Links Informations:

****

Step 9: Remove M code behavior 0/1 if have M code behavior 2/3:

****

Find :

Processing Reports...

Knowledge base examples

Reduction rules (general/specific)

- If the report contains a C42(0/1/2) (Blood/Bone Marrow/Spleen) then assume this is a Leukemia report and remove codes that are non Leukemia morphologies

- IF M-83233 Mixed cell adenocarcinoma THEN REMOVE M-81403 Adenocarcinoma NOS

Deduction rules to add codes

- DEDUCE M-85223 Infiltrating duct and lobular carcinoma (C50._) IF M-82303 Solid adenocarcinoma with mucin formation AND M-85003 Duct adenocarcinoma NOS

Knowledge Acquisition

Report #1..	Morphology: Diagnosis states endocrine carcinoma. This terminology does not exist in ICD-O-3. Consensus opinion is 8246 neuroendocrine carcinoma. this is not certain.
'FWDNORTONC0000008475'	Morphology: adenocarcinoma. code 81403. Do not use 81443 unless stated as "intestinal type".
'FWDNORTONC0000008521'	Morphology: diagnosis states adenocarcinoma in 6a. 81403. do not code 8144 unless stated as adenocarcinoma, intestinal type.
'FWDNORTONC0000008526'	Morphology: Diagnosis 5c states DCIS is solid, cribriform and comedo TYPES. 85232 Using 2007 multiple primary and histology rule H6, if there is intraductal carcinoma and tow or more specific intraductal types from table 3 pg 50, then use 85232.

Knowledge Acquisition

'FWDNORTONC0000008467'	Morphology: Diagnosis states endocrine carcinoma. This terminology does not exist in ICD-O-3. Consensus opinion is 8246 neuroendocrine carcinoma. this is not certain.
'FWDNORTONC0000008475'	Morphology: adenocarcinoma. code 81403. Do not use 81443 unless stated as "intestinal type".
'FWDNORTONC0000008521'	Morphology: diagnosis states adenocarcinoma in 6a. 81403. do not code 8144 unless stated as adenocarcinoma, intestinal type.
'FWDNORTONC0000008526'	Morphology: Diagnosis 5c states DCIS is solid, cribriform and comedo TYPES. 85232 Using 2007 multiple primary and histology rule H6, if there is intraductal carcinoma and tow or more specific intraductal types from table 3 pg 50, then use 85232.

Radiotherapy reports

Extract certain data from radio therapy such as:

- Karnofsky performance status
- Number of fractions (radiation)
- Radiation Dose per fraction
- Symptoms and side effects
- Total Radiation Dose (cGy)

IMPRESSION: Ms. XXXXX is a 70-year-old postmenopausal female with a pT1c N0s M0 right breast infiltrating mammary carcinoma, 1.5 cm, grade I, ER and PR positive, Ki-67 less than 10%, and Her-2/neu not over-amplified, and 0 of 2 lymph nodes involved. She is status post lumpectomy and sentinel lymph node biopsy (1/12/09), and adjuvant accelerated partial breast irradiation prescribed to a dose of 34 Gy in 10 fractions given twice daily delivered using a SAVI catheter (1/30/09). She declined hormonal therapy.

Dear Drs. Rita Lechleitner, George Labban, William Broad, Nathan Becker, Gang Li, and Kim-Phuong Dang:
Ms. XXXXX returns today for follow-up after having last been seen 6 months ago. In the interim, she denies any new problems or complaints. She continues to have discomfort in the right upper outer breast region which remains persistent. Overall, this is stable without any significant worsening. It is described as shock-like sensations starting in the upper outer breast radiating towards the nipple region. She also has tenderness over the fibrotic area. She continues followup with Dr. Gang Li with continued Neurontin, topical lidocaine cream and Vicodin only as needed for severe pain. Her activity is still somewhat limited although she continues to be able to walk and ski without significant difficulty with exacerbation of her pain. She denies any changes on self-breast examination. Her bilateral mammogram in January revealed no evidence of disease. There were no other pertinent positives on the complete review of systems.
PHYSICAL EXAMINATION:
Vitals: Her weight is 150.1 pounds and blood pressure is 136/72.
General: She is a well-developed female in no acute or apparent distress. Her KPS is 90.
Neck: Supple without cervical or supraclavicular lymphadenopathy.
Lungs: Clear to auscultation bilaterally.
Cardiovascular: Examination showed normal sinus rhythm.
Abdomen: Soft, nontender and nondistended.
Musculoskeletal: There is no tenderness to palpation over the spine or long bones.
Breasts: The breasts are examined in the seated and supine positions. She has well healed right upper outer breast lumpectomy and catheter insertion scar that

Processor's results | Processor's log

Question	Answer	External Coding	Value	Rank	Heuristic Vector
Lymph Nodes Positive	Lymph Nodes Positive (#)	-	0	0	(dist:1.00,queExpDnsty:1.00,queEx...
Lymph Nodes Positive	Lymph Nodes Positive (#)	-	0	0	(dist:0.99,queExpDnsty:1.00,queEx...
HX	HX	-	status post	1	(dist:1.00,ansPri:1.00,ansExpDnsty:...
Radiation Dose per fraction (Gy)	Dose per fraction (cGy)	-	34	0	(dist:0.99,queExpDnsty:0.60,queEx...
PRK	PRK		stable	1	(dist:1.00,ansPri:1.00,ansExpDnsty:

Break

Heuristic Scoring – low level

- A heuristic scoring system will allow the expressions to be ranked and compared for overall strength.
- Each component will have an associated scoring method that is normalized to the interval [1, 0).

- **Total score = Score1 * Score2 * Score3…** new scoring components can be added as needed

- **Score scaling:** Not all score components have the same influence on the total score. For example, expression ordering is less significant that density or hierarchy scoring. A scaling factor can be applied to scoring component to reduce the impact of the particular score

Some simple heuristics

Expression density score

The positional length of the expression is calculated as the size of the concept in C-Score co-ordinates.

Expression length = end – start + 1

Carcinoma = 1-1 +1 = 1
Carcinoma = 127 – 127 +1 =1
Squamous cell carcinoma = 127-125 + 1 = 3

The density of the expression is calculated like this

Density = Number of words in expression / Expression length

Some simple heuristics

Word ordering score

Expressions that required a fixed ordering of keywords will not become complete unless the keywords are correctly ordered (when complete, these expressions will receive an order score of 1).

Expressions that do not require keyword ordering will be assigned an order score based on a heuristic like this.

Order score = keywords out of order / total keywords.

Scaling factor:

Scaled order score = (keywords out of order + K)/ (total keywords + K)

Where K is the scaling factor for order scoring from [0, large value]. A larger K value will reduce the impact of the order score to the total score.

Eg: Say K=10, total key words = 4, keywords out of order =2

Order score = 2 / 4 = .5

Scaled order score = (2+10)/(4 + 10) = .86

Example – heuristic scoring

"Carcinoma is often found on the skin. Both Basal and squamous types can appear on skin, but of the two, squamous cell carcinoma is more significant."

But this time we want to score the expression "skin carcinoma"

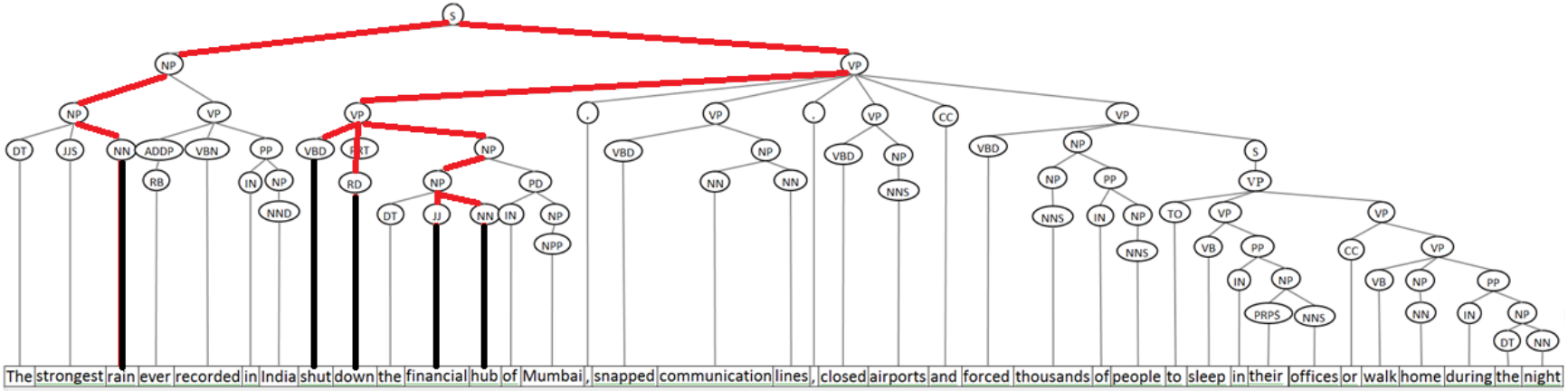
We get the following two expressions:

Carcinoma:1, Skin:7 CScore = 7-1+1 = 7 density = 2/7 (29%)
Skin:67, Carcinoma:127 CScore = 127-67 +1 = 61 **over limit

So in the first result, even though the density is only 29%, it is still a valid expression. If two expressions for the same concept exist but have different densities, the density values are combined with other indicators (such as priorities) to get an overall "strength" measure for each expression.

More experimentation

- Combining the ideas of heuristic segmentation and grammar parsing. Build a constituent tree and use this to



General processor

Redesign AutoCode Engine to incorporate Synoptex technology

- Attribute value identification
- Smart segmentation
- Table detection
- Numeric processing
- Data formatting (general methods)
- Scoring heuristics

New Design – Units Knowledge base

A unit is the representation of an Attribute and its possible values.

- A unit is self contained (contains all the information required to identify and extract the information)
- The processing of a unit is done by the new AutoCode engine
- Units do not care about any particular domain or application
- Units are developed, tested and stored in a library
- Performance measurement is done by the unit

Knowledge representation - UNITS

- New representation scheme is framed around questions and answers (Q/A) such as “histologic type” rather than a “checklist”
- Each unit is defined as the chunk of knowledge that is specific to each question or the data of interest
- Units are developed and performance tested independently of the underlying NLP search algorithms.
- Added inference capabilities allows a higher level of reasoning
  - e.g. recurrence, or automated coding

Stage 1 - Efficient Search

Generate lots of information

Combined techniques – efficient search ning combined with analysis and pattern dete identify the raw data contained in the report

- Text processing
  - Concept identification / value matching
  - Uses language elements to narrow search space
  - Uses context, not just key word identification
- Pattern recognition
  - Identifies various table formats
  - Lists (numeric, labeled, unlabeled)
  - Other unstructured data aggregates
  - Used in further constraining the search algorithms
- Numeric extraction
  - Associates correct numbers to data names
  - Considers units, dimension, type, range
  - Identifies and ignores extraneous numbers

Stage 1 - Heuristic segmentation and search

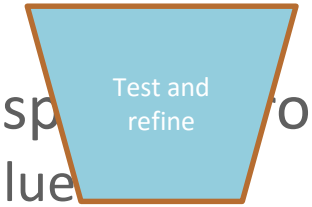
In situations where grammar is not properly formed, this technique helps to bound the concepts as well as assign negation context.

Fast (one pass) search algorithm that use heuristics to reduce the search space

- The UNITS technology implements what we call "heuristic segmentation" where the boundaries of where to search for concepts are defined by heuristics rather than grammar trees.
  - Grows concepts when potential keywords are found
  - Prunes concepts when keywords go out of scope ad defined by the heuristic segmentation

This stage "generates" what can be considered the "RAW" data - all the candidate concepts, numbers and things that can be found within the text.

Stage 2 - Test and Refine



A knowledge base that constrains data sp operties

- Numeric units and ranges, allowed value
- Conflict resolution, negation determination.
- Question/answer mapping

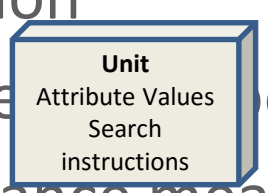
This stage examines the volume of data produced by "Generate" algorithms and refines or removes errant data.

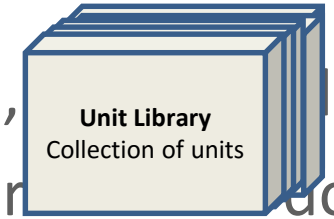
- Each data item has it's own properties and rules – independent of the general NLP methods. We call these pieces of knowledge "UNITS"
- Units are the direct coding of knowledge about how to identify and interpret the correct information for that data item

Stage 2 - Units Knowledge base

A unit is the representation of an Attribute and its possible values.

- A unit is self contained (contains all the information required to identify and extract the information
- Units do not care about any particular domain or application
- Units are [illegible]ed, [illegible] stored in a library
- Performance measur[illegible] done at the unit level





Stage 3 - Inference engine

- This is where the higher level operations take place
- Production rules based system using inference engine to examine the resulting information and make decisions.
- Used to develop expert systems - rules can address any level of logic complexity.
- Examples
  - Case-selection
  - CNS classification,
  - Site identification,
  - Recurrence **
  - Automated coding **

Agenda

- Brief NLP Discussion
- AutoCode Enhancements
- Synoptex Enhancements
- Knowledge Base Manager
- CAP Checklist Updates
- Clinical Trials Matching
- Some Weird Stuff

What about pathology reports?

Can we use NLP techniques for report classification?

AUTOCODE

CLINICAL HISTORY/MACROSCOPY
Right mastectomy and axillary tissue. A right mastectomy specimen with overlying skin measuring 220mm x 85mm and underlying breast tissue measuring 220mm x 100mm x 70mm. The axillary tail measures 125 x 60mm. The nipple is slightly retracted and located centrally. The superior margin is painted red, the inferior margin painted green and the deep cut margin is painted blue. Cut sections of the underlying breast tissue shows an ill-defined grey white yellow lesion with patchy areas of haemorrhage measuring 35 x 35 x 35mm located immediately below the nipple, 20mm from the inferior margin, 45mm from the deep cut margin, 50mm from the superior margin, 85mm from the medial margin and 100mm from the lateral cut margin. A1 - nipple, B1 - upper outer quadrant, C1 - upper inner quadrant, D1 - lower outer quadrant, E1 - lower inner quadrant, F1, G1 - tumour composite blocks, H1, I1 - tumour composite blocks, J1 - deep cut margin, K1 - superior margin, L1 – inferior margin, M4 - lymph nodes, N4 - lymph

M-80103
M-85003
M-80003
C50.9
C77.9

ARTIFICIAL INTELLIGENCE IN MEDICINE INC.

The Challenge: Automated coding

CLINICAL HISTORY/MACROSCOPY
Right mastectomy and axillary tissue. A right mastectomy specimen with overlying skin measuring 220mm x 85mm and underlying breast tissue measuring 220mm x 100mm x 70mm. The axillary tail measures 125 x 60mm. The nipple is slightly retracted and located centrally. The superior margin is painted red, the inferior margin painted green and the deep cut margin is painted blue. Cut sections of the underlying breast tissue shows an ill-defined grey white yellow lesion with patchy areas of haemorrhage measuring 35 x 35 x 35mm located immediately below the nipple, 20mm from the inferior margin, 45mm from the deep cut margin, 50mm from the superior margin, 85mm from the medial margin and 100mm from the lateral cut margin. A1 - nipple, B1 - upper outer quadrant, C1 - upper inner quadrant, D1 - lower outer quadrant, E1 - lower inner quadrant, F1, G1 - tumour composite blocks, H1, I1 - tumour composite blocks, J1 - deep cut margin, K1 - superior margin, L1 – inferior margin, M4 - lymph nodes, N4 - lymph nodes, O - 3 serial slices, lymph node, P - 3 lymph nodes.

M-80103

M-85003

M-80003

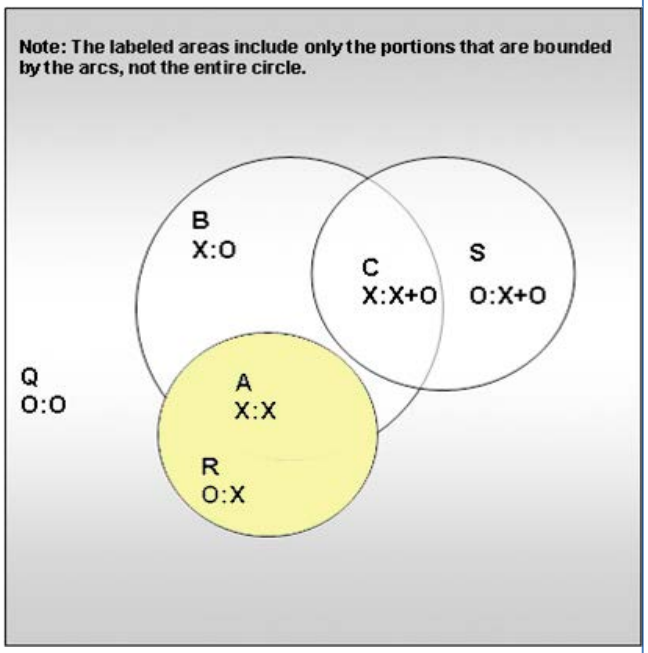
C50.9

C77.9

MICROSCOPY
This right mastectomy specimen demonstrates an invasive ductal carcinoma with the following pathological features:

TUMOUR HISTOLOGY & GRADE
The tumour is of an infiltrating poorly differentiated ductal carcinoma of non-otherwise specified type. The tumour is poorly defined and extremely infiltrative, comprising poorly-formed tubules, nests or strands of cuboidal tumour cells displaying high grade nuclei. The tumour cells are set within fibrotic desmoplastic stroma. Many lactiferous ducts are entrapped within the tumour.
…

Step 1: Create Measuring System



Coding accuracy measures

Sensitivity = (A+C) / (A+B+C)

How often is the reference code returned in those reports where it is the subject code.

Specificity = Q / (Q+R+S)

How often is the subject code not returned in those reports where it is not the reference code.

Reducibility = (A+R) / (A+R+C+S)

How often is the subject code the only code identified when the code is identified.

Confidence = A / (A+R)

How much confidence can we place in the result.

Agenda

- Brief NLP Discussion
- AutoCode Enhancements
- Synoptex Enhancements
- Knowledge Base Manager
- CAP Checklist Updates
- Clinical Trials Matching
- Some Weird Stuff

What about pathology reports?

Can we infer accurate attribute-value data

SYNOPTEX

CLINICAL HISTORY/MACROSCOPY
Right mastectomy and axillary tissue. A right mastectomy specimen with overlying skin measuring 220mm x 85mm and underlying breast tissue measuring 220mm x 100mm x 70mm. The axillary tail measures 125 x 60mm. The nipple is slightly retracted and located centrally. The superior margin is painted red, the inferior margin painted green and the deep cut margin is painted blue. Cut sections of the underlying breast tissue shows an ill-defined grey white yellow lesion with patchy areas of haemorrhage measuring 35 x 35 x 35mm located immediately below the nipple, 20mm from the inferior margin, 45mm from the deep cut margin, 50mm from the superior margin, 85mm from the medial margin and 100mm from the lateral cut margin. A1 - nipple, B1 - upper outer quadrant, C1 - upper inner quadrant, D1 - lower outer quadrant, E1 - lower inner quadrant, F1, G1 - tumour composite blocks, H1, I1 - tumour composite blocks, J1 - deep cut margin, K1 - superior margin, L1 – inferior margin, M4 - lymph nodes, N4 - lymph nodes, O - 3 serial slices, lymph node, P - 3 lymph nodes.

MICROSCOPY
This right mastectomy specimen demonstrates an invasive ductal carcinoma with the following pathological features:

TUMOUR HISTOLOGY & GRADE
The tumour is of an infiltrating poorly differentiated ductal carcinoma of non-otherwise specified

Synoptic Element	Data Value
Specimen Type	Mastectomy
Lymph Node Sampling **	Axillary dissection
Lymph Nodes Examined **	13
Lymph Nodes Positive **	3
Lymph Nodes Negative	
Extranodal extension	
Specimen Size **	220x85 mm
Laterality	Right
Tumor Site **	Lower inner quadrant
Size of Invasive Component **	35x35x35 mm
Invasive Component Greatest dimension	
Histologic Type	Invasive ductal carcinoma
	Ductal carcinoma in situ
	Paget disease without invasive carcinoma
	Cribriform
	Tubular
Grading System	Bloom Richardson
Histological Grade	Grade III
Tubule Formation **	3
Nuclear Pleomorphism **	3
Mitotic Count **	10
Necrosis	
Pathologic Staging (pTNM)	
HER2 Status **	Negative
ER Status **	Positive
PR Status **	Positive
Margin **	
Distance from Deep Margin **	45 mm
Distance from Lateral Margin **	50 mm
Distance from Superior Margin **	50 mm
Distance from Inferior Margin **	20 mm
Distance from Medial Margin **	50 mm
Distance from Anterior Margin	
Distance from Posterior Margin	
Location of involved margin by invasive ...	
DCIS Margin Involvement	
Distance from uninvolved DCIS closest ...	
Location of uninvolved DCIS margin	
Location of involved DCIS margin	
Extent of Margin Involvement for Invasi...	
Extent of Intraductal Component	
DCIS Extent of Intraductal Component	
Extent of Margin Involvement for DCIS	
Lymphatic/Venous Invasion	Lymphovascular Invasion Present
Microcalcifications **	Microcalcifications present
Additional Pathologic Findings	

The Challenge:

Reports do not always follow grammar rules.

Build a table detection system

Specimen type : Hemicolectomy, right, partial excision of urinary bladder, and salpingo-oophorectomy, right
Orientation
Clinical information : Colon, right
Anatomical landmarks : Ileum = proximal
Surgical markings : Absent
Pathlogy markings : Ink = margins of urinary bladder wall

Ileum, terminal
Length : 8 cm
Serosa : Unremarkable
Wall : Thickness: 0.4 cm
Mucosa : Slightly edematous

Vermiform appendix : Absent

Colon, right
Length : 22 cm
Serosa : Cecum: Puckering and attached portion of urinary bladder wall and right adnexa (see below)
Otherwise: Unremarkable
Wall : Thickness: 0.4 cm
Mucosa : Tumour:
- Location: Ileocecal valve
- Configuration: Fungating
- Size:
  - Length: 7 cm
  - Width: 12 cm
  - Thickness: 7 cm
  - Extent of invasion: Urinary bladder
- Margins:
  - Proximal: 8 cm
  - Distal: 15 cm
  - Radial: 0.1 cm`

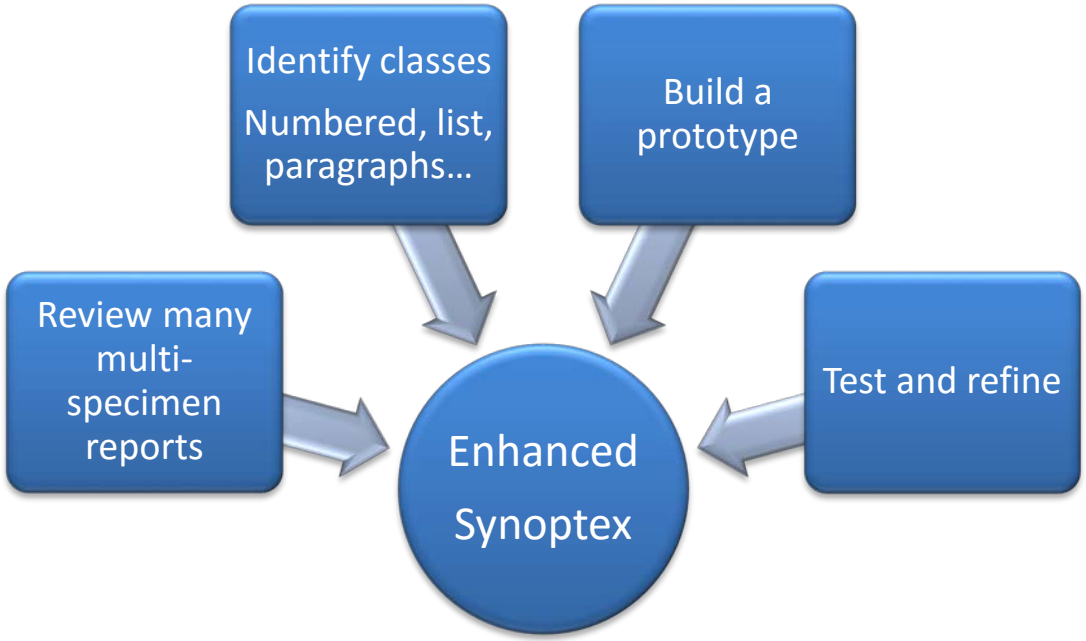
Non-neoplastic colon: Unremarkable

AIM ARTIFICIAL INTELLIGENCE IN MEDICINE INC.

The Challenge:

How do we handle multiple specimen reports?

Build a classifier that identifies and extracts the correct data.



Synoptex - Reference_June_02_2008

File Edit View Tools Help

Breast: General

Auto Detect | Process Multiple Specimens | 1 | 2

1001100000147618
1001100000147714
1001100000147950
1001100000147988
1001100000148170
1001100000148192
1001100000148299
1001100000148316
1001100000148439
1001100000148457
1001100000148458
1001100000148482
1001100000148588
1001100000148661
1001100000148662
1001100000148663
1001100000148718
1001100000148719
1001100000148798
1001100000149128
1001100000149317
1001100000149338
1001100000149344
1001100000149360
1001100000149371
1001100000149398
1001100000149399
1001100000149625
1001100000149629
1001100000149769
1001100000149788
1001100000149810
1001100000149879
1001100000149942
1001100000150009
1001100000150036
1001100000150088
1001100000150171
1001100000150265
1001100000150302
1001100000150374
1001100000150383
1001100000150397
1001100000150398
1001100000150432
1001100000150435
1001100000150441
1001100000150454
1001100000150557
1001100000150571
1001100000150583
1001100000150940

Synoptic Element	Data Value
Specimen Type	Lumpectomy
Lymph Node Sampling	
Lymph Nodes Examined	
Lymph Nodes Positive	
Lymph Nodes Negative	
Extranodal Extension	
Specimen Size **	3.5x3.3x1 cm
Tumor Laterality	Right Tumor Laterality
Tumor Site	
Tumor Size **	3.8 mm
Tumor Greatest dimension **	3.8 mm
Histologic Type	Ductal carcinoma in situ
Grading System	
Histologic Grade	Grade III
Tubule Formation	
Nuclear Pleomorphism	
Mitotic Count	
Necrosis **	Present
Pathologic Staging (pTNM)	
HER2 Status	
ER Status	
PR Status	
Margins **	Margins Uninvolved by Invasive Carcinoma
Distance of DCIS from Margin **	6.5 mm
Distance of Tumor from Deep ...	
Distance of Tumor from Lateral...	6.5 mm
Distance of Tumor from Superi...	
Distance of Tumor from Inferior...	
Distance of Tumor from Medial...	
Distance of Tumor from Anteri...	
Distance of Tumor from Posteri...	
Involved Margins by Invasive ...	
Uninvolved Margins by DCIS	
Involved Margins by DCIS	
Extent of Margin Involvement f...	
Extent of Margin Involvement f...	
Extent of Intraductal Compone...	
Extent of Intraductal Compone...	
Lymphatic Invasion	
Venous Invasion	
Perineural Invasion	
Microcalcifications **	Microcalcifications in DCIS
	Microcalcifications present
Additional Pathologic Findings	Fibroadenomatoid Change
	Fibrosis

Clinical Diagnosis and History
Right breast cancer.
Specimen(s) Received:
1: Right Breast mass
2: Right breast sentinel node
Gross Description
Received fresh for frozen section diagnosis evaluation of margins and designated "right breast **lumpectomy**" is a 2.5 x 0.4 cm tan skin ellipse overlying a 6 x 2 cm pad of yellow fatty tissue excised to a depth of 4 cm. Tags are present on the specimen designating the margins. A firm area is felt in the medial aspect of the specimen. Sectioning through this area reveals a cavity feel with a hematoma which is 3 cm in greatest dimension. A section from this area is frozen on a single block. The remaining frozen tissue is submitted in (Block 1A). The margins of the specimen are painted with blue ink. Additional sections from the medial margin are submitted in (Blocks 1B-1D). Sections from the lateral margins are submitted in (Blocks 1E-1G). Sections from the cranial margins are submitted in (Blocks 1H-1J). Sections from the caudal margins are submitted in (Blocks 1K-1L). Sections from the deep margin are submitted in (Blocks 1M-1P). Sections from the wall of the previous biopsy are submitted in (Blocks 1Q-1S). A section of overlying skin is submitted in (Block 1T).
Submitted in formalin and designated "right breast sentinel node (tissue with purple ink not reactive but in same area as sentinel node)" is yellow fatty tissue measuring 3.5 x 3.3 x 1 cm in aggregate. There appear to be three lymph nodes in this fatty tissue. One of the lymph nodes is identified in the area painted with purple ink. This lymph node is bisected and submitted in (Block 2A). Another lymph node is identified which appears to demonstrate a large amount of fatty infiltration. This lymph node is bisected and submitted in (Block 2B). The remaining lymph node is submitted in (Block 2C).
CSW/sj
Microscopic Description
Sections of the right breast lumpectomy specimen reveal residual high grade ductal carcinoma in situ in two (2) out of twenty (20) sections of breast tissue. The residual ductal carcinoma in situ is identified in the wall of the previous biopsy site. There is high grade cytologic atypia and comedo necrosis is seen. Microcalcifications are associated with the ductal carcinoma in situ. A chronic inflammatory response is also elicited. In one slide, the ductal carcinoma in situ is 1.5 mm in greatest extent. In the other slide, the ductal carcinoma in situ is approximately 3.8 mm in greatest extent. Invasive malignancy is not identified. The inked margins of excision are free of tumor. Ductal carcinoma in situ is identified approximately 6.5 mm from the closest (lateral) margin of excision.
Changes of the previous biopsy site are seen. There is dense fibrosis. Many hemosiderin and xanthrochrome laden macrophages are seen. There are suture granulomata and fat necrosis. Hematoma formation is present.
Also identified in the breast tissue are foci of fibroadenomatoid change.
Section of overlying skin is negative for malignancy.
Three (3) out of three (3) right sentinel lymph nodes are negative for metastatic mammary carcinoma. Fatty infiltration is present with the two largest lymph nodes demonstrating extensive fatty infiltration.
CSW/sj
Final Diagnosis
1. Lumpectomy specimen, right breast (frozen section control).
A. Residual high grade ductal carcinoma in-situ with comedo necrosis identified in wall of previous biopsy site (approximately 3.8 mm in greatest extent).
B. Microcalcifications present.
C. No invasive mammary carcinoma identified.
D. Changes of previous biopsy site including fibrosis, many hemosiderin and xanthrochrome-laden macrophages, suture granulomata, fat necrosis and hematoma formation.
E. Foci of fibroadenomatoid change.
F. Segment of overlying skin, negative for malignancy.
Comment: The inked margins of excision are negative for malignancy. Ductal carcinoma in situ is identified approximately 6.5 mm from the closest inked

Ready 9/43 headers and 13/43 values are found. C:\DEV\Synoptex\Demo 2008-09-22

Synoptex - Reference_June_02_2008

File Edit View Tools Help

Breast: General

Auto Detect | Process Multiple Specimens | 1 | 2

1001100000147618
1001100000147714
1001100000147950
1001100000147988
1001100000148170
1001100000148192
1001100000148299
1001100000148316
1001100000148439
1001100000148457
1001100000148458
1001100000148482
1001100000148588
1001100000148661
1001100000148662
1001100000148663
1001100000148718
1001100000148719
1001100000148798
1001100000149128
1001100000149317
1001100000149338
1001100000149344
1001100000149360
1001100000149371
1001100000149398
1001100000149399
1001100000149625
1001100000149629
1001100000149769
1001100000149788
1001100000149810
1001100000149879
1001100000149942
1001100000150009
1001100000150036
1001100000150088
1001100000150171
1001100000150265
1001100000150302
1001100000150374
1001100000150383
1001100000150397
1001100000150398
1001100000150432
1001100000150435
1001100000150441
1001100000150454
1001100000150557
1001100000150571
1001100000150583
1001100000150940

Synoptic Element	Data Value
Specimen Type	
Lymph Node Sampling **	Sentinel Lymph Nodes
Lymph Nodes Examined **	1
Lymph Nodes Positive **	2
Lymph Nodes Negative **	3
Extranodal Extension	
Specimen Size **	3.5x3.3x1 cm
Tumor Laterality	
Tumor Site	
Tumor Size	
Tumor Greatest dimension	
Histologic Type	Ductal carcinoma in situ
Grading System	
Histologic Grade	
Tubule Formation	
Nuclear Pleomorphism	
Mitotic Count	
Necrosis	
Pathologic Staging (pTNM)	pN0: No regional lymph node metastasis histolo...
	pTis: in situ
HER2 Status	
ER Status	
PR Status	
Margins	
Distance of DCIS from Margin	
Distance of Tumor from Deep ...	
Distance of Tumor from Lateral...	
Distance of Tumor from Superi...	
Distance of Tumor from Inferior...	
Distance of Tumor from Medial...	
Distance of Tumor from Anteri...	
Distance of Tumor from Posteri...	
Involved Margins by Invasive ...	
Uninvolved Margins by DCIS	
Involved Margins by DCIS	
Extent of Margin Involvement f...	
Extent of Margin Involvement f...	
Extent of Intraductal Component	
Extent of Intraductal Compone...	
Lymphatic Invasion	
Venous Invasion	
Perineural Invasion	
Microcalcifications	
Additional Pathologic Findings	

Clinical Diagnosis and History
Right breast cancer.
Specimen(s) Received:
1: Right Breast mass
2: Right breast sentinel node
Gross Description
Received fresh for frozen section diagnosis evaluation of margins and designated "right breast lumpectomy" is a 2.5 x 0.4 cm tan skin ellipse overlying a 6 x 2 cm pad of yellow fatty tissue excised to a depth of 4 cm. Tags are present on the specimen designating the margins. A firm area is felt in the medial aspect of the specimen. Sectioning through this area reveals a cavity feel with a hematoma which is 3 cm in greatest dimension. A section from this area is frozen on a single block. The remaining frozen tissue is submitted in (Block 1A). The margins of the specimen are painted with blue ink. Additional sections from the medial margin are submitted in (Blocks 1B-1D). Sections from the lateral margins are submitted in (Blocks 1E-1G). Sections from the cranial margins are submitted in (Blocks 1H-1J). Sections from the caudal margins are submitted in (Blocks 1K-1L). Sections from the deep margin are submitted in (Blocks 1M-1P). Sections from the wall of the previous biopsy are submitted in (Blocks 1Q-1S). A section of overlying skin is submitted in (Block 1T).
Submitted in formalin and designated "right breast sentinel node (tissue with purple ink not reactive but in same area as sentinel node)" is yellow fatty tissue measuring 3.5 x 3.3 x 1 cm in aggregate. There appear to be three lymph nodes in this fatty tissue. One of the lymph nodes is identified in the area painted with purple ink. This lymph node is bisected and submitted in (Block 2A). Another lymph node is identified which appears to demonstrate a large amount of fatty infiltration. This lymph node is bisected and submitted in (Block 2B). The remaining lymph node is submitted in (Block 2C).
CSW/sj
Microscopic Description
Sections of the right breast lumpectomy specimen reveal residual high grade ductal carcinoma in situ in two (2) out of twenty (20) sections of breast tissue. The residual ductal carcinoma in situ is identified in the wall of the previous biopsy site. There is high grade cytologic atypia and comedo necrosis is seen. Microcalcifications are associated with the ductal carcinoma in situ. A chronic inflammatory response is also elicited. In one slide, the ductal carcinoma in situ is 1.5 mm in greatest extent. In the other slide, the ductal carcinoma in situ is approximately 3.8 mm in greatest extent. Invasive malignancy is not identified. The inked margins of excision are free of tumor. Ductal carcinoma in situ is identified approximately 6.5 mm from the closest (lateral) margin of excision.
Changes of the previous biopsy site are seen. There is dense fibrosis. Many hemosiderin and xanthrochrome laden macrophages are seen. There are suture granulomata and fat necrosis. Hematoma formation is present.
Also identified in the breast tissue are foci of fibroadenomatoid change.
Section of overlying skin is negative for malignancy.
Three (3) out of three (3) right sentinel lymph nodes are negative for metastatic mammary carcinoma. Fatty infiltration is present with the two largest lymph nodes demonstrating extensive fatty infiltration.
CSW/sj
Final Diagnosis
1. Lumpectomy specimen, right breast (frozen section control).
A. Residual high grade ductal carcinoma in-situ with comedo necrosis identified in wall of previous biopsy site (approximately 3.8 mm in greatest extent).
B. Microcalcifications present.
C. No invasive mammary carcinoma identified.
D. Changes of previous biopsy site including fibrosis, many hemosiderin and xanthrochrome-laden macrophages, suture granulomata, fat necrosis and hematoma formation.
E. Foci of fibroadenomatoid change.
F. Segment of overlying skin, negative for malignancy.
Comment: The inked margins of excision are negative for malignancy. Ductal carcinoma in situ is identified approximately 6.5 mm from the closest inked

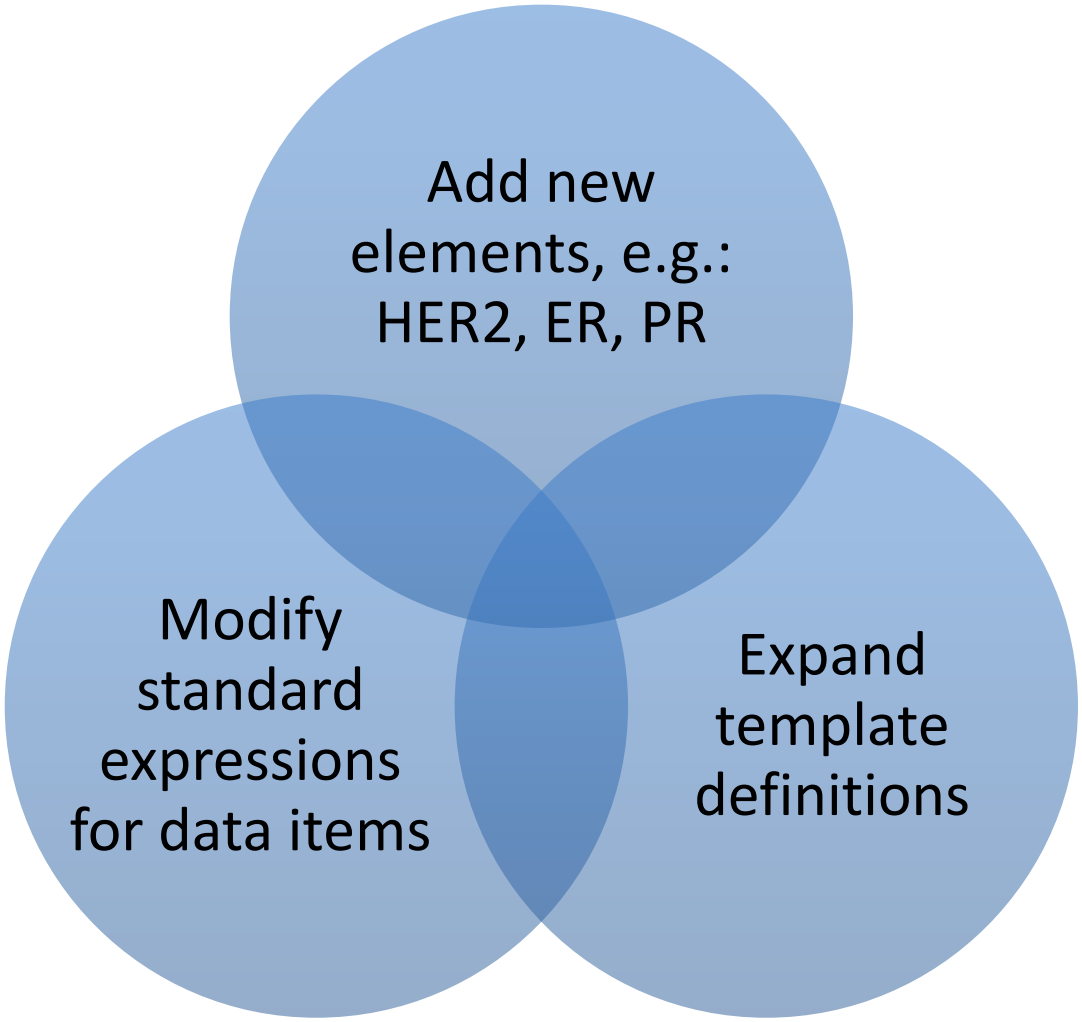
Ready 5/43 headers and 7/43 values are found. C:\DEV\Synoptex\Demo 2008-09-22

Agenda

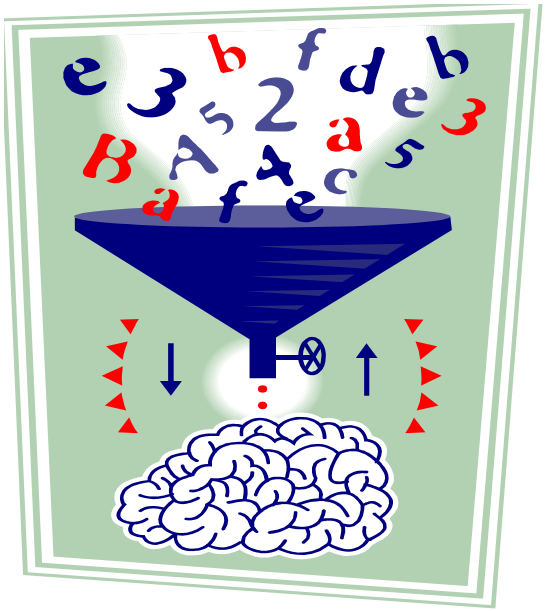
- Brief NLP Discussion
- AutoCode Enhancements
- Synoptex Enhancements
- Knowledge Base Manager
- CAP Checklist Updates
- Clinical Trials Matching
- Some Weird Stuff

The Challenge:

User wants to expand capabilities of Synoptex system



Knowledge base manager



Expand or modify

- Sites / Templates
- Concepts / data items
- Expressions
- Synonyms

User friendly view rather than technology oriented

Allows saving or importing synonyms for language customization

Synoptex Knowledgebase Manager - [Lexicon Editor]

File View Windows Help

Site/Procedure/Header

- Prostate Lexicon
- Lung Lexicon
  - Biopsy
  - Resection
    - Specimen Type
    - Tumor Laterality
    - Tumor Site
    - Tumor Size
    - Histologic Type
    - Histologic Grade
    - Pathologic Staging (pTNM)
    - Lymph Nodes Examined
    - Lymph Nodes Involved
    - Lymph Nodes Negative
    - Margin
    - Closest margin to invasive carcinoma
    - Distance of invasive carcinoma from closest margin
    - Direct Extension of Tumor
    - Venous Invasion
    - Arterial Invasion
    - Lymphatic Invasion
    - Additional Pathologic Findings
- MelanomaSkin Lexicon
- Thyroid Lexicon
- Kidney Lexicon
- UterineCervix Lexicon

Value

Name	Hierarchy
Cannot be determined	0

Header Properties

Numeric Extraction

Number Type: Size (Multi-Dimensional & 1D)

Specify range — Minimum: 0.0 — Maximum: 100.0

Override Default Section Orders

Priority: AD FD FR KPF GD MI SR — Edit

Ignore: — Reset

Specify Cut-Off Length

125

Expressions

Expression	Hierarchy	Enable
tumor [size]	0	☑

Synonyms

size	measure, diameter, size	Edit

Status

ARTIFICIAL INTELLIGENCE IN MEDICINE INC.

Agenda

- Brief NLP Discussion
- AutoCode Enhancements
- Synoptex Enhancements
- Knowledge Base Manager
- CAP Checklist Updates
- Clinical Trials Matching
- Some Weird Stuff

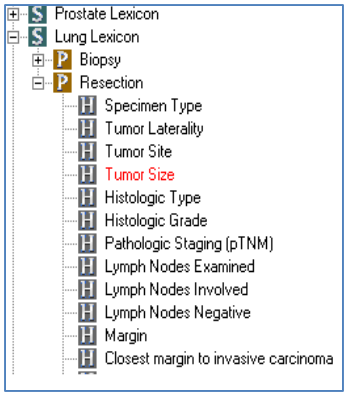
The challenge:

CAP Checklists version updates

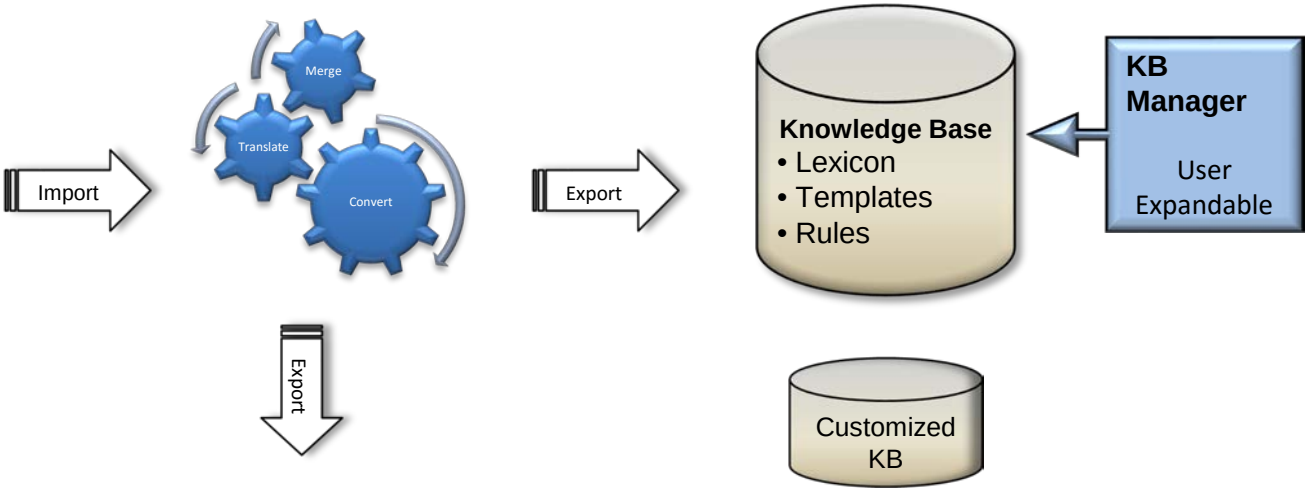
- How frequently are they updated?
- How do we manage updates/versions?
- How can clients upgrade their own systems?
- How can custom additions be kept across versions?

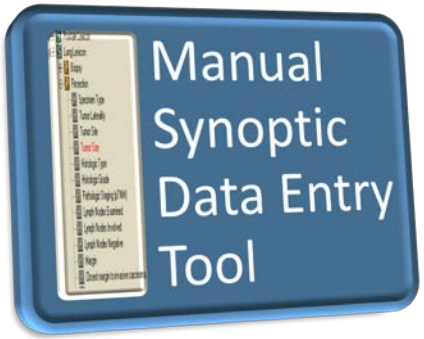
Automatic Updates

CAP Defined Checklists
XML



Synoptic Conversion Tool





Agenda

- Brief NLP Discussion
- AutoCode Enhancements
- Synoptex Enhancements
- Knowledge Base Manager
- CAP Checklist Updates
- Clinical Trials Matching
- Some Weird Stuff

Example trial criteria

Inclusion and exclusion

- Histologic Documentation: Patients must have a histologic (i.e., not just cytologic) diagnosis of invasive breast cancer by core biopsy. Excisional biopsy or incisional biopsy is not allowed. All breast cancer histologic types are allowed.
- **Histologic type : all breast histologic types are allowed**
- **Procedure : diagnosis by core biopsy**
- **No Excisional or incisional biopsy**

- Stage: Any patient with a clinical T1c (>1.5 cm) to T3 invasive breast cancer by the revised TNM staging system (AJCC 6th edition) will be eligible. Any N stage disease is allowed. No distant metastases allowed.
- **Stage: T1c to T3 any N stage is allowed using TNM staging system**

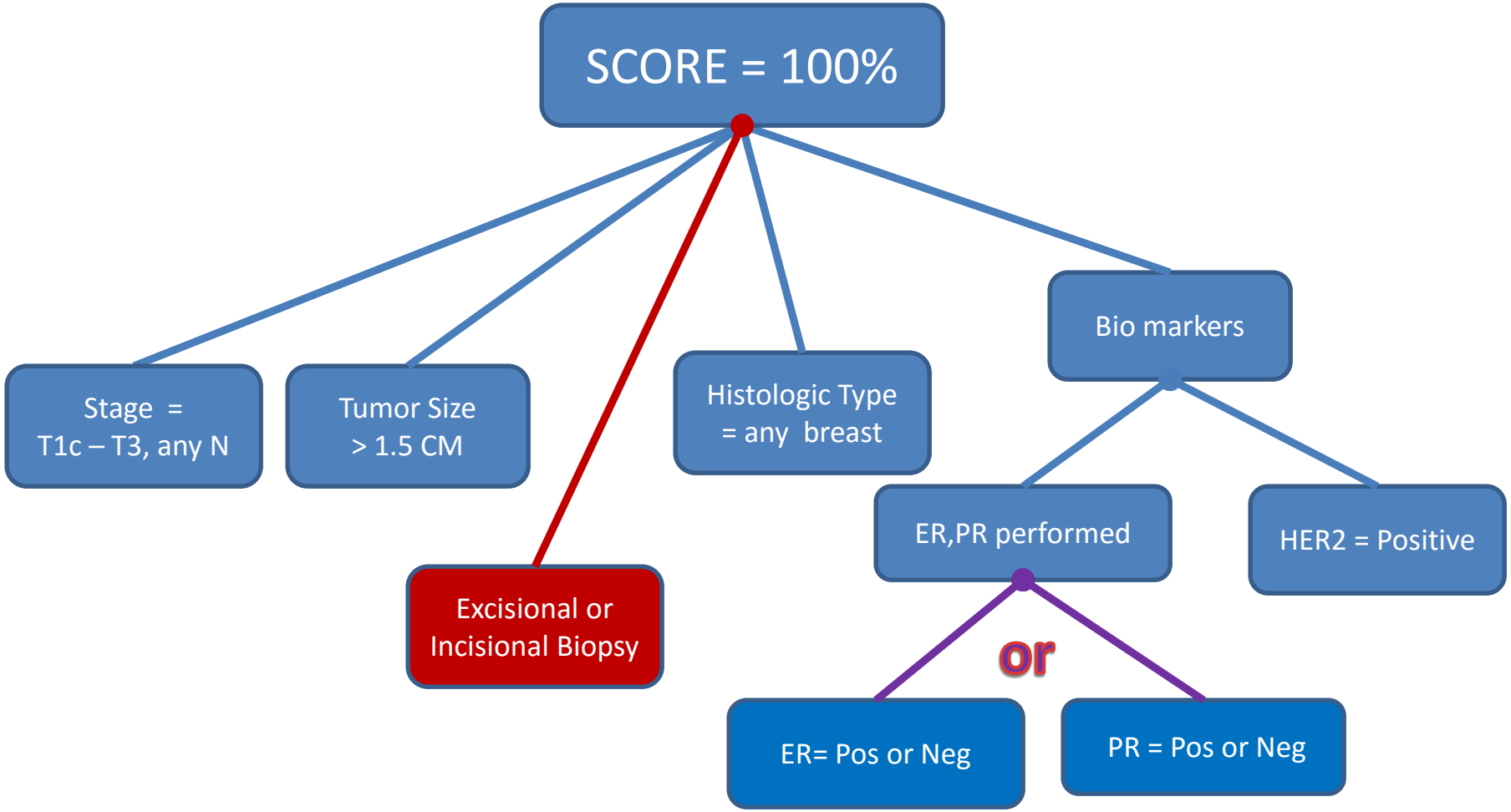
- Tumor Site: Patients must have invasive cancer in the breast. Multifocal disease (i.e. confined to a single quadrant in the same breast) is allowed. Multicentric disease (i.e. disease in multiple breast quadrants) is not allowed. Determination of multifocal and multicentric disease status will be made by the evaluating surgeon; ambiguous cases will be reviewed by the principal investigator. Patients with synchronous contralateral invasive breast cancers are noteligible; prior contralateral breast cancer allowed as long as patient has not received prior chemotherapy or radiation therapy in the past 5 years.
- **Tumor site: Single quadrant only**

- Measurable Disease: Patients must have measurable disease in the breast by imaging studies (mammogram, ultrasound, or MRI), and must be greater than 1.5 cm in at least one dimension by one or more of the imaging assessments.
- **Tumor Size: > 1.5 cm**

- Conventional Biomarker Status: Standard clinical biomarkers for ER, PR, and HER2 must be obtained on the initial diagnostic core biopsy. The invasive cancer must be HER2 negative (i.e. immunohistochemistry score 1-2+ and/or FISH non-amplified). Any ER/PR status is allowed. Patients who are HER2 2+ on initial immunohistochemistry assessment will be further assessed by FISH. In this instance, patient will be consented and further screened for eligibility and have tissue acquired for genomic profiling. If the standard of care additional FISH testing is positive for HER2 gene amplification, the patient will not be randomized and will be treated in the same manner as screen failures.
- **Bio Markers: EP, PR and HER2 must be obtained**
- **HER2 must be negative, any ER, PR is allowed**

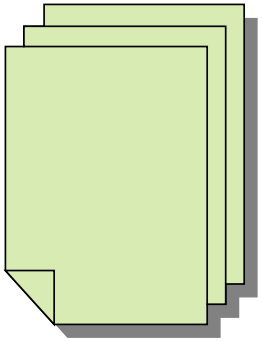
- **Age at least 18 years.**

Scoring Tree Example



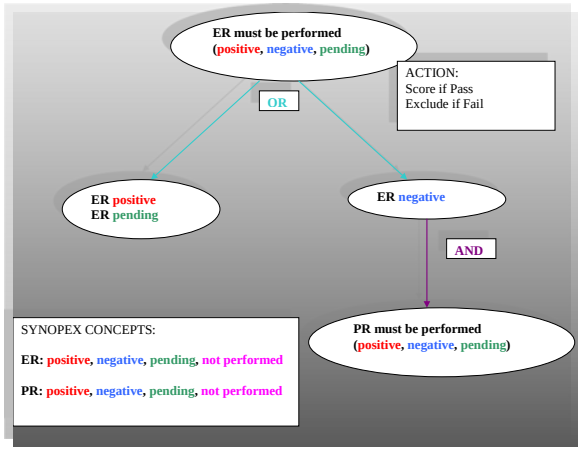
Matching Patients to Clinical Trials?

Protocol descriptions



Available patient data

Machine Readable Search criteria



Clinical Trials Manager

Project View

Search Nodes: Trial 1

Cancer Sites: Breast

Objects: Save Del Add

ER PR and HER2
ER_Status
Grade_and_Histologic_Type
HER2_Status
Histologic_Grade.1
Histologic_Grade.2
Histologic_Type
Histologic_Type_Excludes
Margin
Object3
PR_Status
Specimen_Size

Description
(1) Report: Specimen Size
GREATERTHANOREQUAL 5

TreeView

Names Desc

(-1) Report: Histologic Type
CONTAINS-ANYof these values:
Invasive Carcinoma (NOS)
Invasive Ductal Carcinoma
Invasive Duc
Invasive Duc
Invasive Lob

Compound No
CHILDREN
Histologic_Typ
Histologic_Gra

Compound No
CHILDREN
Specimen_Siz
Margin
ER PR and HE

ConditionEditor

Trial 1

Name: Histologic_Type Ok

lock

On Pass: Add score 10
If Unknown: No Action 0
On Fail: Exclude -1

(10) Report: Histologic Type
CONTAINS-ANYof these values:
Invasive Carcinoma (NOS)
Invasive Ductal Carcinoma
Invasive Ductal Carcinoma With Intraductal Componer
Invasive Ductal Carcinoma With Paget Disease
Invasive Lobular

Patient Data
Histologic Type
Histologic Grade
Histologic Type
Invasive Component Greatest Dimension
Laterality
Location of Involved DCIS Margin
Location of Involved Margin by Invasive Carc
Location of Uninvolved DCIS Margin
Lymph Node Sampling
Lymph Nodes Examined

Operator
contains-any
contains-any
contains-any-ex
contains-all
contains-all-ex
contains-nothing
contains-any-ex-if-any
contains-only-one-of

Values
Carcinoma With Spindle Cell Metaplasia
Carcinoma With Squamous Metaplasia
Cribriform
Ductal Carcinoma In Situ
Invasive Carcinoma (NOS)
Invasive Ductal Carcinoma
Invasive Ductal Carcinoma With Intraductal Component
Invasive Ductal Carcinoma With Paget Disease
Invasive Lobular
Lobular Carcinoma In Situ

Grade
se values:

Grade
se values:

PR Status
-ANYof these values:

ER Status
CONTAINS-ANYof these values:
Pending
Positive

(3) Report: HER2 Status
CONTAINS-ANYof these values:
Negative
Pending
Positive

Clinical Trials Manager

Project View

Search Nodes: Trial 1

Cancer Sites: Breast

Objects: Save Del Add

- ER PR and HER2
- ER_Status
- Grade_and_Histologic_Type
- HER2_Status
- Histologic_Grade.1
- Histologic_Grade.2
- Histologic_Type
- Histologic_Type_Excludes
- Margin
- Object3
- PR_Status
- Specimen_Size

Description

(1) Report: Specimen Size
GREATERTHANOREQUAL 5

TreeView

Names Desc

(-1) Report: Histologic Type
CONTAINS-ANYof these values:
Invasive Carcinoma (NOS)
Invasive Ductal Carcinoma
Invasive Ductal Carcinoma With Intraductal Component
Invasive Ductal Carcinoma With Paget Disease
Invasive Lobular

Compound Node Grade_and_Histologic_Type
CHILDREN
Histologic_Type
Histologic_Grade

(10) Report: Histologic Type
CONTAINS-ANYof these values:
Ductal Carcinoma In Situ
Lobular Carcinoma In Situ
Noninvasive Carcinoma (NOS)
Paget Disease Without Invasive Carcinoma

OR

(1) Report: Histologic Grade
CONTAINS-ANYof these values:
Grade I
Grade II

(1) Report: Histologic Grade
CONTAINS-ANYof these values:
Grade III

Compound Node Object3
CHILDREN
Specimen_Size
Margin
ER PR and HER2

(1) Report: Specimen Size
GREATERTHANOREQUAL 5

(1) Report: Margin
CONTAINS-ANYof these values:
Margin Involved By Invasive Carcinoma

Compound Node ER PR and HER2
CHILDREN
PR_Status
ER_Status
HER2_Status

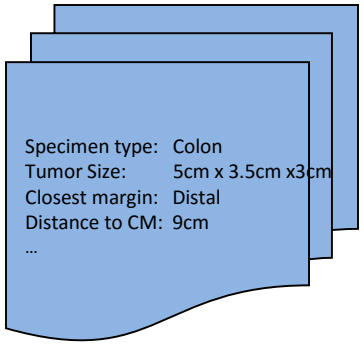
(5) Report: PR Status
CONTAINS-ANYof these values:
Pending
Positive

(3) Report: ER Status
CONTAINS-ANYof these values:
Pending
Positive

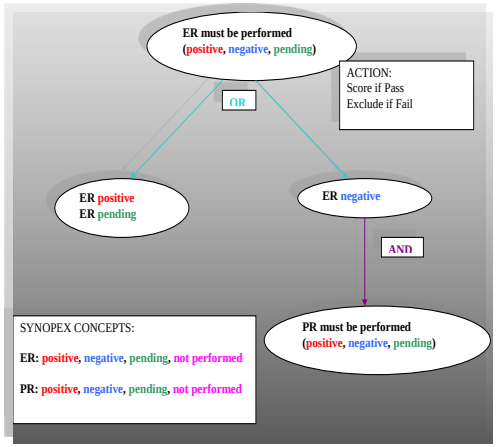
(3) Report: HER2 Status
CONTAINS-ANYof these values:
Negative
Pending
Positive

Matching Engine

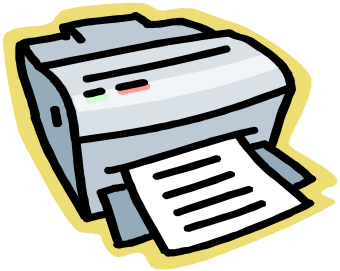
Patient data



Decision Builder



List of patient-trial matches with fitness scores



Results – patient trial matches

Clinical Trials Matching System

File Window

Results

Click on the matches to view details.

AIM ARTIFICIAL INTELLIGENCE IN MEDICINE INC

Matches | Excludes | Statistics

Score	Trial ID	Patient ID
04	B-39	213522420
04	B-39	211159827
04	B-39	172301116
04	B-39	196073520
04	B-39	205982096
04	B-39	200161896
04	B-39	188442972
04	B-39	182641263
04	B-39	171587769
03	B-39	175914605
03	B-39	176035370

Match Details

Score	Type	Description
1	Pass	Breast Cancer
1	Pass	Cancer Stage II, size <= 3
1	Pass	Sex is Female
1	Pass	Age > 17
	Unknown	Cancer Stage is either I or II
	Unknown	Cancer Stage II with size <= 3
	Fail	Exclude Men
	Fail	Positive Sentinel LN
	Fail	non-axillary/intramammary nodes can not be positive

Minimum Score (Threshold) 0

< Back

Combine KB Manager with Decision Builder

Can work on an existing database for Research purposes

Can be implemented in real time to scan and analyze incoming text reports

KB Manager

Define the language concepts and their values.

Discrete data from text

Decision Builder

Define the quantitative or qualitative criteria

Analysis of the data

Search reports for possible cases

Match Engine

Collection of cases

AIM ARTIFICIAL INTELLIGENCE IN MEDICINE INC.

Agenda

- Brief NLP Discussion
- AutoCode Enhancements
- Synoptex Enhancements
- Knowledge Base Manager
- CAP Checklist Updates
- Clinical Trials Matching
- Some Weird Stuff

The Weird Stuff

Can a system identify the relevant concepts in a new domain automatically?

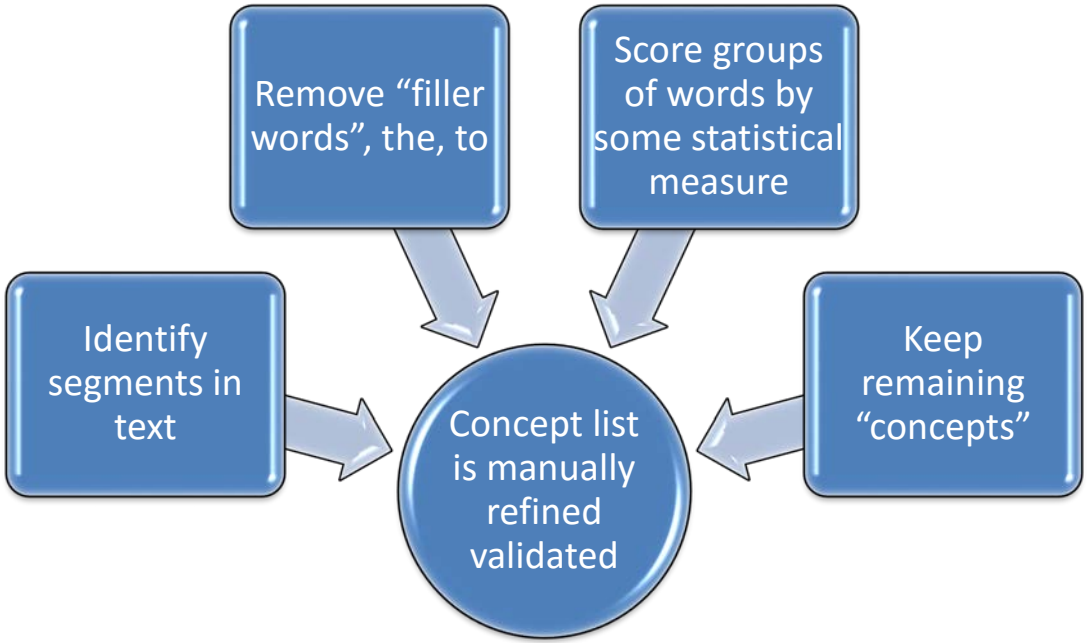
Useful where no prior classification system exists.

May reduce the need to review large numbers of documents or reports.

Would allow the user to browse list of concepts to pick out the relevant ones.

Automatic concept extractor

AutoCode in Reverse?



Concept Recognizer

File Location | Settings

Refresh | Stop | Report Processed: 9760

One File, Single/Multiple Reports

Text

D:\Radiology Reports\DesertRadiolgists.hl7 | HL7

Multiple Files, Single Reports

Text

Score	Count	Words
0.195445272808...	2492	exam ct
0.193624163916...	1647	mastoid air cells
0.193458722014...	783	findings low cuts show normal posterior fossa adequately delineated fourth ventricle
0.193271478736...	1644	no midline shift
0.186688022039...	1588	no abnormal enhancement
0.184935775795...	2358	paranasal sinuses
0.175555577297...	1204	no abnormal enhancement seen
0.169406817522...	2160	location rock
0.165056664322...	1404	no evidence abnormal
0.164586417415...	1400	sections brain obtained
0.163132490947...	2080	brain stem
0.162470306334...	1382	no evidence acute
0.160454156179...	11205	md
0.153956288331...	1963	location palomino
0.153328855674...	1955	restricted diffusion
0.149601347431...	1026	impression normal ct brain
0.149309726871...	1024	periventricular subcortical white matter
0.147211387263...	1877	posterior fossa
0.146599473226...	1247	deep white matter
0.142165023144...	975	mass effect midline shift

Thank You!

- Brief NLP Discussion
- AutoCode Enhancements
- Synoptex Enhancements
- Knowledge Base Manager
- CAP Checklist Updates
- Clinical Trials Matching
- Some Weird Stuff

BREAK

Radiology Reports

HMR 0013 - MRI BRAIN W/WO CONTRAST - Aug 1 2007

FINDINGS: MRI of the brain with contrast.

Clinical Indication: 82-year-old female with **history of meningioma**.

Technique: MRI of the brain was obtained using the following sequences:sagittal T1, axial T1, axial FLAIR, axial T2, axial diffusion, axial ADC, axial postcontrast T1, and coronal postcontrast T1 weighted sequences.

Comparison: comparison is made to an MRI of the brain from an outside institution (Downey Regional Med Ctr) dated 12/5/2006.

There are no extra-axial fluid collections. A right frontal extra-axial parasagittal mass demonstrating isointense T1 signal to gray matter and intermediate to high signal intensity on T2 images with marked homogeneous enhancement postcontrast is noted to measure 2.6 cm transverse x 3.4 cm AP x 3.5 cm craniocaudal. Marrow signal changes in the calvarium abutting the mass are suspicious for interosseous involvement. When compared to the prior study of 12/5/2006, the size and appearance of the mass demonstrates no significant change. There is a minimal to mild mass effect on the adjacent right frontal lobe without evidence of edema.

The ventricles and cortical sulci are prominent consistent with mild to moderate cerebral atrophy. There is no hydrocephalus. The supratentorial brain parenchyma demonstrates scattered foci of T2 and FLAIR hyperintensities noted throughout the periventricular,

The cerebellum demonstrates mild atrophy. The brainstem appears normal.

There are no parenchymal masses or midline shift.

The are no restricted diffusion abnormalities.

Mild right anterior ethmoid sinus mucosal disease is noted. A minimal amount of fluid is noted in the right mastoid air cells. The orbits, remaining paranasal sinuses, and calvarium are unremarkable.

IMPRESSION:

1. Right frontal parasagittal **meningioma**, relatively unchanged in size and appearance when compared to the prior outside study dated 12/5/2006, including marrow signal changes in the calvarium adjacent to the lesion suggestive of intraosseous involvement.
2. Nonspecific white matter disease as described above, likely reflecting chronic ischemic changes.
3. Mild right anterior ethmoid sinus mucosal disease. Minimal amount of fluid noted in the right mastoid air cells..

Radiology Reports

IMPRESSION: Right frontal parasagittal **meningioma**, relatively unchanged in size and appearance when compared to the prior outside study dated 12/5/2006, including marrow signal changes in the calvarium adjacent to the lesion suggestive of intraosseous involvement.

noted throughout the periventricular, subcortical, and deep white matter of both cerebral hemispheres, which are nonspecific, and likely represents chronic ischemic changes. The bilateral basal ganglia demonstrate T2 and FLAIR hypointensity.

Term	Select	Term	Select
hyperintensity FLAIR		white matter change	
hyperintensity T2-weighted		STIR hyperintensity	
T2 hyperintensity		leukomalacic changes	
abnormal density white matter	Yes	choroid plexus cyst	
mass effect	Maybe	demyelinating disease.	
abnormal enhancement	Maybe	brain tumor	Yes
high density lesions		heterogeneous enhancement	Yes
low density lesions	Yes	infarction / infarct	
abnormalities	Maybe	white matter signal abnormality	
abnormal enhancement		chiari malformation	
small vessel disease		chronic ischemic change	
extraaxial mass lesions	Yes	low attenuation changes	Maybe
increased signal intensity	Maybe	Edema / Oedema	
abnormal signal intensity	Yes	vasogenic Edema/Eodema	Yes
enhancement	Yes	BBB (Blood Brain Barrier)	Yes
pathologic mass	Yes	Infiltrative lesion	Yes
increased density	Yes	tumor	Maybe
retention cyst.			

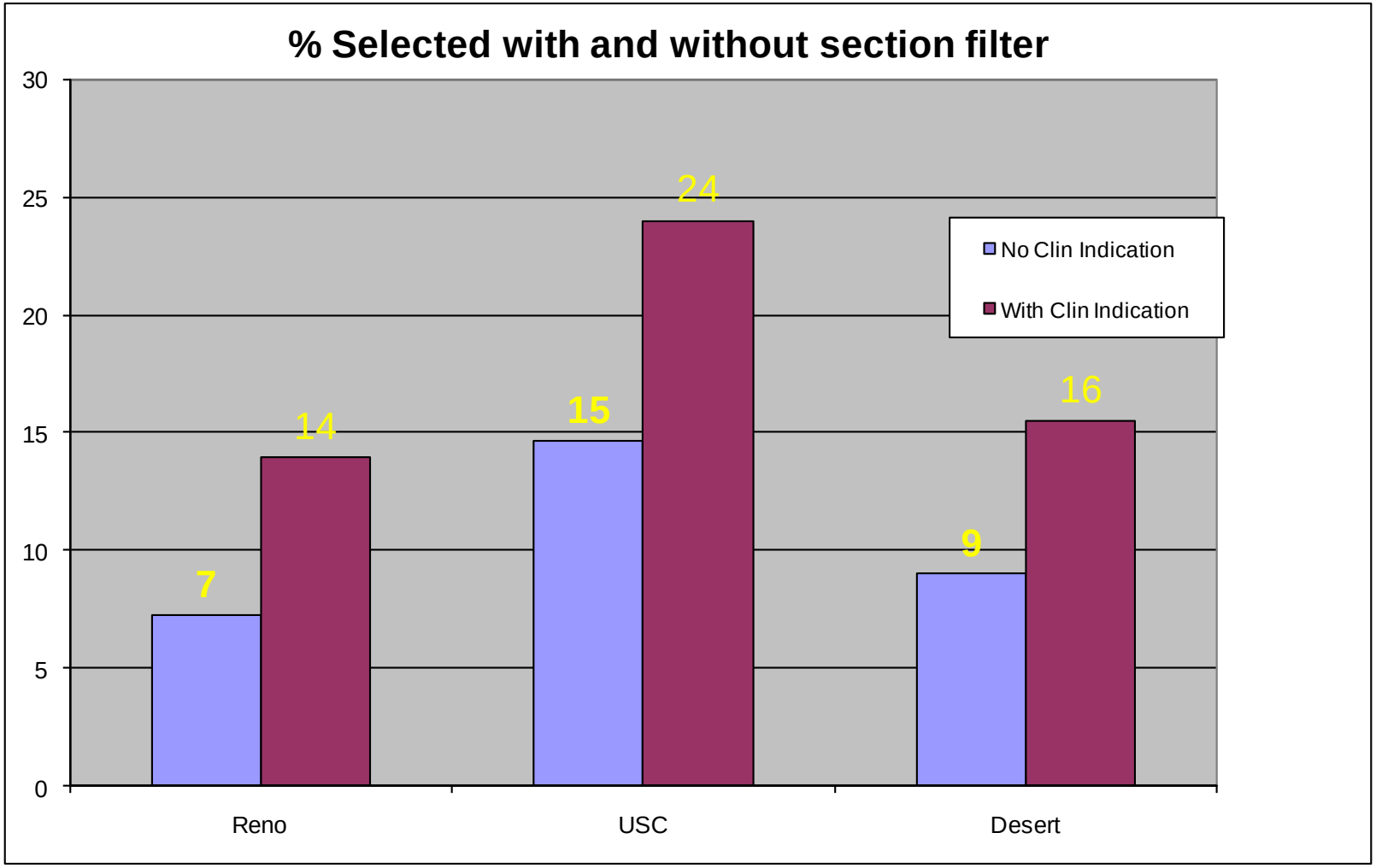
ARTIFICIAL INTELLIGENCE IN MEDICINE

Selection Rate > 30%

Term	Select	Term	Select
hyperintensity FLAIR		white matter change	
hyperintensity T2-weighted		STIR hyperintensity	
T2 hyperintensity		leukomalacic changes	
abnormal density white matter		choroid plexus cyst	
mass effect	Maybe	demyelinating disease.	
abnormal enhancement		brain tumor	Yes
high density lesions		heterogeneous enhancement	
low density lesions		infarction / infarct	
abnormalities		white matter signal abnormality	
abnormal enhancement		chiari malformation	
small vessel disease		chronic ischemic change	
extraaxial mass lesions	Yes	low attenuation changes	
increased signal intensity		Edema / Oedema	
abnormal signal intensity		vasogenci Edema/Eodema	
enhancement		BBB (Blood Brain Barrier)	
pathologic mass	Yes	Infiltrative lesion	Yes
increased density		tumor	Maybe
retention cyst.			

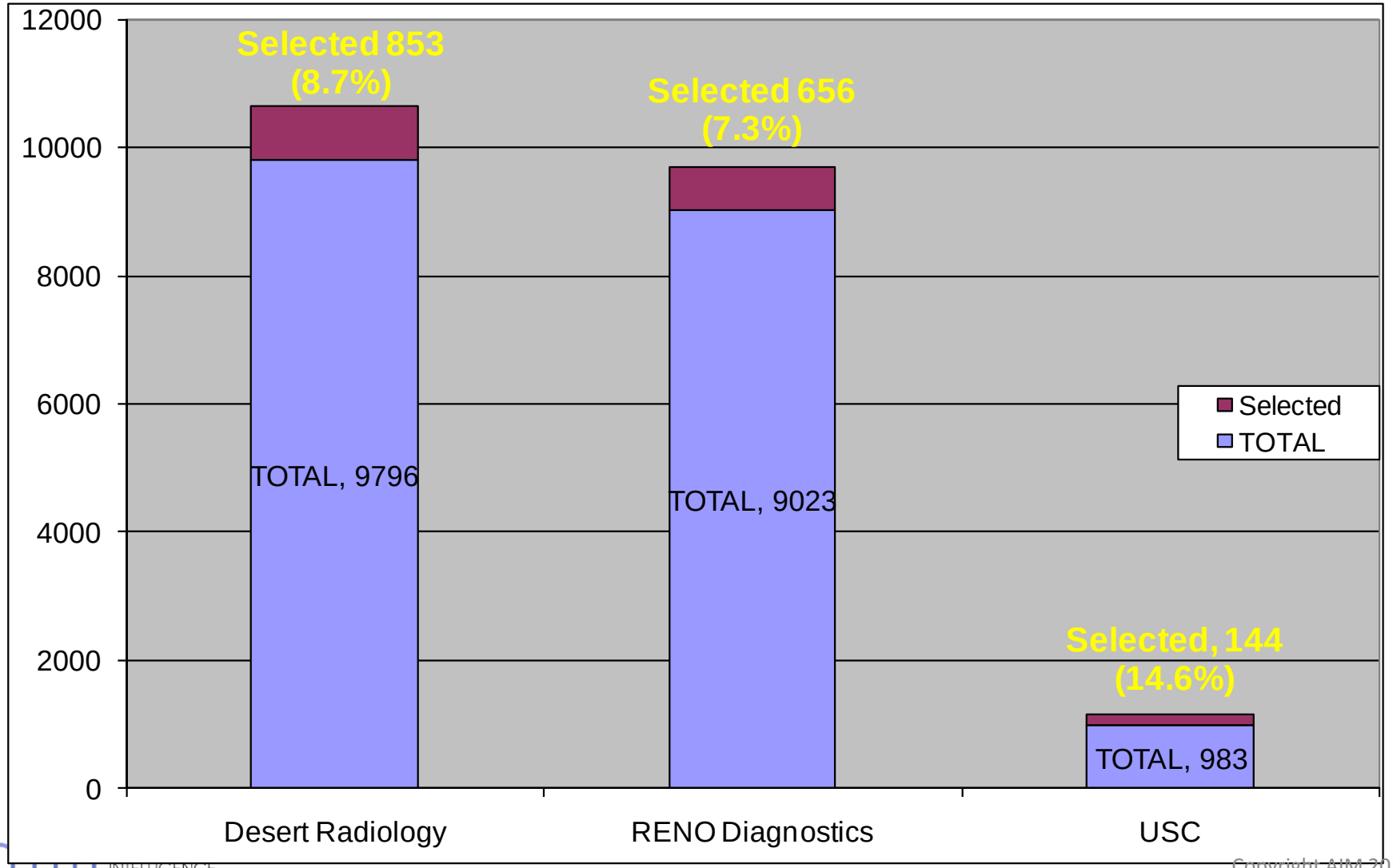
Selection Rate ~ 10%

Process Clinical Indication Section



ARTIFICIAL INTELLIGENCE IN MEDICINE INC.

Phase I Results



Phase II

- Installed into 3 registries and an initial run was done to determine performance and receive feedback.

Different requirements

- One registry was very sensitive to false positives, and in fact wanted only tumors that were new to the registry.

- Another registry wanted all possible cancers for follow-up with existing registry data.

Two different case-findings ?

1. Identify reports of the diagnosis, differential diagnosis, metastasis or history of primary CNS neoplasms or non CNS neoplasms of behavior greater than 2.

2. Identification of a diagnosis or differential diagnosis of a primary CNS neoplasm.

This was eventually extended to 4 categories

Sensitivity and Specificity Results for QC Study 1289 Reports

Bundle	True Positive	True Negative	False Positive	False Negative	Sensitivity	Specificity
B0096	4	90	2	0	100%	97.8%
B0097	16	81	3	0	100%	96.4%
B0098	13	78	4	0	100%	95.1%
B0103	5	93	1	1	83.3%	98.9%
B0108	15	83	2	0	100%	97.6%
B0110	10	88	2	0	100%	97.8%
B0111	1	92	6	0	100%	93.9%
B0112	10	87	2	0	100%	97.8%
B0113	7	90	3	0	100%	96.8%
B0115	15	78	7	0	100%	91.8%
B0116	13	86	1	0	100%	98.9%
B0117	10	86	4	0	100%	95.6%
B0118	5	91	3	1	83.3%	96.8%
Total	124	1123	40	2	98.4%	96.6%

Extended Classifications

0. Negative
1. History of cancer
2. Metastatic tumor
3. Positive previously known
4. Positive

- This information would allow increased decision making as well as provide information to registries.

Definitions

0. Negative – No mention of CNS tumor, no history of tumor, no metastases.

1. History of Cancer – Mention of history of tumor, no current CNS tumor

2. Metastatic Lesions – Reference to the presence of metastases.

3. Positive but Previously Identified – A report of any primary CNS neoplasm but with wording to indicate that is was known to be present when the examination was ordered.

4. Positive – A report of any primary CNS neoplasm that does not contain reference to history of tumor or metastases.

Positive – Previously Known

Report Classifier - All Classifications.csv

File Edit Tools

Aut CF		Code	Description
3	-	AIM-930-...	recurrence
4	+	C09.9	Tonsillar
2	+	C41.2	Spine
1	+	C53.9	Cervical
	+	C71.0	Cerebral [JC - RAD]
	+	C71.3	Parietal lobe
	+	C71.5	Lateral ventricle, NOS
	+	C71.5	Ventricle, NOS
	+	C71.7	Fourth ventricle, NOS
	+	C71.8	Corpus callosum
	+	C71.9	Brain, NOS
	+	C76.0	Head
	+	COVER	MRI brain [AJ - RAD]
	+	HX	History
	-	M-80001	Mass, NOS [JC - RAD]
	+	M-80001	Tumor NOS
	+	M-94703	Medulloblastoma, NOS (...

Case Finding
- Positive
- Positive - Previously Known
- History
- Metastatic
- Negative
- Flag report

Comment
Manual: Positive (Previous

[GROSS PATHOLOGY]

Note is made of a right parietal approach ventricular shunt catheter with the catheter passing through the right lateral ventricle from a posterior to anterior direction, the tip terminating in the region of the anterior aspect of the corpus callosum. There has been interval ventricular dilatation. The ventricles were mildly dilated on the prior study and are currently of normal caliber with the exception of postsurgical widening of the fourth ventricle.
Within the supratentorial space, there has been the interval decrease in degree of edema within the subcortical white matter in the right parietal lobe in the inferior parietal lobule at the site of ventricular shunt placement. There is no evidence of acute cerebral infarction. In this patient with right-sided weakness, there is no evidence of left-sided white matter edema or cortical infarction in the supratentorial space. There is no midline shift.
Within the infratentorial space, note is made of postsurgical change following suboccipital craniectomy for resection of a **medulloblastoma**. They are secondary postoperative widening of the fourth ventricle. There is no evidence of a brainstem infarction. There is cerebellar tonsillar ectopia with secondary narrowing of the CSF spaces of the foramen magnum.

[CLINICAL HISTORY]

UNSPEC SURGERY FOLLOW
ADMIT DATE/TIME:
Brain tumor removed in April 2
MRI of brain performed May 7, 2010. CT head performed April 27, 2010.

Resection medulloblastoma
Post surgical change

[FINAL DIAGNOSIS]

1. Interval resolution of pneumocephalus following surgery with interval ventricular decompression following placement of a ventricular shunt catheter. The ventricles are now of normal caliber.
2. Postoperative change following resection of a **medulloblastoma** at the level the fourth ventricle.
3. Soft tissue at the foramen magnum compatible with moderate cerebellar tonsillar ectopia. Given the patient's new onset right-sided weakness and absence of CT evidence of infarction, recurrent or residual mass or progressive hemorrhage in the brain, further evaluation with an MRI of the cervical spine to evaluate for the degree of cerebellar tonsillar ectopia and for the possibility of cervical syrinx or compression of the cervicomedullary junction may be of benefit.
Dictated by: Robert J Kadner, M.D.
Images and Report reviewed and interpreted by: Robert J Kadner, M.D.
<PS><Electronically signed by: Robert J Kadner, M.D.>
05/25/2010 1252

Positive

Report Classifier - All Classifications.csv

File Edit Tools

Auto CF	Code	Description
3	+ C69.6	Orbit, NOS
4	+ C71.1	Left frontal (Heuristic,...
2	+ C71.1	Frontal lobe
1	+ C71.5	Ventricle, NOS
	+ C71.9	Craniotomy (Heuristic,...
	+ C71.9	Brain, NOS
	+ C72.3	Optic chiasm
	+ C75.1	Pituitary gland
	+ C75.1	Pituitary, NOS
	+ C76.0	Head
	+ COVER	MRI brain [AJ - RAD]
	+ HX	History
	- M-80001	Enhancing mass [AJ -...
	+ M-81400	Microadenoma [JC - ...
	- M-82720	Pituitary adenoma, N...

Case Finding
- (•) Positive
- ( ) Positive - Previously Known
- ( ) History ( ) Metastatic
- ( ) Negative [ ] Flag report

Manual: **Auto:** Positive

Comment

MR BRAIN WITHOUT AND WITH CONTRAST

TECHNIQUE: Coronal T2, sagittal T1, post gadolinium dynamic coronal T1, and post gadolinium sagittal T1 weighted images of the pituitary; coronal FLAIR, axial T2, and post gadolinium axial T1 weighted images of the head; 20 mL intravenous Omniscan given.

[GROSS PATHOLOGY]

FINDINGS: Pituitary gland is bulbous in appearance. It measures 10 x 14 x 9 mm in maximal AP, transverse, and craniocaudal dimensions. In the right side of the gland, there is a 4 mm round hypoenhancing lesion. It is not well visualized on the T1 and T2 weighted images. Infundibulum is midline. Optic chiasm is normal in appearance. There is no enhancing intracranial mass. A developmental venous anomaly is in the posterior left frontal lobe. No definite intracranial hemorrhage. Ventricles differentiation is intact. patent. Orbits are intact inferior left maxillary sin

Pituitary gland microadenoma and no other neoplasm

[CLINICAL HISTORY]

eval for pituitary adenoma.

HISTORY: Increased serum prolactin level. Female problems.

[FINAL DIAGNOSIS]

IMPRESSION:
Pituitary gland **microadenoma**.
Job: 486414

History of cancer

Report Classifier - All Classifications.csv

File Edit Tools

Auto CF	Code	Description
3	- AIM-930-...	mass effect
4	+ C30.1	Middle ear
2	+ C49.9	Vessel, NOS
1	+ C61.9	Prostate, NOS
	+ C71.0	Cerebral [JC - RAD]
	+ C71.5	Ventricle, NOS
	+ C71.9	Brain, NOS
	+ C76.0	Head
	+ C80.9	Fluid
	- COVER	Mass effect [JC - RAD]
	+ COVER	MRI brain [AJ - RAD]
	+ HX	History
	+ M-80003	Cancer

Case Finding
- Positive
- Positive - Previously Known
- History / Metastatic
- Negative / Flag report

Manual: **Auto:** History

Comment

[NATURE OF SPECIMEN]

TECHNIQUE- Axial T1, FLAIR, fat-suppressed T2, axial diffusion, sagittal T1 and postcontrast axial fat-suppressed T1 and coronal and sagittal T1-weighted images were obtained of the entire head.

[GROSS PATHOLOGY]

FINDINGS- There is some minimal T2 high signal in the periventricular white matter compatible with minimal small vessel disease. The remainder of the brain parenchyma is normal in signal intensity. The ventricles are normal in size. There is no localized mass effect. There is no midline shift. No extraaxial fluid collections are identified. No abnormal areas of enhanc
sinuses, mastoid air cells,
voids are demonstrated wi
sinuses.

History of prostate cancer
No other neoplasm

[CLINICAL HISTORY]

CLINICAL HISTORY- Vertigo for a year. Patient has a history of prostate **cancer**.

[FINAL DIAGNOSIS]

IMPRESSION- Normal MRI of the head. The etiology of the vertigo for a year is not established on this exam. Please correlate clinically.

Metastatic

Report Classifier - All Classifications.csv

File Edit Tools

Auto CF	Code	Description
3	+ C64.9	Renal, NOS
4	+ C71.0	Central white matter
2	+ C71.1	Frontal lobe
1	+ C71.1	Left frontal (Heuristic, JC ...
	+ C71.4	Occipital pole
	+ C71.9	Brain, NOS
	+ COVER	MRI brain [AJ - RAD]
	+ HX	History
	+ M-80001	Tumor NOS
	+ M-80001	Enhancing lesion (AJ - R...
	+ M-80003	Cancer
	- M-80006	Metastatic
	+ M-80006	Metastatic

Case Finding
- Positive
- Positive - Previously Known
- History (Metastatic)
- Negative [ ] Flag report

Manual: **Auto:** Metastatic

Comment

[GROSS PATHOLOGY]

FINDINGS- The diffusion sequence demonstrates an area of signal loss centrally within a metastatic deposit involving the white matter of the left frontal lobe centrally. This was present previously. A peripheral zone of vasogenic edema is noted which measures 6.2 cm in AP dimension. On the prior examination this measured 5.1 cm. A irregular and somewhat ring-**enhancing lesion** is present in the left frontal lobe. This measures 2.7 x 2.2 x 2.5 cm in craniocaudal, AP and transverse dimensions respectively. On the prior examination same **enhancing lesion** measured 1.8 x 1.8 x 1.9 cm in the same planes. The prior examination also demonstrated a 3 mm nodular lesion involving the left occipital pole with an area of vasogenic edema measuring approximately 2.2 x 1.0 cm in size. On the current e
size and the area of v
in size. Note is made
intensity on the T1 pr
edema. This may rep
hemorrhage. Further
CT examination of the

History of renal cancer with lesions on frontal lobe

[CLINICAL HISTORY]

CLINICAL HISTORY- 55-year-old with a history of renal cell cancer, metastases, follow up.

[FINAL DIAGNOSIS]

IMPRESSION- Enlargement of a left frontal lesion with associated increasing vasogenic edema. There is also enlargement of a left occipital metastatic deposit with increase in vasogenic edema as well. Areas of increased signal intensity on the precontrast sequence along the margins of the occipital edema may represent elements of petechial hemorrhage. The above information was called to and discussed with Dr. Kommor at the time of the dictation.

Report classification

	Classification	History of Tumor	Metastatic Tumor	CNS Tumor	Non CNS Tumor
0	Negative	-	-	-	-
1	History	yes	-	-	yes
2	Metastatic	implied	Yes	implied	implied
3	Positive previously known	yes	-	yes	-
4	Positive	-	-	yes	-

What about Sensitivity and Specificity?

	Classification	History of Tumor	Metastatic Tumor	CNS Tumor	Non CNS Tumor
0	Negative	-	-	-	-
1	History	yes	-	-	yes
2	Metastatic	implied	Yes	implied	implied
3	Positive previously known	yes	-	yes	-
4	Positive	-	-	yes	-

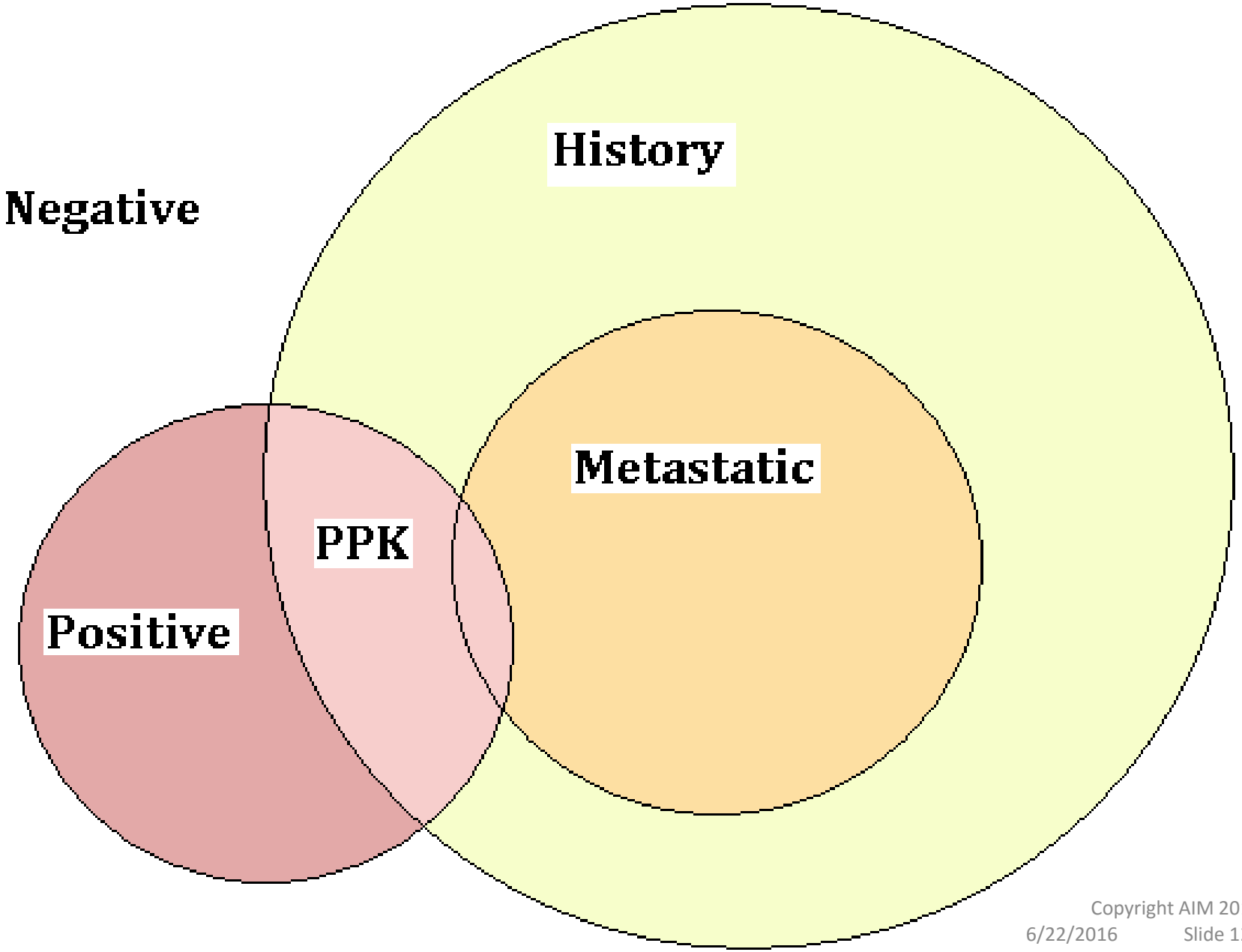
What about Sensitivity and Specificity?

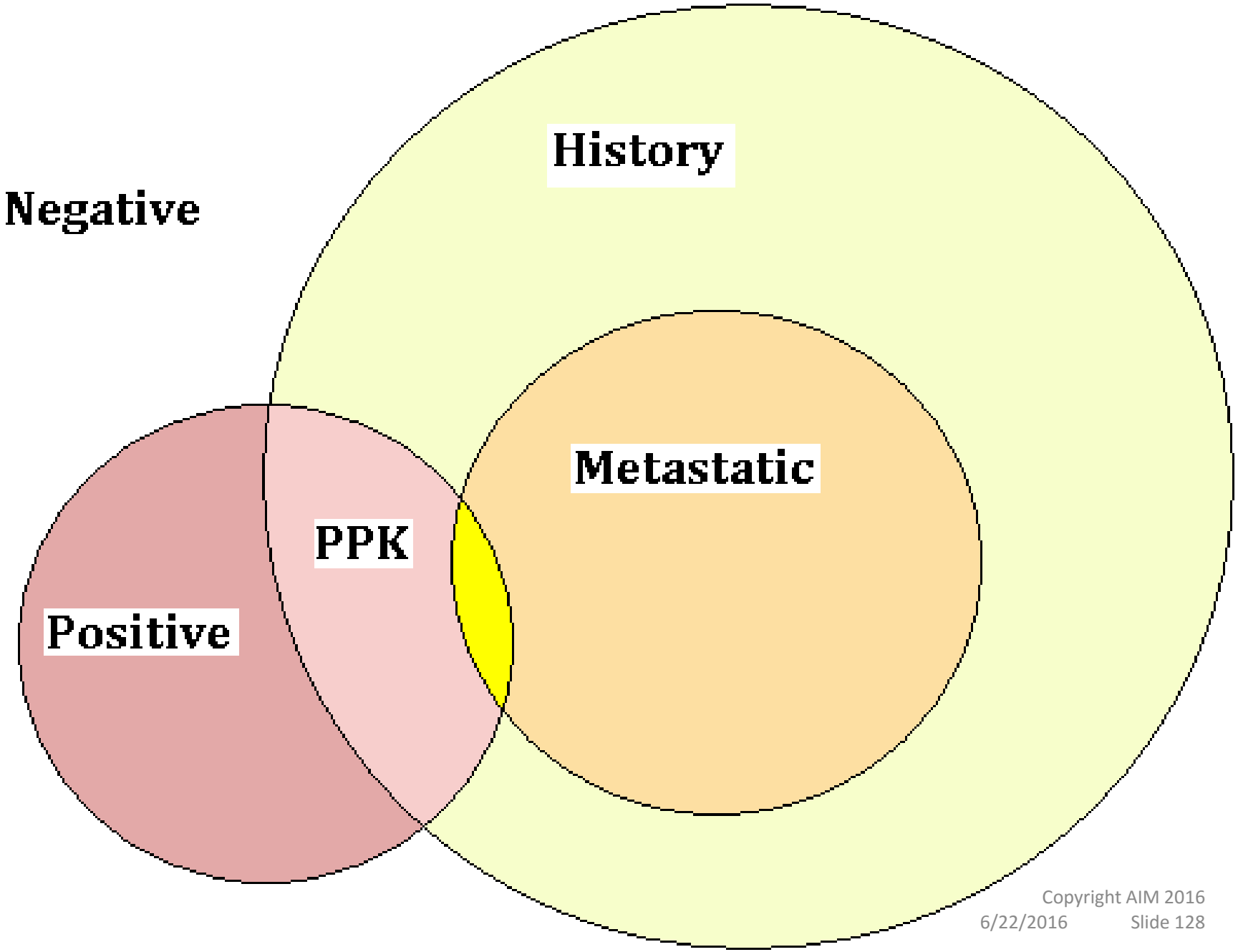
	Classification	History of Tumor	Metastatic Tumor	CNS Tumor	Non CNS Tumor
0	Negative	-	-	-	-
1	History	yes	-	-	yes
2	Metastatic	implied	Yes	implied	implied
3	Positive previously known	yes	-	yes	-
4	Positive	-	-	yes	-

What about Sensitivity and Specificity?

	Classification	History of Tumor	Metastatic Tumor	CNS Tumor	Non CNS Tumor
1	Negative	-	-	-	-
2	History	yes	-	-	yes
3	Metastatic	implied	Yes	implied	implied
4	Positive previously known	yes	-	yes	-
5	Positive	-	-	yes	-

AIM ARTIFICIAL INTELLIGENCE IN MEDICINE INC.





Results

Selection	Specificity	Sensitivity
History of Cancer	98	94
Metastatic cancer	99	89
Positive previously known	99	36
Positive	99	61
All the above	99	96

Results

Selection	Specificity	Sensitivity
History of Cancer	98	94
Metastatic cancer	99	89
Positive previously known	99	36
Positive	99	61
All the above	**99**	**96**

Results

Selection	Specificity	Sensitivity
History of Cancer	98	94
Metastatic cancer	99	89
Positive previously known	99	36
Positive	**99**	**61**
All the above	99	96

Results

Selection	Specificity	Sensitivity
History of Cancer	98	94
Metastatic cancer	99	89
Positive previously known	**99**	**36**
Positive	**99**	**61**
All the above	99	96

What is the registry load?

1422 reports tested – no tuning

True Positive	120
True Negative	1286
False Positive	11
False Negative	5

	Specificity	Sensitivity
All	0.99151889	0.96

	Matches	Manual	Auto
History	46	49	67
Metastatic	16	18	20
Positive(Known)	8	22	18
Positive	22	36	26
Total		125	

What does this mean?

- To reduce false positives – use class 4 only
  - Will miss maybe 25-30% but will get very few false negatives
- To get the best of both worlds – choose 3 and 4
  - Will get some false positive but no too many
- Institutions can choose any combination

Break

Various – Saved slides

Keyword search and concept building

Algorithm for finding concepts

1. Move focus to next sequential item in text (identify a word, segmenter, number, etc)
2. [option 1*: may do number search and marking first before concept search]
3. Calculate normal and cumulative co-ordinates based on segmentation rules and word position
4. Look-up the word in keyword dictionary (may apply suffix reduction here)
   1. if not a keyword or number go to step 1
   2. if number - do numeric processing (see option 1*)
5. With keyword, iterate through the list of active* concept trackers (CT) looking for ones that need this keyword
6. [Option 2: prune dead tracker]
   1. If not found, create a new active tracker and add the new keyword
   2. If found add the new keyword to the tracker – delete any keywords that are out of scope
   3. If tracker is full, mark it complete ie: all key words and in required order (complete trackers become inactive)
7. [Option3: Do sub-set concept check here]
8. Go to step 1 if not end of file
9. Perform negation check (may also do this on the fly in 4.c)

What happens if we find more than one expression for the same attribute?

Scoring – low level

- A heuristic scoring system will allow the expressions to be ranked and compared for overall strength.
- Each component will have an associated scoring method that is normalized to the interval [1, 0).
- **Total score = Score1 * Score2 * Score3…** new scoring components can be added as needed
- **Score scaling:** Not all score components have the same influence on the total score. For example, expression ordering is less significant that density or hierarchy scoring. A scaling factor can be applied to scoring component to reduce the impact of the particular score

Some simple heuristics

Expression density score

Density = Number of words in expression / Expression length

Order score = keywords out of order / total keywords.

Scaled order score = (keywords out of order + K)/ (total keywords + K)

Distance from question (if the reference is present)

Example of heuristic scoring

C. Level 9 lymph node biopsy:
- One lymph node, negative for metastatic malignancy (0/1).

Lymph Nodes Examined
* Number of lymph nodes examined: 30

Lymph Nodes Involved
* Number of lymph nodes involved: 0

Found Units of the

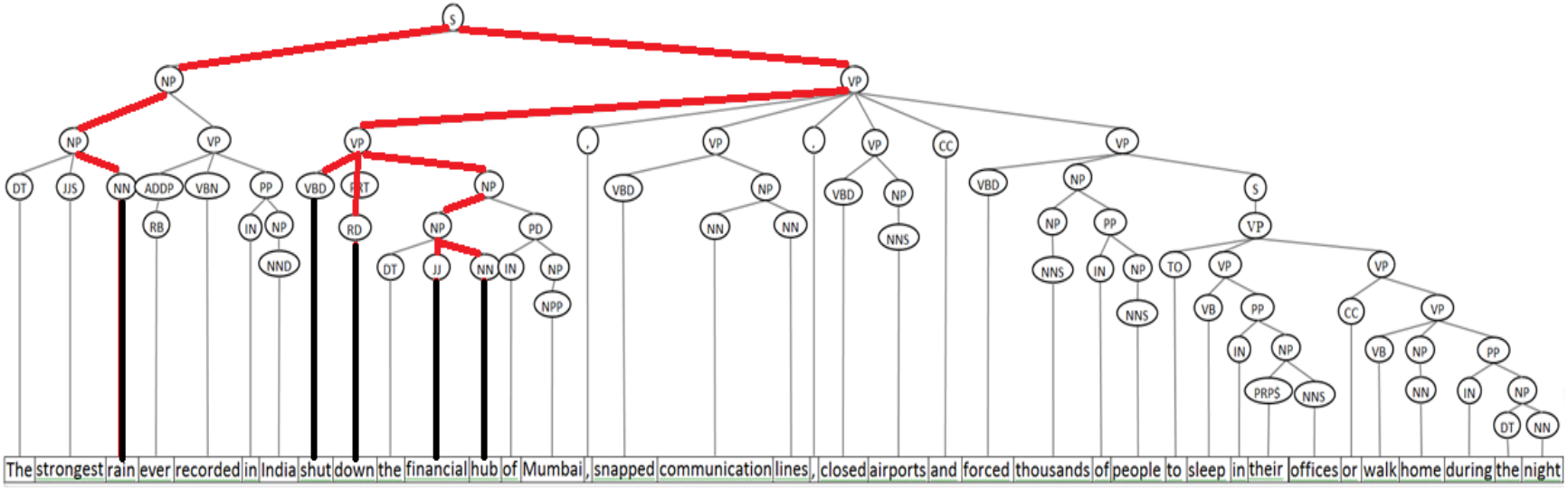
- Found Unit
  - Custom
  - ICD-O-
  - ICD-O-
  - Lymph
  - Lymph
  - Lymph
  - Specim

Processor's results | Processor's log

Question	Answer	External Coding	Value	Rank	Heuristic Vector
Specimen Type	Biopsy	-	[biopsy]	1	(ansPri:1.00,expDnsty:1.00,expPri:1.00) -> 1.0000
Lymph Nodes Positive	Lymph Nodes Positive (#)	-	1	0	(dist:0.90,queExpDnsty:0.75,queExpPri:1.00) -> 0.8833
Lymph Nodes Positive	Lymph Nodes Positive (#)	-	0	0	(dist:0.97,queExpDnsty:1.00,queExpPri:1.00) -> 0.9915
Lymph Nodes Negative	Lymph Nodes Negative (#)	-	1	0	(dist:0.98,queExpDnsty:0.85,queExpPri:1.00) -> 0.9444
Lymph Nodes Examined	Lymph Nodes Examined	-	30	0	(dist:0.98,queExpDnsty:1.00,queExpPri:1.00) -> 0.9917
Lymph Nodes Examined	Lymph Nodes Examined	-	30	0	(dist:0.95,queExpDnsty:0.95,queExpPri:1.00) -> 0.9667
Lymph Nodes Examined	Lymph Nodes Examined	-	0	0	(dist:0.95,queExpDnsty:0.95,queExpPri:1.00) -> 0.9662

More experimentation

- Combining the ideas of heuristic segmentation and grammar parsing. Build a constituent tree and use this to augment the heuristic "distance"



Study 1:
Patient and provider influences on disparities in colorectal cancer care

RCA Test Location:
Emory University
Rollins School of Public Health

The goal of this study is to understand how clinical and non-clinical factors impact disparities in chemotherapy use among colorectal cancer patients. The specific aims of the study are: 1) to examine the association between clinical, socioeconomic, and psychosocial patient factors and racial differences in use of chemotherapy; and 2) to examine the association between patient-provider relationship factors and use of chemotherapy.

University of Michigan
RSG CPHPS-121236
American Cancer Society

Study Requirements
Patients with invasive pathologic stage III colorectal cancer excluding appendix and lymphoma histologies.
Age > 20 years

Traditional Method of Case Identification
On-screen search and review of electronic pathology reports received at the registry.

Computer-Assisted Method
RCA automated screening of incoming pathology for colorectal cancer with one or more local lymph nodes positive. Followed by manual review of cases to verify absence of distant metastases.

Patient and provider influences on disparities in colorectal cancer care

RCA Test Location:
Emory University
Rollins School of Public Health

The traditional method was performed on two weeks of pathology reports received in January 2013

The computer-assisted method was performed on two weeks of pathology reports received in February 2013

Notes:
1. Manual pathology review did not require a separate step to identify reports with lymph nodes positive. Manual review assessed each case as a whole.

Impact of RCA

	Traditional Method	RCA assisted
Pathology Reports Reviewed	2,110	2,150
Colorectal Cancer Reports	203	200
Lymph Nodes Positive	n/a[1]	32
True Stage III Cases	9	11
Total Human Effort (min)	306 min	32 min

- Saved 4.5 hours of work in the test scenario
- = 30 minutes per candidate
- For 100 candidates = 50 hours of labor savings!

Study 2:
Comparative analysis of effectiveness of surgery and radiation for localized prostate cancer (CAESAR)

RCA Test Location:
Emory University
Rollins School of Public Health

A prospective observational cohort study to compare the effectiveness of surgery and radiation for localized prostate cancer, the most common male cancer. The study focuses on modern technologies and control for differences in patients and treatments that may affect both cancer-related and patient-reported outcomes (such as impotence and incontinence).

ClinicalTrials.gov ID:
NCT01326286
Vanderbilt University

Study Requirements
Men with a pathologic diagnosis of adenocarcinoma of the prostate with clinically localized stage.
Age 18-79

Traditional Method of Case Identification
Paper-based and/or database search and review of pathology reports

Computer-assisted Method
RCA scanning and filtering of electronic pathology reports received in real-time and containing a diagnosis of adenocarcinoma of the prostate

Study 2:
Comparative analysis of effectiveness of surgery and radiation for localized prostate cancer (CAESAR)

RCA Test Location:

Emory University

Rollins School of Public Health

A prospective observational cohort study to compare the effectiveness of surgery and radiation for localized prostate cancer, the most common male cancer. The study focuses on modern technologies and control for differences in patients and treatments that may affect both cancer-related and patient-reported outcomes (such as impotence and incontinence).

Impact of RCA

- Needed to accrue 500 subjects by 2013
- Had accrued 260 subject by November 2011
- RCA assisted accrual rate of approximately 70 subjects per month allowed the study to be filled by March 2012. Almost 1 year ahead of schedule!

"It allows us to identify newly diagnosed prostate cancer patients very quickly—within days of diagnosis"

"If we were to retrieve all pathology reports manually and fulfill the requirements we would need to visit each facility on a weekly basis. That would require a lot more personnel time plus travel and expenses"

Michael Goodman, MD, MPH
Department of Epidemiology

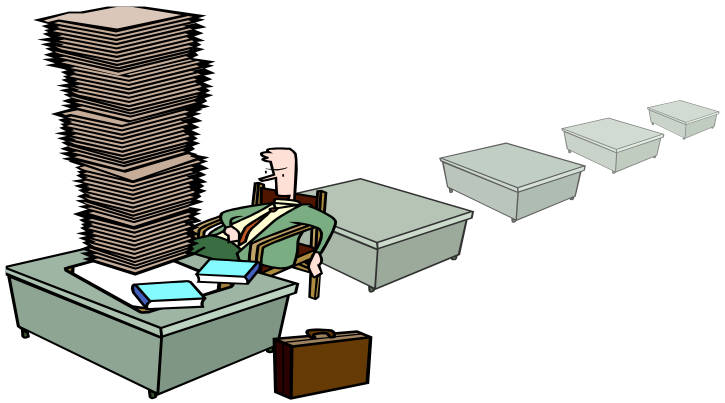
AIM ARTIFICIAL INTELLIGENCE IN MEDICINE INC.

ClinicalTrials.gov ID:

Retrospective study of HER2 testing patterns in breast cancer

Picture this scenario

- A key project to collect HER2 reporting data for analysis.
- Data exist in the text of 27,000 breast cancer pathology reports.
- Budget to hire 6 students.
- Need data in a standardized, machine readable form.
- Have a 3 month deadline!



Manual data extraction:

10 to 12 minutes per report by a trained reviewer

10 mins * 27,000 = 270,000 mins = 562 days

McMaster HER2 study

- 27,051 narrative surgical pathology reports for all incident breast cancer patients diagnosed in 2006 or 2007 in Ontario, Canada.

Data required	
tumour size	HER2 test provision
nodal status	type of test (IHC or FISH)
metastasis	HER2 status (negative, equivocal, positive)
histologic tumour grade, histologic tumour type,	HER2 overexpression measurements (0, 1+, 2+, 3+, % staining, HER2:CEP17 gene ratio)

Constraint: 6 students with 3 month time limit to review reports, tabulate and verify data

Manual: 14 minutes per report :: Synoptex: 2 ½ minutes per report
A six fold improvement in productivity

Stage 2 – Knowledge base

Unit Library Manager - [KnowledgeBase Explorer | Opened from DB]

File Edit View Global Domain Manager Window

Name
- Distance of Tumor from Clos...
- Distance of Tumor from De...
- Distance of Tumor from Inf...
- Distance of Tumor from Lat...
- Distance of Tumor from Me...
- Distance of Tumor from Pos...
- Distance of Tumor from Sup...
- Distant Metastasis (pM)
- EGFR
- ER - Allred Score
- ER Status
- Examination type
- Extranodal Extension
- FISH Results
- Focality
- Gleason Grade - Primary Pat...
- Gleason Grade - Secondary ...
- Gleason Grade - Tertiary Pat...
- Gleason Score
- Grade of dysplasia
- Grading System
- HER2 % cells stained
  - % cells stained
- HER2 gene copy number
- HER2 Result
  - HER2 FISH equivocal
  - HER2 FISH negative
  - HER2 FISH positive
  - HER2 IHC equivocal
  - HER2 IHC negative
  - HER2 IHC positive
  - HER2 indeterminate
  - HER2 Negative
  - HER2 Positive
- HER2:CEP 17 ratio
- Histologic Grade
- Histologic Type
- HX
- ICD-O-3 Morphology AIM Inc.
- ICD-O-3 Topography AIM Inc.

Concept | Numeric Answer

Name: HER2 IHC positive

Internal Code: 0001

Negative Value: HER2 IHC negative

☑ Use Global Negators ... Show Links...

External Codings

Name	Value

Local Negators

En	NegativeExp	Direction	Strength	Ordered	Contiguous
☐				☐	☐

Concept Answer

Answer Priority: 0

Answer Excludes

Concept To Exclude
HER2 FISH equivocal
HER2 FISH negative
HER2 FISH positive
HER2 IHC equivocal
HER2 IHC negative
HER2 indeterminate

Answer Excluded By

Excluded By

Expressions

EN	Expression	Rank	Normalization	Ordered	Contiguous	Domain Links Summary
☑	[HER2] [IHC] positive	1	☑	☐	☐	
☑	[HER2] [IHC] amplification [identified/found]	1	☑	☐	☐	
☑	[HER2] [IHC] amplified	1	☑	☐	☐	
☑	[HER2] [IHC] overexpression [identified/found]	1	☑	☐	☐	
☑	[HER2] [IHC] overexpressed	1	☑	☐	☐	
☐			☐	☐	☐	

Active KB Location : Opened from DB | Login: vli

Stage 2 – Knowledge base

Unit Library Manager - [KnowledgeBase Explorer | Opened from DB]

File Edit View Global Domain Manager Window

Name

- Distance of Tumor from Clos...
- Distance of Tumor from De...
- Distance of Tumor from Inf...
- Distance of Tumor from Lat
- Distance of Tumor from Me
- Distance of Tumor from Po
- Distance of Tumor from Su
- Distant Metastasis (pM)
- EGFR
- ER - Allred Score
- ER Status
- Examination type
- Extranodal Extension
- FISH Results
- Focality
- Gleason Grade - Primary Pat
- Gleason Grade - Secondary
- Gleason Grade - Tertiary Pa
- Gleason Score
- Grade of dysplasia
- Grading System
- HER2 % cells stained
  - % cells stained
- HER2 gene copy number
- HER2 Result
  - HER2 FISH equivocal
  - HER2 FISH negative
  - HER2 FISH positive
  - HER2 IHC equivocal
  - HER2 IHC negative
  - HER2 IHC positive
  - HER2 indeterminate
  - HER2 Negative
  - HER2 Positive
- HER2:CEP 17 ratio
- Histologic Grade
- Histologic Type
- HX
- ICD-O-3 Morphology AIM Inc.

Concept | Numeric Answer

Name: HER2 IHC positive

Concept Answer

Answer Priority: 0

Define Expression

Description:

Rank: 1

Ordered Contiguous Enabled

Add Concept Add Synonym

Searchable Entity	Normalized View	Normalize	Further Info
[HER2]		☐	Syn:{her2,her 2,her2neu,erbb 2,erbb2,neu,tkr1,ngl,p185erbB2,mln 19,her2 neu,cerb...
[IHC]		☐	Syn:{IHC,immunohistochemistry,herceptest,hercep test}
amplification	amplification	☑	Denorm:{amplification,amplifications}
[identified/found]		☐	Syn:{identified,discovered,found,recovered,noted,observed,detected,seen,present,...
		☑	

Close

Expressions

EN	Expression	Rank	Normalization	Ordered	Contiguous	Domain Links Summary
☑	[HER2] [IHC] positive	1	☑	☐	☐	
☑	[HER2] [IHC] amplification [identified/found]	1	☑	☐	☐	
☑	[HER2] [IHC] amplified	1	☑	☐	☐	
☑	[HER2] [IHC] overexpression [identified/found]	1	☑	☐	☐	
☑	[HER2] [IHC] overexpressed	1	☑	☐	☐	
☐			☐	☐	☐	

Active KB Location : Opened from DB | Login: vli

Stage 2 – Knowledge base

Unit Library Manager - [KnowledgeBase Explorer | Opened from DB]

File Edit View Global Domain Manager Window

Name

- Distance of Tumor from Clos...
- Distance of Tumor from De...
- Distance of Tumor from Inf...
- Distance of Tumor from Lat...
- Distance of Tumor from Me...
- Distance of Tumor from Pos...
- Distance of Tumor from Sup...
- Distant Metastasis (pM)
- EGFR
- ER - Allred Score
- ER Status
- Examination type
- Extranodal Extension
- FISH Results
- Focality
- Gleason Grade - Primary Pat...
- Gleason Grade - Secondary ...
- Gleason Grade - Tertiary Pat...
- Gleason Score
- Grade of dysplasia
- Grading System
- HER2 % cells stained
  - % cells stained
- HER2 gene copy number
- HER2 Result
  - HER2 FISH equivocal
  - HER2 FISH negative
  - HER2 FISH positive
  - HER2 IHC equivocal
  - HER2 IHC negative
  - HER2 IHC positive
  - HER2 indeterminate
  - HER2 Negative
  - HER2 Positive
- HER2:CEP 17 ratio
- Histologic Grade
- Histologic Type
- HX
- ICD-O-3 Morphology AIM Inc.

Concept | Numeric Answer

Name: % cells stained

Show Links...

Internal Code: 0001

Number Type

- UnsignedInteger
- RomanNumeral
- UnsignedDecimal
- Length
- Volume
- Weight
- MultiDimensional
- Concentration
- Percentage

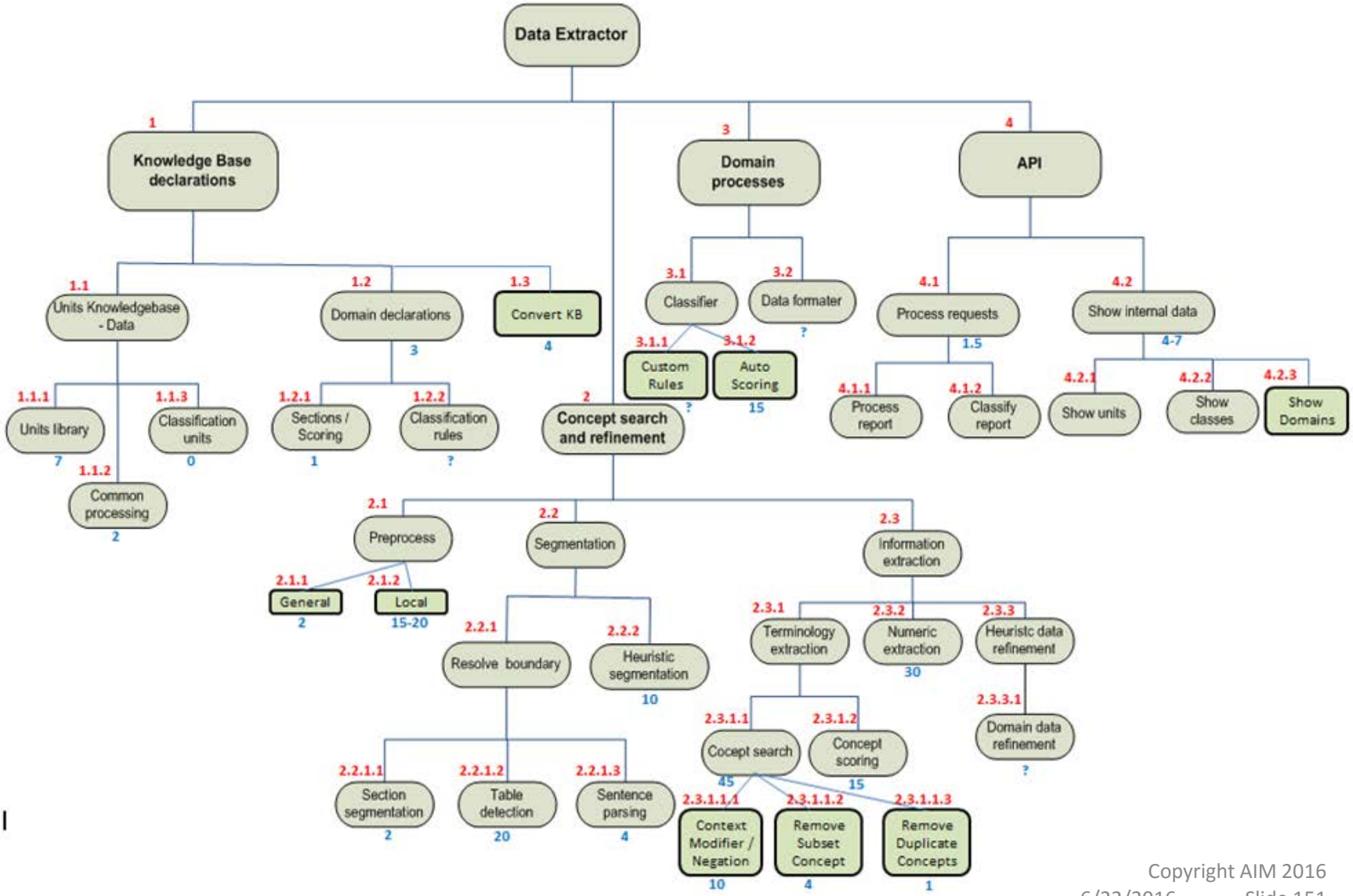
Specify Range

Minimum: 0.00

Maximum: 100.00

Active KB Location : Opened from DB | Login: vli

Units Extractor – Data and Functions



History of cancer

Report Classifier - All Classifications.csv

File Edit Tools

Auto CF	Code	Description
3	- AIM-930-...	mass effect
4	+ C30.1	Middle ear
2	+ C49.9	Vessel, NOS
1	+ C61.9	Prostate, NOS
	+ C71.0	Cerebral [JC - RAD]
	+ C71.5	Ventricle, NOS
	+ C71.9	Brain, NOS
	+ C76.0	Head
	+ C80.9	Fluid
	- COVER	Mass effect [JC - RAD]
	+ COVER	MRI brain [AJ - RAD]
	+ HX	History
	+ M-80003	Cancer

Case Finding
- Positive
- Positive - Previously Known
- History / Metastatic
- Negative / Flag report

Manual: **Auto:** History

Comment

[NATURE OF SPECIMEN]

TECHNIQUE- Axial T1, FLAIR, fat-suppressed T2, axial diffusion, sagittal T1 and postcontrast axial fat-suppressed T1 and coronal and sagittal T1-weighted images were obtained of the entire head.

[GROSS PATHOLOGY]

FINDINGS- There is some minimal T2 high signal in the periventricular white matter compatible with minimal small vessel disease. The remainder of the brain parenchyma is normal in signal intensity. The ventricles are normal in size. There is no localized mass effect. There is no midline shift. No extraaxial fluid collections are identified. No abnormal areas of enhanc
sinuses, mastoid air cells,
voids are demonstrated wi
sinuses.

History of prostate cancer
No other neoplasm

[CLINICAL HISTORY]

CLINICAL HISTORY- Vertigo for a year. Patient has a history of prostate **cancer**.

[FINAL DIAGNOSIS]

IMPRESSION- Normal MRI of the head. The etiology of the vertigo for a year is not established on this exam. Please correlate clinically.

Positive

Report Classifier - All Classifications.csv

File Edit Tools

Auto CF	Code	Description
3	+ C69.6	Orbit, NOS
4	+ C71.1	Left frontal (Heuristic,...
2	+ C71.1	Frontal lobe
1	+ C71.5	Ventricle, NOS
	+ C71.9	Craniotomy (Heuristic,...
	+ C71.9	Brain, NOS
	+ C72.3	Optic chiasm
	+ C75.1	Pituitary gland
	+ C75.1	Pituitary, NOS
	+ C76.0	Head
	+ COVER	MRI brain [AJ - RAD]
	+ HX	History
	- M-80001	Enhancing mass [AJ -...
	+ M-81400	Microadenoma [JC - ...
	- M-82720	Pituitary adenoma, N...

Case Finding
- (•) Positive
- ( ) Positive - Previously Known
- ( ) History ( ) Metastatic
- ( ) Negative [ ] Flag report

Manual: **Auto:** Positive

Comment

MR BRAIN WITHOUT AND WITH CONTRAST

TECHNIQUE: Coronal T2, sagittal T1, post gadolinium dynamic coronal T1, and post gadolinium sagittal T1 weighted images of the pituitary; coronal FLAIR, axial T2, and post gadolinium axial T1 weighted images of the head; 20 mL intravenous Omniscan given.

[GROSS PATHOLOGY]

FINDINGS: Pituitary gland is bulbous in appearance. It measures 10 x 14 x 9 mm in maximal AP, transverse, and craniocaudal dimensions. In the right side of the gland, there is a 4 mm round hypoenhancing lesion. It is not well visualized on the T1 and T2 weighted images. Infundibulum is midline. Optic chiasm is normal in appearance. There is no enhancing intracranial mass. A developmental venous anomaly is in the posterior left frontal lobe. No definite intracranial hemorrhage. Ventricles
differentiation is intact.
patent. Orbits are intact
inferior left maxillary sin

Pituitary gland microadenoma and no other neoplasm

[CLINICAL HISTORY]

eval for pituitary adenoma.

HISTORY: Increased serum prolactin level. Female problems.

[FINAL DIAGNOSIS]

IMPRESSION:
Pituitary gland **microadenoma**.
Job: 486414

Report classification

	Classification	History of Tumor	Metastatic Tumor	CNS Tumor	Non CNS Tumor
0	Negative	-	-	-	-
1	History	yes	-	-	yes
2	Metastatic	implied	yes	implied	implied
3	Positive previously known	yes	-	yes	-
4	Positive	-	-	yes	-

[COMMENT]
pt is being en route to floor .
@ 1946 RN, Allie doesnt know what size nor location for pt IV; will call as soon as she finds out
CT Thorax w/ IV Contrast .
********** FINAL ********** .
Transcribed By: JCM .
Transcribed Date/Time: 09-NOV-2015 22:46 .
Signed Date/Time: 09-NOV-2015 22:56.
[FINAL DIAGNOSIS]
.
FINDINGS: There are scattered **ground glass** opacities with superimposed interlobular septal thickening (crazy paving pattern). There are multiple small nodules with surrounding groundglass opacity, X few which appear cavitary. These are concerning for septic emboli. Tiny left pleural effusion with associated atelectasis. No evidence of loculation.
No evidence of hilar or mediastinal adenopathy. Coronary calcifications. .
IMPRESSION:
1. Multiple suspected septic emboli.
2. Scattered mixed groundglass and interstitial opacities bilaterally, which are nonspecific but may reflect edema or pneumonitis.
3. Tiny left pleural effusion without evidence of loculation.
BUN: 8, CREAT 0.64, Was Iodinated IV Contrast Used? (excluding gastroview) Yes, Contrast Given OMNI, Amount 70CC.
[CLINICAL HISTORY]
.
CLINICAL INDICATION: Pleural Effusion .
COMPARISON: Correlation with chest radiograph 11/8/2015. .
TECHNIQUE: X CT scan of the chest was performed after IV contrast administration.

Found Units of the Report
- Found Units
  - Additional Path
  - Custom Select
  - HX
  - ICD-O-3 Topog

Update Resu

Processor's results | Processor's log

	Question	Answer	External Coding	Value	Rank	Heuristic Vector	Domain Info	Po
	HX	HX	-	clinical history	1	(dist:1.00,ansPri:1.00,ansExpDnsty:...		+
	HX	HX	-	indication	1	(ansPri:1.00,expDnsty:1.00,expPri:1....		+
	ICD-O-3 Topography AIM Inc.	Thorax, NOS	C76.1	thorax	1	(ansPri:1.00,expDnsty:1.00,expPri:1....		+
	ICD-O-3 Topography AIM Inc.	Mediastinum, NOS	C38.3	mediastinum	1	(ansPri:1.00,expDnsty:1.00,expPri:1....		+
	ICD-O-3 Topography AIM Inc.	Thorax, NOS	C76.1	chest	1	(ansPri:1.00,expDnsty:1.00,expPri:1....		+
	ICD-O-3 Topography AIM Inc.	Thorax, NOS	C76.1	chest	1	(ansPri:1.00,expDnsty:1.00,expPri:1....		+
▶	Custom Selectable Terms	Groundglass Opacities (GGO)	AIM330	ground glass op...	1	(ansPri:1.00,expDnsty:1.00,expPri:1....		+
	Custom Selectable Terms	Pulmonary nodule	AIM340	nodule	1	(ansPri:1.00,expDnsty:1.00,expPri:1....		+
	Additional Pathologic Findings	Calcification	-	calcification	1	(ansPri:1.00,expDnsty:1.00,expPri:1....	Brain and Spinal Cord(14), General(...	+

Which approach works best ?

There is a trade-off in generality vs accuracy.

- NLP works great when sentence are well formed – falls down when information is not in sentences.
- Heuristic approaches are robust – but require knowledgebase, experimentation and much testing.
- Use a hybrid approach – pattern match with heuristics and incorporate formal NLP when optimal.

Question	Answer
Additional dimension...	1.0
Additional dimension...	1.3
HISTOLOGIC TYPE (chec...	Invasive ductal carcinoma
HISTOLOGIC GRADE...	
Type of Grading System	
Tubule formation	
Nuclear pleomorphism	
Mitotic count	
For a 25x objective ...	
For a 40x objective ...	
Nottingham System	
Nottingham Total Sc...	
Nottingham Grade	
Bloom Richardson ...	
Bloom Richardson S...	
Bloom Richardson G...	
Number of mitoses p...	
PATHOLOGIC STAGI...	
PRIMARY TUMOR (pT)	pT1mic: Microinvasion 0.1 cm or less in greatest dim...
REGIONAL LYMPH NO...	pN1a: Metastasis in 1-3 axillary lymph nodes (at least...
Number of regional l...	12
Number of regional l...	3
DISTANT METASTASI...	pMX: Cannot be assessed
MARGINS AND INVA...	
MARGINS (check all th...	
Status of margin inv...	Margins uninvolved by invasive carcinoma
Distance from clo...	0.09
Uninvolved margi...	
Status of margin inv...	
VENOUS/LYMPHATIC (L...	Absent

1. Left breast, lumpectomy: Diminutive aggregates of atypical cells consistent with residual invasive ductal carcinoma, grade 3, are identified (tissue section 1E). Immunohistochemistry reveals expression of pan-cytokeratin by these atypical cells, a finding that confirms their epithelial nature. The largest aggregate of atypical cells measures 0.7 mm in greatest dimension. These aggregates are embedded in a fibrotic tumor bed that measures 18 x 13 x 10 mm. The surgical margin is uninvolved by invasive carcinoma; the closest approximation of the residual foci of invasive carcinoma to the surgical margin (posterior) is 0.9 mm. The fibrotic tumor bed contacts the posterior surgical margin. No residual ductal carcinoma in situ is identified. Extensive fibrocystic change and sclerosing adenosis are observed elsewhere in the specimen. 2. Extra posterior margin, resection: No evidence of malignancy. Skeletal muscle is identified. 3. Extra medial margin, resection: No evidence of malignancy. 4. Extra superior margin, resection: No evidence of malignancy. 5. Extra inferior margin, resection: No evidence of malignancy. 6. Extra lateral margin, resection: No evidence of malignancy. 7. Extra anterior margin, resection: No evidence of malignancy. 8. Axillary dissection: Twelve lymph nodes are identified microscopically. Two are involved by tumoral metastasis. The largest continuous focus of tumoral metastasis measures approximately 4 mm in greatest dimension. There is no evidence of extra nodal extension. Surgical Pathology Cancer Case Summary Procedure: Excision without wire guided localization Lymph node sampling: Axillary dissection (partial or complete dissection) Specimen laterality: Left Tumor site: Not specified Histologic type: Invasive ductal carcinoma (no special type) Tumor size: Diminutive aggregates of invasive carcinoma, the largest measures 0.7 mm in greatest dimension Histologic grade: Grade 3. Given the small tumoral size, the Nottingham histologic score cannot be accurately established Tumor focality: (unifocal) few residual aggregates of invasive carcinoma are identified in a single tumoral bed Ductal carcinoma in situ: Not identified Lobular carcinoma in situ: Not identified Macroscopic and microscopic extent of tumor Skin: Skin is not present in the specimen Nipple: The nipple is not present in the specimen Skeletal muscle: The skeletal muscle is uninvolved Margins Margins are uninvolved by invasive carcinoma The distance of invasive carcinoma to the closest margin in the lumpectomy specimen is 0.9 mm (posterior margin); the additional posterior surgical margin obtained as specimen #2 measures 1 cm in width Lymph nodes Total number of lymph nodes examined (sentinel and nonsentinel): 12 Number of lymph nodes with macrometastasis: 2 Number of lymph nodes with micrometastasis: 0 Number of lymph nodes with isolated tumor cells: 0 Extranodal extension: Not identified Treatment effect In breast: Probable or definite response to presurgical therapy in invasive carcinoma In lymph nodes: Probable or definite response to presurgical therapy in metastatic carcinoma Lymphovascular invasion: Not identified Pathologic Staging Primary tumor: ypT1mi Regional lymph nodes: ypN1a Distant metastasis: Not applicable Additional pathologic findings: Fibrocystic change and sclerosing adenosis Ancillary studies: Performed in previous biopsy (see specimen AS13-8050) Microcalcifications: Not identified Clinical history: Not provided BR/sw

M69700	M80103	M85003	P1100	T04030	M49000	M74220	M74320
M80001	T00100	M09400	T13000	M09400	M09400	M09400	

1. In formalin, labeled with the patient's name and medical record number, and designated as "left breast", is a flat, roughly oval segment of tan-yellow fibroadipose tissue, partially fragmented, weighing 72 gm. A short single stitch marks the superior aspect, a long single stitch the lateral and a short double stitch the anterior. The specimen measures 6.5 cm from superior to inferior, 5.4 cm from medial to lateral, 2.8 cm from anterior to posterior. The surface is inked as follows: posterior black, anterior blue, superior green, inferior red, medial orange; lateral yellow. Sectioning reveals extensive areas of fibrosis spanning the entire specimen. A tan-

Results of recent E-Path QC studies – Sept 2015 ?

Registry	Hospital/Lab	Sensitivity	Specificity
Louisiana Tumor Registry	Our Lady of the Lake	0.989	0.991
Louisiana Tumor Registry	The Pathology Lab (TPL)	1.000	0.982
Louisiana Tumor Registry	Thibodaux	0.993	0.919
Louisiana Tumor Registry	HCSD	0.986	0.995
Louisiana Tumor Registry	Delta Pathology	0.993	0.989
Detroit	Oakwood Hospital	0.992	0.991
Iowa	UnityPoint Des Moines	1.000	0.993
Iowa	Mercy Iowa City	1.000	0.984
Iowa	United Clinical Lab	0.991	0.992
Iowa	Iowa Pathology Associates IPA Des Moines	0.988	0.994
Utah	Intermountain Medical Center	0.986	0.978
Utah	Huntsman Cancer Hospital/U of Utah	0.992	0.966
Utah	St. Mark's Hospital	1.000	0.974
Kentucky Tumor Registry	Norton	0.977	0.987
Kentucky Tumor Registry	Baptist	1.000	0.989
Kentucky Tumor Registry	Univ of Louisville	0.993	0.987
Hawaii Registry	Kuanani	1.000	1.000
Hawaii Registry	Clinical Labs of Hawaii	1.000	0.985
Hawaii Registry	Queens Medical labs	1.000	0.992
Hawaii Registry	Kaiser Moanala	1.000	0.929
New Mexico Cancer Registry	Presbyterian Healthcare System	1.000	0.986
New Mexico Cancer Registry	UNM Cancer Cente	0.996	0.984
New Jersey State Cancer Registry	Virtua	1.000	0.991
New Jersey State Cancer Registry	RWJ - Hamilton	1.000	0.992
New Jersey State Cancer Registry	St Peter's	1.000	0.971
	Average	0.995	0.982

Diagnostic imaging examinations - CNS ?

HMR 0013 - MRI BRAIN W/WO CONTRAST - Aug 1 2007

FINDINGS: MRI of the brain with contrast.

Clinical Indication: 82-year-old female with **history of meningioma**.

Technique: MRI of the brain was obtained using the following sequences:sagittal T1, axial T1, axial FLAIR, axial T2, axial diffusion, axial ADC, axial postcontrast T1, and coronal postcontrast T1 weighted sequences.

Comparison: comparison is made to an MRI of the brain from an outside institution (Downey Regional Med Ctr) dated 12/5/2006.

There are no extra-axial fluid collections. A right frontal extra-axial parasagittal mass demonstrating isointense T1 signal to gray matter and intermediate to high signal intensity on T2 images with marked homogeneous enhancement postcontrast is noted to measure 2.6 cm transverse x 3.4 cm AP x 3.5 cm craniocaudal. Marrow signal changes in the calvarium abutting the mass are suspicious for interosseous involvement. When compared to the prior study of 12/5/2006, the size and appearance of the mass demonstrates no significant change. There is a minimal to mild mass effect on the adjacent right frontal lobe without evidence of edema.

The ventricles and cortical sulci are prominent consistent with mild to moderate cerebral atrophy. There is no hydrocephalus. The supratentorial brain parenchyma demonstrates scattered foci of T2 and FLAIR hyperintensities noted throughout the periventricular, subcortical, and deep white matter of both cerebral hemispheres, which are nonspecific, and likely represents chronic ischemic changes. The bilateral basal ganglia demonstrate T2 and FLAIR hypointensity.

The cerebellum demonstrates mild atrophy. The brainstem appears normal.

There are no parenchymal masses or midline shift.

The are no restricted diffusion abnormalities.

Mild right anterior ethmoid sinus mucosal disease is noted. A minimal amount of fluid is noted in the right mastoid air cells. The orbits, remaining paranasal sinuses, and calvarium are unremarkable.

IMPRESSION:

1. Right frontal parasagittal **meningioma**, relatively unchanged in size and appearance when compared to the prior outside study dated 12/5/2006, including marrow signal changes in the calvarium adjacent to the lesion suggestive of intraosseous involvement.
2. Nonspecific white matter disease as described above, likely reflecting chronic ischemic changes.
3. Mild right anterior ethmoid sinus mucosal disease. Minimal amount of fluid noted in the right mastoid air cells..

NEW: Radiology scans of the chest

Classify same as in CNS – but also look for recurrence

	XRAY	CT	MRI	TOTAL
QC Report Sample	749	749	2	1500
Classification > 0	21	225	0	246
1 History of cancer	18	130	-	148
2 Metastases	1	62	-	63
3 Previously known	2	23	-	25
4 Positive (new)	-	9	-	9
5 - Recurrent	-	1	-	1

Classification	# OF POSITIVE REPORTS	SENSITIVITY	SPECIFICITY
1 History	148	0.973	0.997
2 Metastatic	63	0.940	1
3 Previously known	25	0.889	0.994
4 Positive	9	0.75	1
5 Recurrent	1	1	1
1+2+3+4+5	246	0.947	0.988
2+3+4	97	0.907	0.994
3+4	34	0.83	0.994

Practical AI – Development, Implementation, Support

- Focus on the practical – which problems to address?
- Need trained people – AIM trains in AI
- Need approaches to managing AI systems, esp with multiple knowledge or rule bases.
- Need approaches to the testing of systems in the field.
  - QC studies: manual check for each new (or operating) implementation
  - Case finding logic : thousands of reference examples
  - Testing AI engine and KnowledgeBase: (~70,000) samples to reference and compare

What about Sensitivity and Specificity?

	Classification	History of Tumor	Metastatic Tumor	CNS Tumor	Non CNS Tumor
0	Negative	-	-	-	-
1	History	yes	-	-	yes
2	Metastatic	implied	yes	implied	implied
3	Positive previously known	yes	-	yes	-
4	Positive	-	-	yes	-

Advances in AI today

- Mostly statistics based – machine learning
  - Neural models
  - Statistical rules, Bayesian, machine learning
  - Deep learning (new techniques for training multi-layer neural networks)
  - Gives “black box” model of AI – does not address underlying principles
  - Natural progression of the field – the “Physics” of AI dealt with simpler problems – knowledge representation in languages and logic techniques.
  - Problems tackled by ML are more complex - so we use statistics rather than fundamental models

BUT! Don’t throw out the power of heuristics quite yet!