Informatics in Radiation Oncology

Ramon Alfredo Carvalho Siochi, PhD
University of Iowa
Hospitals and Clinics

Informatics

- Information Technology
  - Infrastructure: hardware, applications, networks
  - Storage, security and flow of information in the form of data
- Information Science
  - Extracting information from data
  - Generate knowledge

Learning Objectives

1. Review the information technology and infrastructure required to deliver radiation therapy.
2. Understand the use of informatics in radiotherapy clinical trials.
3. Learn the basics of data mining and its applicability to radiation oncology.
4. Become familiar with the information technology that can be leveraged to improve clinical operations.

Overview

- IT infrastructure for RT
  - Data Flow
  - Database Basics
  - Turning Data into Information
- Clinical Trials
  - Protocols
  - National Databases
- Data Mining
- IT for safety and quality

Infrastructure ⇔ Process Map

- Radiation Oncology is Data Intensive
- Follow the Data
- Processes: What do we do with this data?
- Paper vs Paperless
  - Different processes
  - Different IT infrastructure

Data Flow in RO

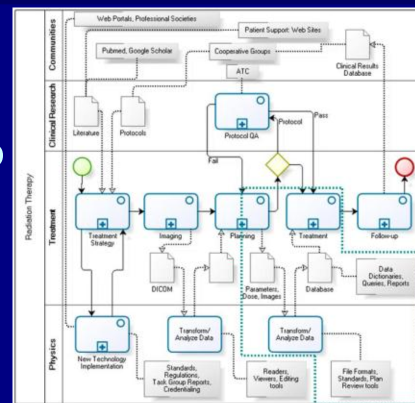
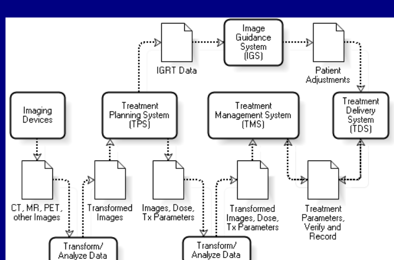


**Fig. 11.1 from Siochi, Information resources for radiation oncology, Ch. 11 of a forthcoming book: Informatics in Radiation Oncology, G. Starkschall, B. Curran, editors.*

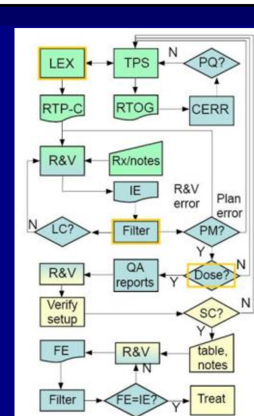
Distributed system data flow



Redundant data living in many places: INFORMATION should match.
(Data might be stored in different forms but mean the same thing.)

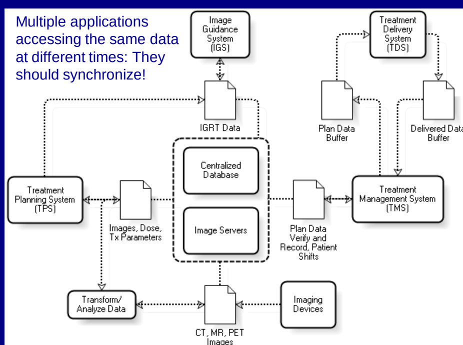
Clinical Interactions, paperless checks

Physicists
Dosimetrists/Physicians
Therapists
In-House Software



Adapted from Fig 5. Siochi, et al. Radiation therapy plan checks in a paperless clinic, J. App. Clin. Med. Phys., 10(1):43-62.

Centralized DB dataflow



Database basics

- DB consists of Tables
- Table: consists of rows (also called records)
- Row: contains column elements (also called fields)
- Queries
  - E.g. how many patients had IMRT this month?
  - SQL (Structured Query Language)

Database Table

Field or Column Names define the table

Key must be unique

T_ID	First	Last	MI	MRN
1	Alpha	Omega		9876
2	Primero	Ultimo	M	5432

Record (row)

Field (column)

Free Text in a DB

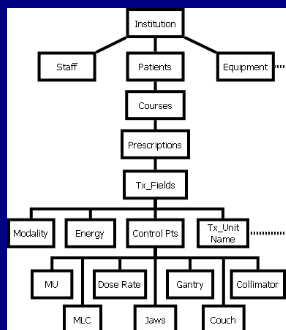
- Not easy to query
- Free text in a column called “Notes”
  - form can vary:
    - The patient is feeling nauseous
    - Symptoms include nausea
- Single element:
  - column “HasNausea” is set to true.
  - Alternatively create a Table Called “Symptoms” and assign each symptom a code; the table will contain codes rather than free text.

Typical Tables in an RT DB

In order to “incorporate” tables into other tables, foreign keys are used to point back to the related tables.

Here, each record in the Tx_Fields table consists of parameters that describe Linac settings. One of the parameters, control points, is a set of records in another table, with a “foreign key” that points back to the Tx_Field record to which it belongs.

DATA DICTIONARY – provides the definitions of the tables and the relationships among them.



Going Paperless

- Digitize paper into images
  - Stop gap measure. OCR?
  - Proof documents (e.g. consent forms)
- Generate files rather than printing
  - No paper involved, data imported into DB
  - proof documents (MU calc results, etc)
- Populate fields in database records
  - Depending on EMR, may let you define your own
- Online Documentation (Wiki)

UIHC Wiki

Infrastructure Summary

- Computers with RT applications:
  - Treatment Planning System
  - Treatment Management System (“V&R”)
  - Treatment Delivery System (Linac control console)
- Servers
  - DB servers
  - Web server
  - Wiki host server
- Archiving and Backup
- Networks
  - Data transfers, e.g. DICOM: images and RT plans
  - Access to servers

II. Clinical Trials

- Well defined
  - process: protocol
  - deliverables (data and formats)
  - goals
- QA of data
- Safe way to implement new treatment schemes

Oncology Groups

TABLE 11.2 Organizations participating in the NIH cooperative group program

Organization	URL
American College of Radiology Imaging Network	http://www.acrin.org
American College of Surgeons Oncology Group	http://www.acosog.org
Cancer and Leukemia Group B	http://www.calgb.org
Children's Oncology Group	http://www.childrensoncologygroup.org
Eastern Cooperative Oncology Group	http://www.ecog.org
European Organisation for Research and Treatment of Cancer	http://www.eortc.be
Gynecologic Oncology Group	http://www.gog.org
National Cancer Institute of Canada, Clinical Trials Group	http://www.ctg.queensu.ca
National Surgical Adjuvant Breast and Bowel Project	http://www.nsabp.pitt.edu
North Central Cancer Treatment Group	http://ncctg.mayo.edu
Radiation Therapy Oncology Group	http://www.rtog.org
Southwest Oncology Group	http://www.swog.org

**Table 11.2 from Siochi, Information resources for radiation oncology, Ch. 11 of a forthcoming book: Informatics in Radiation Oncology, G. Starkschall, B. Curran, editors.*

Participation

- Submission of test (credentialing) data
- Ensures ability of clinic to:
  - Meet the planning objectives
  - Provide the data in the correct format
  - Perform the process correctly
- Example: http://atc.wustl.edu/home/about.html
- "…support NCI sponsored advanced technology clinical trials, particularly those requiring digital data submission…"

Data formats

- DICOM
- RTOG
- Advanced Technology Consortium http://atc.wustl.edu/resources/index.html
- For tools to help with submission of data in these formats

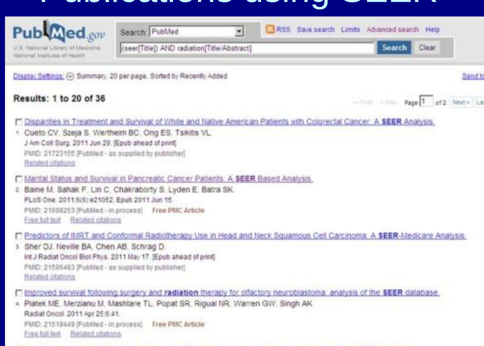
Results of Clinical Trials

- Published data in journals
  - http://atc.wustl.edu/publications/ATC_publications.pdf
    Example:
  - Bradley J, Graham M, Winter K, Purdy J, Komaki R, Roa W, Ryu J, Bosch W, and Emami B: Toxicity and Outcome Results of RTOG 9311; a Phase I/II Dose Escalation Study using Three-Dimensional Conformal Radiotherapy in Patients with Inoperable Non-Small Cell Lung Carcinoma. Int J Radiat Onco Biol Phys, 61 (2):318-328, 2005.
- Some databases available for research
  - http://atc.wustl.edu/resources/data_request.html
  - Request includes a research plan
- http://www.cancer.gov/clinicaltrials/results

Other Clinical Data

- Resources for meta-analysis
  - Using multiple databases from multiple, related clinical trials
- Databases of Cancer Registries
  - Surveillance, Epidemiology and End Results (SEER)
  - Cancer incidence, survival statistics
  - Submission to state/national registries: tedious manual entry, some tools to automate
    - METRIQ for MOSAIQ using Mosaiq Connect
    - Aria, HL7 Information Exchange Manager, OncoLog

Publications using SEER



III. Data Mining

- Often misused phrase
  - Incorrectly applied to standard queries (e.g. incidence of a particular type of cancer by region)
  - Does not apply when there is a specific question to answer
- Data Mining:
  - Knowledge Discovery
  - Looking for patterns
  - More generic question
  - Multiple queries plus algorithms to analyze results

Data mining tasks

- Supervised Learning
  - Classification (predicting “labels” e.g. high, medium, low risk)
  - Numerical prediction (regression) (e.g. using neural nets)
- Unsupervised Learning
  - Clustering
  - Association Rules

Classification

- One attribute is the classifier
  - E.g. risk: high, medium, low
- Training set consists of
  - Previously classified patients
  - variety of attributes (columns in record)
- Algorithms find combinations of attributes most likely to yield a given classification

Classification in RT: example


BMC Med Inform Decis Mak. 2008 Sep 21;8:41.

Non-compliance with a postmastectomy radiotherapy guideline: decision tree and cause analysis.

Razavi AR, Gill H, Ahlfeldt H, Shahsavar N.

Department of Biomedical Engineering, Division of Medical Informatics, Linköping University, Sweden. amira@imt.liu.se

Abstract

BACKGROUND: The guideline for postmastectomy **radiotherapy** (PMRT), which is prescribed to reduce recurrence of breast cancer in the chest wall and improve overall survival, is not always followed. Identifying and extracting important patterns of non-compliance are crucial in maintaining the quality of care in Oncology.

METHODS: Analysis of 759 patients with malignant breast cancer using decision tree induction (DTI) found patterns of non-compliance with the guideline. The PMRT guideline was used to separate cases according to the recommendation to receive or not receive PMRT. The two groups of patients were analyzed separately. Resulting patterns were transformed into rules that were then compared with the reasons that were extracted by manual inspection of records for the non-compliant cases.

RESULTS: Analyzing patients in the group who should receive PMRT according to the guideline did not result in a robust decision tree. However, classification of the other group, patients who should not receive PMRT treatment according to the guideline, resulted in a tree with nine leaves and three of them were representing non-compliance with the guideline. In a comparison between rules resulting from these three non-compliant patterns and manual inspection of patient records, the following was found: In the decision tree, presence of perigland growth is the most important variable followed by number of malignantly invaded lymph nodes and level of Progesterone receptor. DNA index, age, size of the tumor and level of Estrogen receptor are also involved but with less importance. From manual inspection of the cases, the most frequent pattern for non-compliance is age above the threshold followed by near cut-off values for risk factors and unknown reasons.

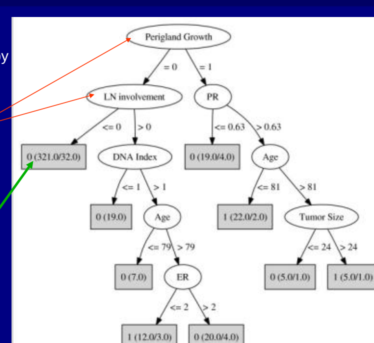
CONCLUSION: Comparison of patterns of non-compliance acquired from **data mining** and manual inspection of patient records demonstrates that not all of the non-compliances are repetitive or important. There are some overlaps between important variables acquired from manual inspection of patient records and **data mining** but they are not identical. **Data mining** can highlight non-compliance patterns valuable for guideline authors and for medical audit. Improving guidelines by using feedback from **data mining** can improve the quality of care in oncology.

Classification by decision tree induction

Classification by decision tree induction

Attributes are in ovals

Classifier: compliance=0 Non-compliance = 1



How do you determine which attribute to start with in order to create a tree that reliably predicts new cases?

From: BMC Med Inform Decis Mak. 2008; 8: 41.

Published online 2008 September 21. doi: 10.1186/1472-6947-8-41

Clustering: Error detection

Phys Med Biol. 2007 Nov 7;52(21):6511-24. Epub 2007 Oct 16.

Towards the development of an error checker for radiotherapy treatment plans: a preliminary study.

Azmandian F, Kaeli D, Dy JG, Hutchinson E, Ancukiewicz M, Niemierko A, Jiang SB.

Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115, USA.

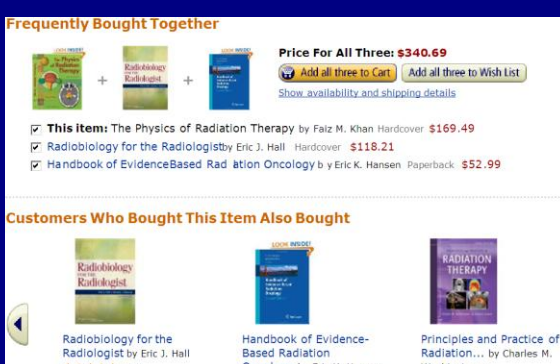
"…The basic idea of using clustering algorithms for outlier detection is to first cluster (based on the treatment parameters) a large number of patient treatment plans. Then, when checking a new treatment plan, the parameters of the plan will be tested to see whether or not they belong to the established clusters. If not, they will be considered as 'outliers' and therefore highlighted to catch the attention of the human chart checkers…"

AAPM/COMP 2011 Vancouver 29

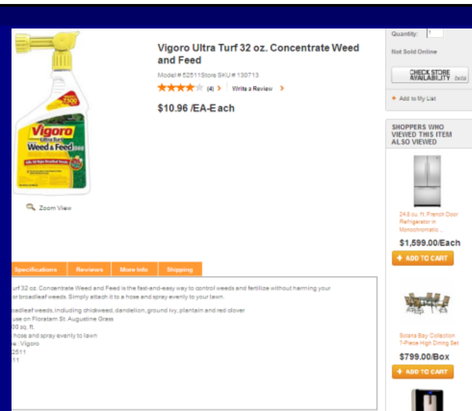
Association

- Combinations of attributes are examined to see if they occur together
- If A then B
- Market Basket Analysis

AAPM/COMP 2011 Vancouver 30



AAPM/COMP 2011 Vancouver 31



AAPM/COMP 2011 Vancouver 32

Association: SRS Planning

Med Phys. 2006 Jul;33(7):2532-40.

Automated gamma knife radiosurgery treatment planning with image registration, data-mining, and Nelder-Mead simplex optimization.

Lee KJ, Barber DC, Walton L.

Unit of Academic Radiology, University of Sheffield, Sheffield, South Yorkshire, United Kingdom. k.j.lee@sheffield.ac.uk

Abstract

Gamma knife treatments are usually planned manually, requiring much expertise and time. We describe a new, fully automatic method of treatment planning. The treatment volume to be planned is first compared with a database of past treatments to find volumes closely matching in size and shape. The treatment parameters of the closest matches are used as starting points for the new treatment plan. Further optimization is performed with the Nelder-Mead simplex method: the coordinates and weight of the isocenters are allowed to vary until a maximally conformal plan specific to the new treatment volume is found. The method was tested on a randomly selected set of 10 acoustic neuromas and 10 meningiomas. Typically, matching a new volume took under 30 seconds. The time for simplex optimization, on a 3 GHz Xeon processor, ranged from under a minute for small volumes (<1000 cubic mm, 2-3 isocenters), to several tens of hours for large volumes (>30,000 cubic mm, >20 isocenters). In 8/10 acoustic neuromas and 8/10 meningiomas, the automatic method found plans with conformation number equal or better than that of the manual plan. In 4/10 acoustic neuromas and 5/10 meningiomas, both overtreatment and undertreatment ratios were equal or better in automated plans. In conclusion, **data-mining** of past treatments can be used to derive starting parameters for treatment planning. These parameters can then be computer optimized to give good plans automatically.

Plans with tumors having a certain size and shape also have certain treatment parameters.

Past treatments were data mined to derive starting parameters for plan optimization.

Not exactly "association" since the desired rule was known up front or assumed.

Massive Amounts of Data

- Patterns hidden in large amounts of data
- Uncovered by algorithms
- Data still needs to be prepared (e.g. selection of attributes and classifier)
- Can come from many related sources
- Similar to meta-analysis
- Data pooling
- NEED INFRASTRUCTURE!

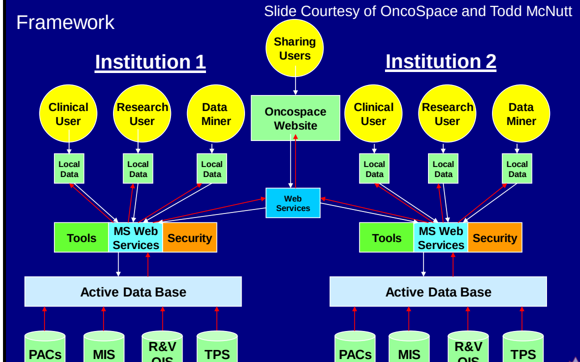
Infrastructure for Large Data Sets

- Too much data to move around, take the analysis to the data (Sky Server approach)
- OncoSpace (McNutt, Wong at Johns Hopkins)
- Custom procedures and functions are part of the database
- Web Services
- Multiple databases

OncoSpace Example Infrastructure

Slide Courtesy of OncoSpace and Todd McNutt

Framework

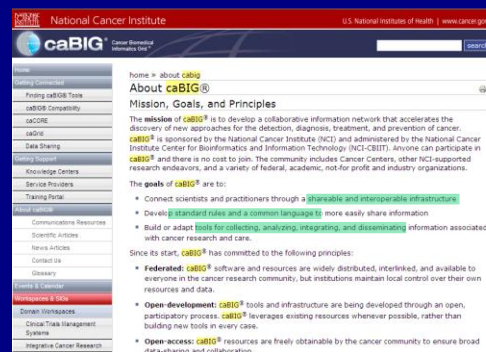


Radiation Oncology Data Alliance (RODA)

- radiation oncology specific data registry
- Data collected during routine use of the EMR (Mosaiq) in practices world-wide
- aggregated to a central data repository
- analyze and correlate the data with patient outcomes

From:
http://www.elekta.com/healthcare_international_radiation_oncology_data_alliance.php

Cancer Biomedical Informatics Grid



Comparative Effectiveness Research

- Systematic Literature Review
  - Randomized controlled trials
  - Observational, non-randomized studies
  - Determine gaps in knowledge
- New Studies
- Resources to help conduct research through AHRQ
- http://www.effectivehealthcare.ahrq.gov/index.cfm

CER data sources

- often depend on aggregating data housed in disparate databases
- federal, state and Thomson Reuters MarketScan® databases to be made accessible to researchers using next-generation analytical applications
- Thomson Reuters will develop a pilot system linking multiple healthcare data sources

From Thomson Reuters' Press Release

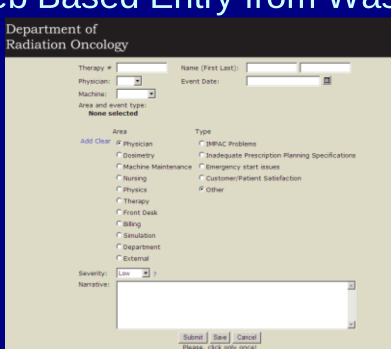
IV. IT for clinical operations

- Databases are subject to business analytics in successful companies
- Health care has been slow to adopt tools that have been developed to improve business operations
- DBs can be used to improve safety and quality

Event Log Database

- Web based system from Wash U.
- Record process deviations
- UIHC application to analyze deviations
  - Standard queries: filters
  - Intermediate complexity: simple association rules to find interactions that are more prone to error
  - Future: Tools for data mining

Web Based Entry from Wash U


Department of
Radiation Oncology

Therapy #
Name (First Last):
Physician:
Event Date:
Machine:
Area and event type:
None selected

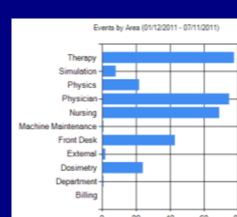
Area
Add Clear
Physician
Dosimetry
Machine Maintenance
Nursing
Physics
Therapy
Front Desk
Billing
Simulation
Department
External

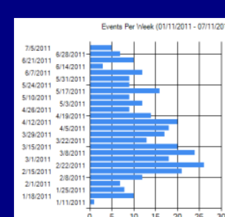
Type
IMPAC Problems
Inadequate Prescription Planning Specifications
Emergency start issues
Customer/Patient Satisfaction
Other

Severity: Low ?
Narrative:

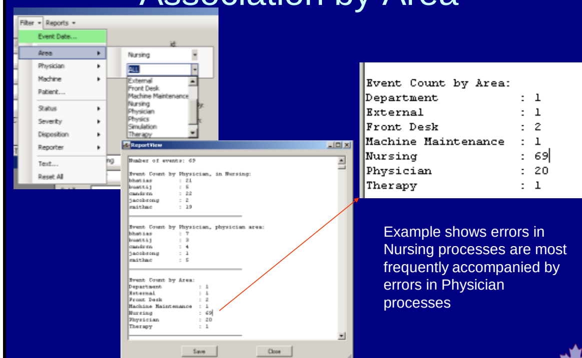
Submit Save Cancel
Please, click only once!

Event Log Viewer – UIHC Events by Area and Per Week





Association by Area



Example shows errors in Nursing processes are most frequently accompanied by errors in Physician processes

RFID databases

- DBs of real-time location
  - Resources (staff, equipment)
  - Patients
- Read RFID DB and compare to schedules
- Optimize the assignment of resources to patients

Business Intelligence

- Decision Support System
  - Online analytical processing, analytics, data mining, text mining, predictive analytics, benchmarking, business performance management
  - Often uses data warehouses
  - Commercially available tools
- Concepts could be adopted in Healthcare
  - Safety, quality, cost effectiveness, CER
- HOWEVER we need to be collecting data!

Summary

- Radiation Oncology Informatics deals with
  - The IT infrastructure to plan and deliver radiotherapy and participate in clinical trials
  - The information science needed to analyze clinical data
  - The infrastructure to gather massive amounts of data
  - The improvement of clinical practice, safety, and quality