

## Informatics in Radiation Oncology

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

AAPM/COMP 2011 Vancouver 

## 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

AAPM/COMP 2011 Vancouver 

## 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.

AAPM/COMP 2011 Vancouver 

## 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

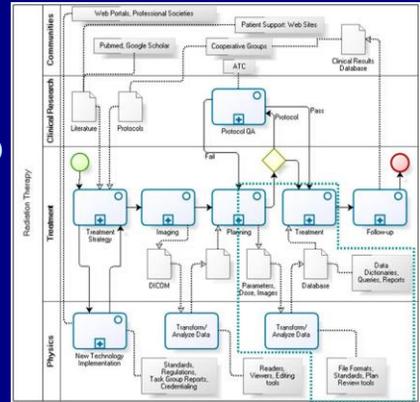
AAPM/COMP 2011 Vancouver 

## Infrastructure ⇔ Process Map

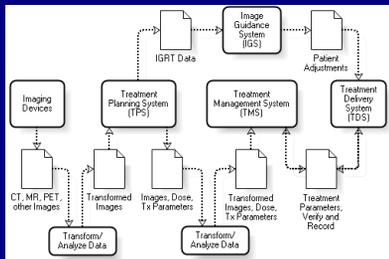
- Radiation Oncology is Data Intensive
- Follow the Data
- Processes: What do we do with this data?
- Paper vs Paperless
  - Different processes
  - Different IT infrastructure

## Data Flow in RO

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



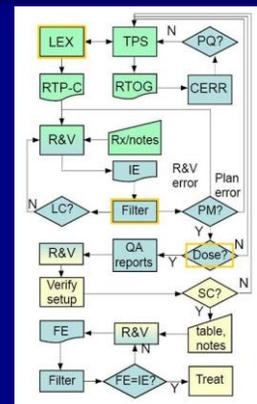
## Distributed system data flow



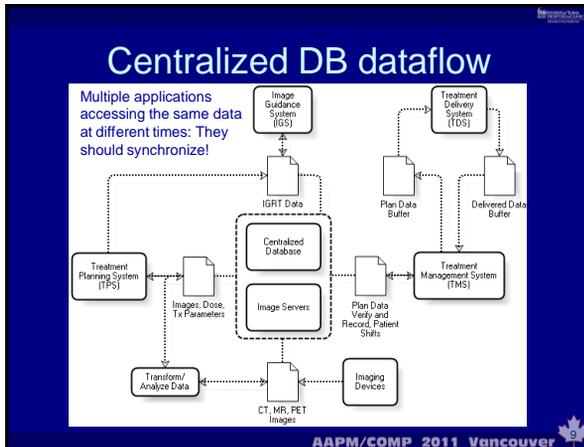
Redundant data living in many places: INFORMATION should match. (Data might be stored in different forms but mean the same thing.)

## Clinical Interactions, paperless checks

Physicists  
Dosimetrists/Physicians  
Therapists  
**In-House Software**



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



## 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)

AAPM/COMP 2011 Vancouver

## 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)

AAPM/COMP 2011 Vancouver

## 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.

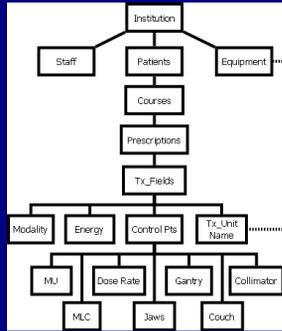
AAPM/COMP 2011 Vancouver

## 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.



AAPM/COMP 2011 Vancouver

## 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)

AAPM/COMP 2011 Vancouver

## UIHC Wiki

AAPM/COMP 2011 Vancouver

## 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

AAPM/COMP 2011 Vancouver

## 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	<a href="http://www.acrin.org">http://www.acrin.org</a>
American College of Surgeons Oncology Group	<a href="http://www.acosog.org">http://www.acosog.org</a>
Cancer and Leukemia Group B	<a href="http://www.calgb.org">http://www.calgb.org</a>
Children's Oncology Group	<a href="http://www.childrensoncologygroup.org">http://www.childrensoncologygroup.org</a>
Eastern Cooperative Oncology Group	<a href="http://www.ecog.org">http://www.ecog.org</a>
European Organisation for Research and Treatment of Cancer	<a href="http://www.eortc.be">http://www.eortc.be</a>
Gynaecologic Oncology Group	<a href="http://www.gog.org">http://www.gog.org</a>
National Cancer Institute of Canada, Clinical Trials Group	<a href="http://www.ctg.queensu.ca">http://www.ctg.queensu.ca</a>
National Surgical Adjuvant Breast and Bowel Project	<a href="http://www.nsabp.pitt.edu">http://www.nsabp.pitt.edu</a>
North Central Cancer Treatment Group	<a href="http://nccctg.mayo.edu">http://nccctg.mayo.edu</a>
Radiation Therapy Oncology Group	<a href="http://www.rtog.org">http://www.rtog.org</a>
Southwest Oncology Group	<a href="http://www.swog.org">http://www.swog.org</a>

*\*Table 11.2 from Stoeckl, 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](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](http://atc.wustl.edu/resources/data_request.html)
  - Request includes a research plan
- <http://www.cancer.gov/clinicaltrials/results>

AAPM/COMP 2011 Vancouver

## 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 Mosaic Connect
    - Aria, HL7 Information Exchange Manager, OncoLog

AAPM/COMP 2011 Vancouver

## Publications using SEER

The screenshot shows a PubMed search interface with the search term 'SEER'. The search criteria are 'see[Title] AND radiation[Title:Abstract]'. The results are sorted by 'Recently Added' and show 36 results. The first three results are:

1. **Disparities in Treatment and Survival of White and Native American Patients with Colorectal Cancer: A SEER Analysis.** Cuello CV, Szaga S, Wertheim BC, Ong ES, Takkis VL. *J Am Coll Surg*. 2011 Jun 29. [Epub ahead of print]. PMID: 21723350. Published - as supplied by publisher. [Related citations](#)
2. **Mental Status and Survival in Pancreatic Cancer Patients: A SEER-Based Analysis.** Baine M, Sahak F, Lin C, Chakraborty S, Lyden E, Batra SK. *PLoS One*. 2011;6(6):e21910. Epub 2011 Jun 15. PMID: 21698253. Published - in process. Free PMC Article. [View full text](#) [Related citations](#)
3. **Prevalence of IMRT and Conformal Radiotherapy Use in Head and Neck Squamous Cell Carcinoma: A SEER-Medicare Analysis.** Sher DJ, Henley SA, Chen AB, Sroczyn D. *Int J Radiat Oncol Biol Phys*. 2011 May 17. [Epub ahead of print]. PMID: 21556463. Published - as supplied by publisher. [Related citations](#)

AAPM/COMP 2011 Vancouver

## 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

AAPM/COMP 2011 Vancouver

## 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

AAPM/COMP 2011 Vancouver

## 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

AAPM/COMP 2011 Vancouver

## 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.**  
 Balasub G, Gillis A, Andersson J, Sjöström A, J  
 Department of Biomedical Engineering, Division of Medical Informatics, Linköping University, Sweden. amra@mit.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 peritumoral growth is the most important variable followed by number of marginally 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.  
**CONCLUSIONS:** 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

AAPM/COMP 2011 Vancouver

## Classification in RT: example

From: [BMC Med Inform Decis Mak. 2008; 8: 41.](#)

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

AAPM/COMP 2011 Vancouver

## Clustering: Error detection

Proc Med Biol, 2007 Nov; 7:5221-5511-24. Epub 2007 Oct 18.

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

Somandottil F, Knafl D, O'Leary J, Hutchinson E, Appalrajesh K, Namdeo A, Jago 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

## Association

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

AAPM/COMP 2011 Vancouver

### Frequently Bought Together



Price For All Three: **\$340.69**

[Add all three to Cart](#) [Add all three to Wish List](#)

[Show availability and shipping details](#)

- This item:** The Physics of Radiation Therapy by Faiz M. Khan Hardcover **\$169.49**
- Radiobiology for the Radiologist by Eric J. Hall Hardcover **\$118.21**
- Handbook of Evidence-Based Radiation Oncology by Eric K. Hansen Paperback **\$52.99**

### Customers Who Bought This Item Also Bought



Radiobiology for the Radiologist by Eric J. Hall  
★★★★★ (14)



Handbook of Evidence-Based Radiation Oncology by Eric K. Hansen



Principles and Practice of Radiation... by Charles M. Washington...

AAPM/COMP 2011 Vancouver



### Vigoro Ultra Turf 32 oz. Concentrate Weed and Feed

Model# 52211506 SKU# 150713

★★★★★ (6) [Write a Review](#)

**\$10.96/EA-Each**

[Zoom View](#)

[Specifications](#) [Reviews](#) [More Info](#) [Shipping](#)

4932 oz. Concentrate Weed and Feed is the best choice only to control weeds and fertilizer without harming your or landscape needs. Simply attach to a hose and spray evenly to your lawn.

Control weeds including crabgrass, dandelion, ground ivy, plantain and red clover.

Use on: Turfgrass: St. Augustine Grass

32 oz. B.

Feed and spray evenly to lawn.

Vigoro

2011

11

Quantity:

Not Sold Online

[CHECK OUT NOW](#) [DONE](#)

[Add to My List](#)

SHoppers WHO VIEWED THIS ITEM ALSO VIEWED



24 in. x 36 in. French Door Refrigerator

Monochrome

**\$1,899.00/Each**

[ADD TO CART](#)



30 in. Bay Cabinet

7-Phase High Dining Set

**\$799.00/Box**

[ADD TO CART](#)

AAPM/COMP 2011 Vancouver

## Association: SRS Planning

Med Phys. 2009 Jun;36(7):2832-40.

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

Lee S, Saher DC, Wilson L.

Unit of Academic Radiology, University of Sheffield, Sheffield, South Yorkshire, United Kingdom, s.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 Neider-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 over-treatment and under-treatment 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.

AAPM/COMP 2011 Vancouver

## 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!**

AAPM/COMP 2011 Vancouver

## Infrastructure for Large Data Sets

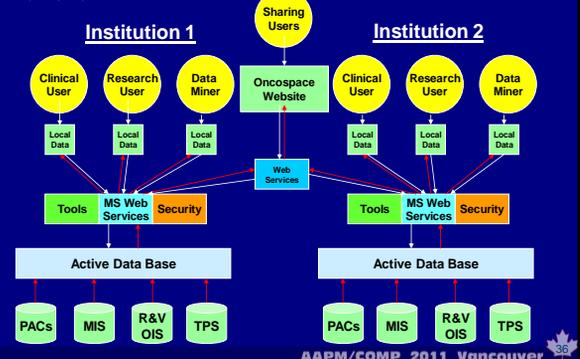
- Too much data to move around, take the analysis to the data (Sky Server approach)
- OncoSpace (McNutt, Wong at Johns Hopkins)
- Custom procedures and functions are part of the database
- Web Services
- Multiple databases

AAPM/COMP 2011 Vancouver

## OncoSpace Example Infrastructure

Framework

Slide Courtesy of OncoSpace and Todd McNutt



AAPM/COMP 2011 Vancouver

## Radiation Oncology Data Alliance (RODA)

- radiation oncology specific data registry
- Data collected during routine use of the EMR (Mosaik) 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](http://www.elekta.com/healthcare_international_radiation_oncology_data_alliance.php)

AAPM/COMP 2011 Vancouver 37

## Cancer Biomedical Informatics Grid

home » about caBIG

### About caBIG®

#### Mission, Goals, and Principles

The **mission** of caBIG® is to develop a collaborative information network that accelerates the discovery of new approaches for the detection, diagnosis, treatment, and prevention of cancer. caBIG® is sponsored by the National Cancer Institute (NCI) and administered by the National Cancer Institute Center for Biinformatics and Information Technology (NCI-CBIT). Anyone can participate in caBIG® and there is no cost to join. The community includes Cancer Centers, other NCI-supported research endeavors, and a variety of federal, academic, not-for-profit and industry organizations.

The **goals** of caBIG® are to:

- Connect scientists and practitioners through a **shareable and interoperable infrastructure**.
- Develop **standard rules and a common language** to more easily share information.
- Build or adapt **tools for collecting, analyzing, integrating, and disseminating** information associated with cancer research and care.

Since its start, caBIG® has committed to the following principles:

- **Federated:** caBIG® software and resources are widely distributed, interlinked, and available to everyone in the cancer research community, but institutions maintain local control over their own resources and data.
- **Open-development:** caBIG® tools and infrastructure are being developed through an open, participatory process. caBIG® leverages existing resources wherever possible, rather than building new tools in every case.
- **Open-access:** caBIG® resources are freely obtainable by the cancer community to ensure broad data-sharing and collaboration.

AAPM/COMP 2011 Vancouver 38

## 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>

AAPM/COMP 2011 Vancouver 39

## 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

AAPM/COMP 2011 Vancouver 40

## 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

AAPM/COMP 2011 Vancouver

## 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

AAPM/COMP 2011 Vancouver

## Web Based Entry from Wash U

Department of Radiation Oncology

Therapy # [ ] Name (First Last): [ ]  
 Physician: [ ] Event Date: [ ]  
 Machine: [ ]

Area and event type:  
 Name selected

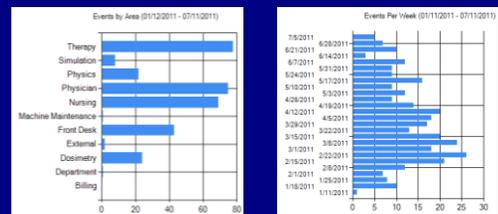
Area	Type
<input type="checkbox"/> Physician	<input type="checkbox"/> IMAC Problems
<input type="checkbox"/> Dosimetry	<input type="checkbox"/> Inadequate Prescription Planning Specifications
<input type="checkbox"/> Machine Maintenance	<input type="checkbox"/> Emergency start issues
<input type="checkbox"/> Nursing	<input type="checkbox"/> Customer/Patient Satisfaction
<input type="checkbox"/> Physics	<input type="checkbox"/> Other
<input type="checkbox"/> Therapy	
<input type="checkbox"/> Front Desk	
<input type="checkbox"/> Calling	
<input type="checkbox"/> Simulation	
<input type="checkbox"/> Department	
<input type="checkbox"/> External	

Severity: [ ]  
 Narrative: [ ]

[ Save ] [ Save ] [ Cancel ]  
 Please, click only once!

AAPM/COMP 2011 Vancouver

## Event Log Viewer – UIHC Events by Area and Per Week



AAPM/COMP 2011 Vancouver

## Association by Area

Event Count by Area:

Department	: 1
External	: 1
Front Desk	: 2
Machine Maintenance	: 1
Nursing	: 69
Physician	: 20
Therapy	: 1

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

AAPM/COMP 2011 Vancouver

## 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

AAPM/COMP 2011 Vancouver

## 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!

AAPM/COMP 2011 Vancouver

## 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

AAPM/COMP 2011 Vancouver