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 in radiation oncology, Ch. 11 of a forthcoming book, Informatics in Radiation Oncology, G. Starkschall, B. Currin, 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

Centralized DB dataflow

Multiple applications accessing the same data at different times: They should synchronize!

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

<table>
<thead>
<tr>
<th>T_ID</th>
<th>First</th>
<th>Last</th>
<th>MI</th>
<th>MRN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Alpha</td>
<td>Omega</td>
<td>9876</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Primero</td>
<td>Ultimo</td>
<td>M</td>
<td>5432</td>
</tr>
</tbody>
</table>

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

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](http://atc.wustl.edu/home/about.html)
- “…support NCI sponsored advanced technology clinical trials, particularly those requiring digital data submission…”

Data formats

- DICOM
- RTOG
- For tools to help with submission of data in these formats

Oncology Groups

<table>
<thead>
<tr>
<th>Organization</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>American College of Radiology Imaging Network</td>
<td><a href="http://www.acrino.org">http://www.acrino.org</a></td>
</tr>
<tr>
<td>American College of Surgeons Oncology Group</td>
<td><a href="http://www.acgos.org">http://www.acgos.org</a></td>
</tr>
<tr>
<td>Children’s Oncology Group</td>
<td><a href="http://www.childrensoncologygroup.org">http://www.childrensoncologygroup.org</a></td>
</tr>
<tr>
<td>Eastern Cooperative Oncology Group</td>
<td><a href="http://www.ecog.org">http://www.ecog.org</a></td>
</tr>
<tr>
<td>Gynecologic Oncology Group</td>
<td><a href="http://www.gog.org">http://www.gog.org</a></td>
</tr>
<tr>
<td>National Cancer Institute of Canada, Clinical Trials Group</td>
<td><a href="http://www.nci.org.cn">http://www.nci.org.cn</a></td>
</tr>
<tr>
<td>National Surgical Adjuvant Breast and Bowel Project</td>
<td><a href="http://www.nsabp.pitt.edu">http://www.nsabp.pitt.edu</a></td>
</tr>
<tr>
<td>North Central Cancer Treatment Group</td>
<td><a href="http://www.ncctg.mayo.edu">http://www.ncctg.mayo.edu</a></td>
</tr>
<tr>
<td>Radiation Therapy Oncology Group</td>
<td><a href="http://www.rtog.org">http://www.rtog.org</a></td>
</tr>
<tr>
<td>Southwest Oncology Group</td>
<td><a href="http://www.swog.org">http://www.swog.org</a></td>
</tr>
</tbody>
</table>

*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.*
Results of Clinical Trials

• Published data in journals
  Example:

• 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 in RT: example

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?

Clustering: Error detection

...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...

Association

• Combinations of attributes are examined to see if they occur together
• If A then B
• Market Basket Analysis
Association: SRS Planning

Plants 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
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:

Cancer Biomedical Informatics Grid

Cancer Biomedical Informatics Grid (caBIG)

- National Cancer Institute
- US National Institutes of Health
- National Library of Medicine

Mission, Goals, and Principles

- to develop a collaborative information network that accelerates the discovery of new approaches for the detection, diagnosis, treatment, and prevention of cancer
- to sponsor the Cancer Biomedical Informatics Grid (caBIG) to provide an infrastructure for cancer researchinformatics and information technology (IT) to support cancer research in the areas of collaboration, analysis, and dissemination of research data, and to coordinate the use of IT tools and resources for cancer research and care

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

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